



Article **PDD-ET: Parkinson's Disease Detection Using ML Ensemble Techniques and Customized Big Dataset**

Kalyan Chatterjee ¹, Ramagiri Praveen Kumar ¹, Anjan Bandyopadhyay ², Sujata Swain ², Saurav Mallik ^{3,4,*}, Aimin Li ⁵ and Kanad Ray ^{6,7,8,*}

- ¹ Department of Computer Science & Engineering, Nalla Malla Reddy Engineering College, Hyderabad 500088, India
- ² School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar 751024, India; anjan.bandyopadhyayfcs@kiit.ac.in (A.B.); sujata.swainfcs@kiit.ac.in (S.S.)
- ³ Department of Environmental Health, Harvard T H Chan School of Public Health, Boston, MA 02115, USA
- ⁴ Department of Pharmacology & Toxicology, The University of Arizona, Tucson, MA 85721, USA
- ⁵ School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; liaiminmail@gmail.com
- ⁶ Amity School of Applied Sciences, Amity University Rajasthan, Jaipur 303002, India
- ⁷ Facultad de CienciasFisico-Matematicas, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y AV. 18 sur, Col. San Manuel Ciudad Universitaria, Pueble Pue 72570, Mexico
- ⁸ Faubert Lab, Ecole d'optométrie, Université de Montréal, Montréal, QC H3T1P1, Canada
- Correspondence: sauravmtech2@gmail.com or smallik@hsph.harvard.edu (S.M.); kanadray00@gmail.com (K.R.)

Abstract: Parkinson's disease (PD) is a neurological disorder affecting the nerve cells. PD gives rise to various neurological conditions, including gradual reduction in movement speed, tremors, limb stiffness, and alterations in walking patterns. Identifying Parkinson's disease in its initial phases is crucial to preserving the well-being of those afflicted. However, accurately identifying PD in its early phases is intricate due to the aging population. Therefore, in this paper, we harnessed machine learning-based ensemble methodologies and focused on the premotor stage of PD to create a precise and reliable early-stage PD detection model named PDD-ET. We compiled a tailored, extensive dataset encompassing patient mobility, medication habits, prior medical history, rigidity, gender, and age group. The PDD-ET model amalgamates the outcomes of various ML techniques, resulting in an impressive 97.52% accuracy in early-stage PD detection. Furthermore, the PDD-ET model effectively distinguishes between multiple stages of PD and accurately categorizes the severity levels of patients affected by PD. The evaluation findings demonstrate that the PDD-ET model outperforms the SVR, CNN, Stacked LSTM, LSTM, GRU, Alex Net, [Decision Tree, RF, and SVR], Deep Neural Network, HOG, Quantum ReLU Activator, Improved KNN, Adaptive Boosting, RF, and Deep Learning Model techniques by the approximate margins of 37%, 30%, 20%, 27%, 25%, 18%, 19%, 27%, 25%, 23%, 45%, 40%, 42%, and 16%, respectively.

Keywords: Parkinson's disease detection (PDD); ensemble techniques (ETs); big dataset; adaptive boosting (AB); random forest (RF); LSTM; GRU; SVR; KNN; premotor

1. Introduction

Parkinson's disease (PD) represents a severe form of neurocellular disorder that can lead to the deterioration of the central nervous system in affected individuals [1]. The disease is transmissible among patients and exhibits varying symptoms across different cases, underscoring the necessity for early PD detection, benefiting both patients and others. To create an effective PD detection model, careful attention must be given to monitoring the premotor phase of the disease.

PD-affected individuals may exhibit recognizable symptoms such as tremors, limb stiffness, changes in walking patterns, and balance issues. However, these symptoms



Citation: Chatterjee, K.; Kumar, R.P.; Bandyopadhyay, A.; Swain, S.; Mallik, S.; Li, A.; Ray, K. PDD-ET: Parkinson's Disease Detection Using ML Ensemble Techniques and Customized Big Dataset. *Information* **2023**, *14*, 502. https://doi.org/ 10.3390/info14090502

Academic Editor: Francesco Beritelli

Received: 13 August 2023 Revised: 31 August 2023 Accepted: 6 September 2023 Published: 13 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). develop gradually within patients. Additionally, there are two distinct categories of PD symptoms: (i) motor-related symptoms tied to movement and (ii) non-motor symptoms unrelated to motion. Non-motor symptoms tend to be more severe and encompass aspects including depression, sleep pattern abnormalities, diminished olfactory function, and cognitive decline. Over 52 billion individuals in the USA are affected by PD annually, with a global impact surpassing 10 million cases. Detecting PD early, categorizing its stages within patients, initiating prompt treatment, and identifying alleviating symptoms and treatments are essential strategies to combat PD [2]. Hence, early detection is vital to slowing PD progression and potentially developing curative therapies.

Despite several PD symptoms and diagnostic tests, the exact diagnosis of PD remains elusive [2]. These diagnostic indicators are employed collectively to identify PD, with various biomarkers scrutinized to recognize the disease's early stages. PD treatments focus on alleviating symptoms rather than halting or reversing the disease's progression within patients.

Before developing the novel PDD-ET model, existing PD detection models were explored. This exploration revealed various measurement methods, including speech data [3–6], gait dynamics [7], force trajectory information [8], olfactory recognition measurements [9], and involuntary fluctuations in cardiovascular activity [10]. In 2020, Illner et al. employed a serrulate-based pitch estimator (SBPE) to detect PD-related speech abnormalities [11]. Although SBPE demonstrated some promising experimental outcomes distinguishing PD-affected patients, noise resilience remains a concern. Improved, robust PD detection algorithms/models are needed to address this challenge. Solana-Lavalle et al. developed PDD models using vocal features [12]. Maachi et al. introduced a 1D convolutional neural network (CNN) for PD detection based on gait signals [13]. However, both speech and gait techniques' PD detection performance could be more reliable, especially in the presence of background noise. Additionally, these techniques demand dedicated equipment and controlled environments [14]. Wagner et al. proposed a wavelet-based PD detection method using data collected from PD-affected patients [15]. Gallego et al. presented a motor impairment-based PDD model using smartphones [16]. This model also incorporated statistics such as covariance, skewness, and temporal data. Dinov et al. utilized diverse data sources for their PDD model, including imaging, genetics, clinical data, and demographics [17]. Other handwriting evaluation-based PDD models are also discussed [18,19].

The literature survey underscores the need for an accurate, early-stage PDD model to acquire crucial information about PD patients for controlling the disease's progression within their bodies. Data-driven and model-based methods have been established for this purpose, with substantial involvement from machine learning (ML) and deep learning (DL) techniques [20–23]. These tools offer valuable insights for the classification and early identification of PD. Various ML and DL approaches have been utilized in prior research to tackle the issue of the PDD challenge. For example, SVM has been used to detect and classify PD-related dysphonia problems by extracting nonlinear features through SVM's nonlinear kernels [24]. Random forest (RF) and neural networks have also been applied for PD detection using acoustic analysis of articulation. RF and SVM combined exhibit respectable PDD results. In a comparison of decision tree (DT), regression (Reg.), DM neural network (DMNN), and neural network (NN), NN achieved around 93% accuracy, outperforming other algorithms [25].

In recent times, DL techniques have gained prominence in addressing the PDD challenge due to their ability to handle sequential time-series data and large datasets while managing overfitting, underfitting, and long-term dependencies [26,27]. The memorization capacity of LSTM makes it effective in handling the long-term dependencies inherent in time-series sequential datasets, resulting in improved performance, as seen in detecting *Freezing of Gait* issues. The *Freezing of Gait* phenomenon is significant as a PD indicator. Furthermore, we also consider the premotor phase as a PD indicator [28], encompassing symptoms like rapid eye movement (REM), sleep behavior disorder (SBD), and olfactory loss (OL). The advantages and disadvantages of our proposed PDD-ET model and the existing models are outlined in Tables 1 and 2.

Table 1. Comparison of advantages and disadvantages of the PDD-ET model and the existing models.

Models	Advantages	Disadvantages		
SVR [29]	 (i) It has capability to handle non-linear relationships between input features and the target output. (ii) It can effectively capture complex patterns and variations in the data, making it suitable for situations where the underlying relationships are not linear. 	 (i) Its sensitivity to the selection of hyperparameters, such as the kernel type and regularization parameters. (ii) Poorly tuned hyperparameters can lead to overfitting or underfitting, impacting the model's generalization performance. 		
CNN [30]	 (i) Its capacity to autonomously discern pertinent attributes from unprocessed data, like images, without the need for manual feature engineering. (ii) CNNs are particularly adept at capturing spatial patterns and hierarchies of features, making them well-suited for image-based data like brain scans or medical images commonly used in PD diagnosis. 	 (i) Its susceptibility to overfitting, especially when dealing with limited training data. CNNs contain a large number of learnable parameters, and without enough diverse data, the model might generalize poorly to new, unseen examples. (ii) Regularization techniques and data augmentation can mitigate this issue, but careful attention to dataset size and quality is necessary. 		
Stacked-LSTM [31]	 (i) Its proficiency in capturing and learning long-term dependencies within sequential data. (ii) It can effectively handle complex temporal patterns in time series data related to PD symptoms. 	(i) Its susceptibility to overfitting, especially when dealing with limited training data.		
LSTM [26]	(i) Its aptitude to apprehend extended correlations and sequential patterns within time series data.	 (i) Its sensitivity to hyperparameter tuning. (ii) It struggle when dealing with very short sequences and the temporal dependencies. 		
Decision Tree, RF, SVR [32]	 (i) Diverse Learning, (ii) Bias Reduction, (iii) Robustness, (iv) Capturing Complex Patterns, and (v) High Accuracy. 	 (i) Complexity, (ii) Hyperparameter Tuning, (iii) Computational Intensity, (iv) Data Requirements, and (v) Risk of Overfitting. 		
GRU [33]	(1) It has the capacity to encompass distant connections within sequential data.	(i) It might not capture very long-term dependencies as effectively as more complex models like LSTMs.		
Alex Net [34]	 (i) Its capacity to effectively extract features from images and visual data. (ii) It can automatically learn hierarchical features, making it suitable for processing visual data such as brain scans to PD diagnosis. 	 (i) Its relatively large number of parameters, which can result in higher computational requirements and increased training times. (ii) Alex Net may not be as efficient in capturing intricate spatial patterns for complex image recognition tasks 		
Deep Neural Network [35]	 (i) Feature Learning, (ii) Hierarchy of Features, (iii) Versatility, (iv) Performance, and (v) Transfer Learning. 	 (i) Data Hunger, (ii) Complexity, (iii) Hyperparameter Tuning, (iv) Black Box Nature, and (v) Computational Intensity. 		

Models	Advantages	Disadvantages
HOG [36]	 (i) Efficient Feature Extraction, (ii) Robustness to Illumination and Color Variations, (iii) Interpretability, (iv) Low-Dimensional Representation, and (v) Applicability to Different Data Modalities. 	 (i) Limited Representation of Complex Patterns, (ii) Limited Contextual Information, (iii) Dependency on Image Quality, (iv) Feature Engineering Required, and (v) Limited Application to
Quantum ReLU Activator [37]	(i) Quantum Advantage, (ii) Feature Transformation, and (iii) Non-linearity.	Non-Visual Data. (i) Complexity, (ii) Hardware and Infrastructure, (iii) Lack of Quantum Data, (iv) Interpretability, and (v) Limited Adoption.
Proposed PDD-ET model	 (i) Improved Accuracy and Robustness, (ii) Reduced Overfitting, (iii) Capturing Diverse Patterns, (iv) Handling Noisy Data, (v) Model Flexibility, (vi) Interpretable Insights, (vii) Scalability, Parallelism, and Adaptability. 	(i) Architecture of the model is Complex.(ii) Training of the model is complex and time consuming.

Table 2. Comparison of advantages and disadvantages of the PDD-ET model and the existing models.

The key contributions of this paper are four-fold and stated as follows:

- 1. To the best of our understanding, we are pioneers in constructing a customized big dataset encompassing diverse attributes from both individuals affected by Parkinson's disease and those who are in good health. Furthermore, we are introducing the novel notion of the customized expansive dataset into PD detection and classification.
- 2. To identify PD during its initial phases within the bodies of afflicted individuals to manage its progression, we employed premotor, cerebrospinal fluid (CF), and SPECT indicators to achieve a proficient and precise detection of PD.
- 3. Employing ensemble techniques (ET) to categorize the distinct levels of PD.
- 4. We extensively examine the latest machine learning and deep learning techniques alongside our proposed ensemble technique-driven PDD-ET model. Consequently, the assessed methods encompass shallow machine learning (SML), deep learning (DL), and ensemble learning (EL) approaches.

In this paper, we have formulated an EL-based PDD-ET model designed to gain insights into PD by undergoing training procedures. Through this training process, our model acquires the ability to distinguish between healthy and PD-affected people. The main aim of this paper is to offer a meticulous and comprehensive evaluation of a spectrum of ML, DL, and EL techniques for early-stage PD detection. Additionally, we shed light on the performance of these techniques using our unique and tailored extensive PD dataset, as discussed in Section 2.1.

The subsequent sections of the paper are organized as follows. Section 2 delves into creating our customized big PD dataset and the corresponding ensemble learning methodologies applied. Section 3 describes the system architecture of the PDD-ET model. Section 4 outlines the particulars of our proposed Parkinson's Disease Detection (PDD) model. Moving forward, Section 5 visually presents our experiments' results and facilitates comparison between techniques. Lastly, our paper culminates with concluding remarks and future prospects in Sections 6 and 7.

2. Customized Big PD Dataset and EL Technique

This section outlines the framework based on EL that is employed for the classification of distinct Parkinson's disease (PD) levels and the early-stage detection of PD. Figure 1 illustrates the general progression of the PDD-ET model.



Figure 1. General flow diagram of PDD-ET model.

2.1. Customized Big PD Dataset (CBPDD)

PD is a progressive neurological condition characterized by reduced dopamine levels within the brain. It becomes evident through the deterioration of movement capabilities, leading to symptoms like tremors and stiffness. Speech is also notably affected, with difficulties such as trouble articulating sounds (dysarthria), decreased volume (hypophonia), and limited pitch range (monotone). Moreover, cognitive decline and shifts in emotional state may arise, and the risk of developing dementia is heightened.

The traditional approach to diagnosing PD entails a clinician compiling the patient's neurological history and assessing their motor abilities in various scenarios. Given the lack of a definitive laboratory test for PD diagnosis, this process can be intricate, particularly in the early stages when motor symptoms are not yet pronounced. Monitoring the disease's progression often involves recurrent visits to the clinic by the patient. A valuable enhancement could be developing an efficient screening technique that eliminates the need for in-person clinic visits. Given that individuals with PD exhibit distinct vocal traits, the analysis of voice recordings offers a non-invasive and informative diagnostic tool. Applying ML algorithms to a voice recording dataset makes it possible to achieve accurate PD diagnosis. This approach presents a practical preliminary screening measure to be taken before seeking consultation with a clinician.

To curate our customized big PD dataset, we sourced the PD dataset from Kaggle's official website [38] and the collected patient dataset from various hospitals, as clinical and behavioral data play a crucial role in PD detection and diagnosis. Figure 2 describes the construction of the customized big dataset. Within this dataset, we incorporated a more comprehensive set of features, encompassing 50,583 samples from healthy individuals and 60,958 samples from individuals afflicted by PD. Beyond these two groups, an additional 6000 samples were gathered, encompassing attributes such as patient movement, drug habits, medical history, flexibility, gender, and age group. These data points were meticulously collected from medical clinics. Subsequently, all these samples were amalgamated with the existing PD dataset, introducing them as noise or outliers. Following this, the dataset underwent preprocessing and zero-score normalization, culminating in the creation of our distinctive Customized Big Parkinson's Disease Dataset (CBPDD). This CBPDD was developed with the intent to facilitate the detection and classification of PD.



Figure 2. Construction of Customized Big Dataset for the PDD-ET model.

The used clinical and behavioral data for PD detection are described as follows:

- 1. *Demographic Information:* Basic patient details such as age, gender, and ethnicity are often considered, as they can provide insights into potential risk factors.
- 2. *Medical History:* Previous medical conditions, surgeries, medication history, and family history of PD or related neurological disorders are essential for understanding a patient's overall health.
- 3. *Symptom Profiles:* Detailed descriptions of motor symptoms like tremors, rigidity, bradykinesia (slowness of movement), and postural instability are fundamental indicators of PD.
- 4. *Non-Motor Symptoms:* These include cognitive impairments, sleep disturbances, mood changes, loss of smell (anosmia), and autonomic dysfunctions.
- 5. *UPDRS Assessment:* The Unified Parkinson's Disease Rating Scale (UPDRS) finds extensive application as a commonly employed instrument to evaluate the severity of PD symptoms. It covers both motor and non-motor symptoms.
- 6. *Gait Analysis:* Gait abnormalities are common in PD. Analyzing gait patterns and abnormalities can aid in early detection.
- 7. *Speech Patterns:* PD often affects speech, leading to changes in volume, pitch, articulation, and rhythm. Speech analysis can provide valuable diagnostic insights.
- 8. *Fine Motor Skills:* Assessments of handwriting, finger tapping speed, and dexterity can reveal motor impairments indicative of PD.
- 9. *Reaction Time and Movement Speed:* Slowed reaction times and reduced movement speed can be early indicators of PD.
- 10. *Response to Levodopa:* Observing how a patient responds to levodopa, a common PD medication, can help confirm the diagnosis.
- 11. *Self-Reported Questionnaires:* Patients' self-reported questionnaires about their quality of life, daily activities, and emotional state can contribute to behavioral data.
- 12. *Neuropsychological Testing:* Assessments of cognitive functions like memory, attention, and executive functions can provide additional diagnostic information.
- 13. *Electrophysiological Data:* Electroencephalography (EEG), electromyography (EMG), and other electrophysiological tests can reveal abnormal brain activity and muscle responses.
- 14. *Imaging Data:* Neuroimaging techniques such as MRI, DAT scans, and PET scans can detect structural and functional changes in the brain associated with PD.

Attribute Information:

- MDVP: Fo(Hz)—Average vocal fundamental frequency;
- MDVP: Fhi(Hz)—Maximum vocal fundamental frequency;
- MDVP: Flo(Hz)—Minimum vocal fundamental frequency;
- MDVP: Jitter(%);
- MDVP: Jitter(Abs);

- MDVP: RAP;
- MDVP: PPQ;
- Jitter: DDP—Several measures of variation in fundamental frequency;
- MDVP: Shimmer;
- MDVP: Shimmer(dB);
- Shimmer: APQ3;
- Shimmer: APQ5;
- MDVP: APQ;
- Shimmer: DDA—Several measures of variation in amplitude NHR;
- HNR—Two measures of ratio of noise to tonal components in the voice status:
 - Health status of the subject (one)-Parkinson's;
 - (zero)-healthy RPDE.
- D2—Two nonlinear dynamical complexity measures;
- DFA—Signal fractal scaling exponent spread1, spread2;
- PPE—Three nonlinear measures of fundamental frequency variation.

Collecting, integrating, analyzing these diverse clinical and behavioral data and attribute information can enable more accurate and early PD detection, leading to timely interventions and improved patient outcomes.

2.2. Ensemble Learning Technique (ELT)

In this context, we used the power of ensemble learning to facilitate the detection of PD. EL amalgamates diverse algorithms to yield superior predictive performance compared to utilizing individual algorithms in isolation [39]. Leveraging the advantages inherent to EL, our ultimate ensemble is determined following a series of meticulous experiments. The selection of the final ensemble is detailed in Table 3, which outlines the ensemble assessment. From Table 3, we discern that the performance enhancements are notable in the case of ensembles utilizing algorithms like adaptive boosting [40], random forest (RF) [41], support vector regressor (SVR) [29], long short-term memory (LSTM) [20], gated recurrent unit (GRU) [27], and Stacked LSTM [31]. Consequently, the ensemble chosen as our ultimate selection encompasses the adaptive boosting, RF, SVR, LSTM, GRU, and Stacked LSTM algorithms.

Select ML algorithms for Ensemble	Accuracy	Sensitivity	Precision	F1-Score
Adaptive boosting, RF	65.93%	67.23%	68.69%	69.23%
Adaptive boosting, RF, and KNN	75.223%	75.98%	76.23%	78.23%
Adaptive boosting, RF, KNN, and LSTM	80.12%	80.26%	80.32%	80.95%
Adaptive boosting, RF, Stacked RF, GRU, and KNN	85.12%	86.26%	86.32%	86.95%
Adaptive boosting, RF, Stacked RF, GRU, and Improved KNN	90.0123%	89.926%	89.932%	89.95%
Adaptive boosting, RF, Stacked KNN, Stacked LSTM, and SVR	90.12%	91.26%	91.32%	92.95%
Adaptive boosting, RF, SVR, LSTM, GRU, and Stacked LSTM	96.12%	97.26%	98.32%	98.05%

Table 3. Ensemble status.

3. System Architecture of Our Proposed PDD-ET Model

Figure 3 outlines the framework of the PDD-ET model. By observing the system diagram in Figure 3, we can understand the progression of early PD detection using our proposed PDD-ET model.



Figure 3. System architecture of the PDD-ET model. Here, I_i and H_i denote the input and hidden layers of the proposed PDD-ET model.

4. Proposed PDD-ET Model

The impact of Parkinson's disease detection using ML ensemble techniques and a Customized Big Dataset (PDD-ET) can be profound and far-reaching, contributing to both the medical field and society. Our work lies in its potential to transform Parkinson's disease diagnosis and treatment. By leveraging advanced ML techniques and a customized big dataset, we can contribute to more accurate diagnosis, improved patient care, and the advancement of medical research. Our work can change lives, enhance medical practices, and inspire further innovation in neurology and AI in healthcare. Therefore, this section expounds upon our conceptualized *Parkinson's Disease Detection Ensemble Technique* by meticulously elucidating the model's overarching operational methodology, as visually represented in Figure 3. The comprehensive workflow of our model for early-stage Parkinson's disease detection (PDD) through the amalgamation of machine learning ensemble techniques and the utilization of premotor stage characteristics unfolds across a sequence of distinct phases.

- Initially, the customized big PD dataset is subjected to normalization using the zeroscore normalization technique. Subsequently, a pre-processing step is implemented to eliminate any absent or inaccurate values from the training dataset before initiating the training of the proposed Parkinson's Disease Detection Ensemble Technique (PDD-ET) model.
- After that, we divided the customized big dataset into training, testing, and validation datasets (80%:10%:10%) and fed them into the PDD-ET model.
- With this setup, we start the training procedure of the PDD-ET model.
- Finally, we evaluate the results through comparison with the predicted and observed outcomes.

4.1. Construction of the PDD-ET Model

The focal objective of this research is to establish a connection between PD and PDD, aiming to achieve the early identification of PD. Our approach involves the construction of the PDD-ET model, which is founded on an ensemble of diverse algorithms, including adaptive boosting, RF, SVR, LSTM, GRU, and Stacked LSTM, illustrated in Figure 4. Commencing with individual training for each regressor module, we attain a total of six distinct models (referred to as Model₁, Model₂, Model₃, Model₄, Model₅, and Model₆). These models are subsequently amalgamated via adjustments in weight parameters, culminating in the creation of the requisite PDD-ET model.

4.2. Training of the PDD-ET Model

Before commencing the training process, we partition the training dataset into six separate sub-training datasets. Each sub-training dataset is then utilized to train six individual models, encompassing adaptive boosting, random forest, SVR, LSTM, GRU, and stacked LSTM algorithms.

Training		Trai Sam (14	ning ple ₁ %)	Training Sample ₂ (14%)	Training Sample ₃ (13%)	Training Sample ₄ (13%)	Training Sample ₅ (13%)	Training Sample ₆ (13%)	
(80%)		Ada boo	ptive sting	Random Forest	SVR	LSTM	GRU	Stacked LSTM	
					Training	Procedure			
	Model ₁ Model ₂ Model ₃ Model ₄ Model ₅ Model ₆								
Weight Adjustment									
Ensembling & Combining									
↓									
				En	semble PD	D-ET Mode	1		

Figure 4. Construction of PDD-ET model.

The training procedure yields optimal outcomes by adhering to the configured hyperparameter setup, as outlined in Table 4.

Table 4. Best hyperparameter configuration.

Hyperparameter	Value		
Loss	MSE and MAE		
Optimizer	stochastic gradient descent (SGD)		
Batch size	68		
Time step	1		
Epochs	1000		
Learning Rate	0.0003		
Dropout	0.03		

4.3. Deployment of the PDD-ET Model

The construction of the PDD-ET model involves the integration of individual algorithms, including adaptive boosting, RF, SVR, LSTM, GRU, and stacked LSTM. Each algorithm features five hidden layers, comprising fifty neurons each, designed to accommodate the inputs from the meticulously curated customized extensive PD dataset. The ensemble algorithms employed are detailed as follows:

- 1. Adaptive Boosting algorithm [40]:
 - Adaptive boosting algorithm is an ensemble technique used to improve weak models' performance.
 - Here, we designed three adaptive boosting algorithms with different models. The first boosting algorithm deals with the classification tree. The second and third boosting algorithms are concerned with different linear models.
- 2. Random Forest algorithm [41]:
 - Random forest (RF) also belongs to the class of ML ensemble techniques. It aggregates the results in a cluster form. RF proceeds using the bootstrap mechanism and de-correlates the classification trees via random splits during training.
- 3. SVR algorithm [29]:
 - SVR is an offshoot of support vector machines (SVM), originating from its principles. Employed to transform features from a lower-dimensional realm to a higher-dimensional one through the utilization of kernel functions, SVR constructs a hyperplane that maximizes optimization outcomes.
- 4. LSTM algorithm [26]:

- LSTM is a variant of recurrent neural network (RNN) having the memorization capacity with its cell units [42]. It works efficiently by using its complex input, forget, and output gates at each layer of the hierarchy. The output of every layer is treated as the input of its next immediate layer during training.
- 5. GRU algorithm [33]:
 - GRU is a modified version of the LSTM neural network.
 - It improves the complexity and gate architecture of standard LSTM by using update and reset gates.
- 6. Stacked LSTM algorithm [31]:
 - Stacked LSTM (SLSTM) is a special kind of ensemble technique to enhance the performance of a standard LSTM neural network. Inside this SLSTM neural network, the memory states (i.e., cells) are reset at each layer to obtain an accurate result at every step.
 - Stacking architecture makes the model deeper and reaches for improved performance.
 - The SLSTM neural network also reduces the accumulation and propagation errors for long-term detection.
 - It also minimizes the computational complexity of the iterative strategy for long-term multi-step prognosis.

5. Result Analysis and Discussion

5.1. Implementation Details

To assess the efficacy of the PDD-ET model's performance, we gauge the accuracy of both the individual ensemble models and the comparative models during the training phase. The training procedure involves segmenting the customized extensive PD dataset into proportions of 80% for training, 10% for testing, and 10% for validation. During both the training and testing stages, we uphold the consistency of the proportion of PD-affected and non-affected individuals within the dataset, achieved through the implementation of stratified sampling. Our ensemble models are trained using the training data to discern whether patients are afflicted by PD or not in the testing data. To ensure computational efficiency, this entire process is iterated 500 times.

To comprehensively depict the robustness and precision of our proposed PDD-ET model compared to other models, we uniformly train all compared models with the same configuration. The models subjected to comparison include (i) SVR [29], (ii) CNN [30], (iii) Stacked-LSTM [31], (iv) LSTM [26], (v) GRU [33], (vi) Alex Net [34], (vii) DT+RF+SVR [32], (viii) Deep Neural Network [35], (ix) HOG [36], (x) Quantum ReLU Activator [37], (xi) Improved KNN [43], (xii) Adaptive Boosting [40], (xiii) RF [41], and (xiv) Deep Learning [44] Models.

5.2. Experimental Setup

The experiments were conducted on a server housing an Intel i7-8700K CPU and an NVIDIA GeForce GTX 1080 GPU. The software environment employed was Python 3.7, facilitated by mini Anaconda.

5.3. Model Evaluation

PDD-ET and all compared models are evaluated based on the following metrics: (i) accuracy (acc), (ii) sensitivity (senv), (iii) specificity (spec), (iv) precision (prec), and (v) F1-Score (F1). These evaluation metrics are computed from [45].

Algorithm 1 PDD-ET: Parkinson's Disease Detection using ML Ensemble Techniques and Customized Big Dataset

- 1: INPUT: Customized Big Dataset.
- 2: **OUTPUT:** Parkinson's disease detection.
- 3: Training of PDD-ET Model
- 4: Begin
- 5: Initialize raw input data for every model.
- 6: Apply zero-score normalization followed by pre-processing.
- 7: Training of Adaptive Boosting algorithm:
- 8: Assign equal weight for every dataset (5%, 5%, and 4%) of classification tree and linear models.
- 9: Identify the miss classified samples.
- 10: Increase the weight parameter for the miss classified samples.
- 11: Ensemble three models.
- 12: Obtained Model₁.
- 13: if obtained desired results then
- 14: {
- 15: Ensemble three models.
- 16: Obtained Model₁.
- 17:
- 18: else
- 19:
- 20: Identify the miss classified samples.
- 21: Increase the weight parameter for the miss classified samples.
- 22: Ensemble three models.
- 23: Obtained Model₁.
- 24:
- 25: Training of Random Forest algorithm:
- 26: Select random samples through a bagging classifier from the training set (14% of training data).
- 27: Generate decision trees for every training data.
- 28: Voting process will be considered for averaging the decision trees.
- 29: Select the most voted result as the final desired result in terms of Model₂.
- 30: Training of SVR algorithm:
- 31: Set the hyperplane function, kernel, and boundary lines for SVR.
- 32: Perform feature extraction.
- 33: Perform fitting operation to generate the output in terms of Model₃.
- 34: Training of LSTM & GRU algorithms:
- 35: Construct an instance of the sequential classes.
- 36: Create 50 layers and connect each of them in the required sequence.
- 37: Compile the neural networks.
- 38: Perform fitting operation to generate the desired result in terms of Model₄, and Model₅.
- 39: Training of Stacked LSTM algorithm:
- 40: Create a stack and perform push() operation for every LSTMs.
- 41: Follow the training procedure of the LSTM neural network to get the desired result in terms of Model₆.
- 42: Add the desired results from every model by adjusting weight parameters.
- 43: Formed Ensemble PDD-ET Model.
- 44: Evaluate the ensemble model on the validation set using *acc*, *senv*, *spec*, *prec*, *F*1, *Loss*, and *AUC* metrics.
- 45: Early Detection of PD based on Algorithm 2.
- 46: Interpretability and Deployment.
- 47: Continual Monitoring and Improvement.
- 48: End

Algorithm 2 Early PD Detection Criteria.

- 1: INPUT: Clinical observations, diagnostic test results, patient data.
- 2: **OUTPUT:** Early PD detection criteria.
- 3: Begin
- 4: Initialize criteria list: $criteriaList \leftarrow \emptyset$
- 5: Clinical Symptoms and Signs:
- 6: *criteriaList* ← *criteriaList* ∪ Specific motor and non-motor symptoms associated with PD.
- 7: Diagnostic Tests:
- 8: *criteriaList* \leftarrow *criteriaList* \cup Imaging test results (MRI, DAT scans).
- 9: *criteriaList* \leftarrow *criteriaList* \cup UPDRS scores and assessments.
- 10: Biomarkers:
- 11: $criteriaList \leftarrow criteriaList \cup$ Identified biomarkers in bodily fluids.
- 12: **Response to Treatment:**
- 13: $criteriaList \leftarrow criteriaList \cup$ Positive response to dopaminergic medications.
- 14: Longitudinal Monitoring:
- 15: *criteriaList* \leftarrow *criteriaList* \cup Gradual worsening of symptoms and test results over time.
- 16: Machine Learning Algorithms:
- 17: $criteriaList \leftarrow criteriaList \cup$ Machine learning predictions based on diverse data.
- 18: Combining Criteria:
- 19: Define comprehensive criteria by combining elements from *criteriaList*.
- 20: End

In addition to the aforementioned five assessment metrics, we incorporated the AUC curve. Accuracy is a measure of correctly identifying individuals with Parkinson's disease. A higher accuracy value indicates the superior performance of the PDD overall. The *senv* metric quantifies the PDD-ET model's proficiency in detecting individuals with PD. On the other hand, the *spec* metric gauges the PDD-ET model's capability to accurately identify individuals without the condition. The *prec* metric establishes the significance of positive detections pertaining to PD-affected patients. The *F*1 score represents the harmonic mean between *prec* and *senv*.

5.4. Performance Analysis

As depicted in Figure 5, we can discern the PDD-ED model's capability to detect individuals affected by PD by utilizing PD features. Within Figure 5, our proposed PDD-ET model's feature importance score, indicated by the F1-Score, is presented.



Figure 5. PD detection of the proposed PDD-ET model.

Figure 6 illustrates the loss trends of the proposed PDD-ET model across both the training and testing datasets. Upon observation of Figure 6, it becomes evident that the training and testing losses experience swift declines during the initial epochs. Following the first 20 epochs, both plots stabilize, indicating that the training loss and testing loss have reached a comparable equilibrium. This pattern signifies a consistent trend between the training and testing datasets, with the loss maintaining a relatively stable pattern over a span of 50 epochs.





Figure 7 illustrates the distribution of Parkinson's disease within the proposed PDD-ET model, employing the density feature.



Figure 7. Detection of healthy and PD-affected patients based on PD density of PDD-ET model.

Figure 8 illustrates the detection of PD using the proposed PDD-ET model, employing the spiral images from the healthy and PD-affected patients.

 Healthy
 Parkinsons
 Parkinsons
 Healthy
 Healthy
 Healthy

 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy
 Healthy

Figure 8. Detection of healthy and PD-affected patients based on spiral images of PDD-ET model.

Figure 9 delineates the representation of PD status within the proposed PDD-ET model.



PD status with different mdvp_fhi_hz group

Figure 9. Status of PD detection of the PDD-ET model.

Figure 10 delineates the representation of PD spread status within the proposed PDD-ET model based on all used features of PD.



Figure 10. Spread status of Parkinson's disease using the PDD-ET model.

5.5. PD Detection of PDD-ET Model

Figure 11 presents the AUC value associated with the PDD-ET model. Upon examination of Figure 11, it becomes apparent that enhanced receiver operating characteristics are evident within the testing dataset. This substantiates the heightened PD detection proficiency achieved by the proposed PDD-ET model.



Figure 11. AUC curve of the proposed PDD-ET model.

Table 5 displays the distribution of PD status as depicted by the proposed PDD-ET model.

	117,541 Units of Samples			
PD Status	PD Positive	PD Negative		
Training Set	25%	10.6%		
Validation Set	5%	6%		
Testing Set	45.4%	8%		
Total Set	75.4%	24.6%		

Table 5. Distribution of PD status of the PDD-ET model.

5.6. Comparing the PD Detection with Other ML/DL Models

Table 6 demonstrates the efficacy of our proposed PDD-ET model compared to other models, showcasing the minimal generated loss (i.e., 19.325%). Upon examination of Table 6, it becomes evident that our proposed PDD-ET model showcases both efficiency and robustness compared to the state-of-the-art models. Moreover, we can observe that the PDD-ET model aligns with the advancements attributed to ensemble techniques. The fusion of models significantly enhances the neural network's performance compared to each ML and DL algorithm.

Model	acc (%)	senv (%)	spec (%)	prec (%)	F1 (%)	Loss (%)	AUC (%)
SVR [29]	58.236	59.986	50.236	50.6921	55.236	40.23	78.23
CNN [30]	65.329	66.159	60.956	65.2369	65.659	30.26	85.655
Stacked-LSTM [31]	75.415	76.652	75.236	74.489	75.658	20.365	80.369
LSTM [26]	68.968	68.265	68.266	67.856	68.236	30.256	80.235
GRU [33]	70.968	70.265	70.256	71.256	71.658	26.23	82.365
Alex Net [34]	77.235	76.235	76.556	76.698	76.569	30.123	83.569
Decision Tree, RF, SVR [32]	76.652	76.231	76.359	76.589	76.658	26.256	82.698
Deep Neural Network [35]	68.568	68.236	68.356	68. 432	68.569	32.123	84.236
HOG [36]	70.236	70.569	70.165	70.3215	70.213	30.215	81.023
Quantum ReLU Activator [37]	72.123	72.369	72.456	72.325	72.658	25.369	78.256
Improved KNN [43]	50.9658	50.3256	50.4568	50.231	50.562	45.236	75.9869
Adaptive Boosting [40]	55.658	55.956	55.7858	55.9831	55.7562	40.136	78.169
RF [41]	53.1258	53.366	53.556	53.111	53.1262	32.236	79.1169
Deep Learning Model [44]	79.918	79.561	79.116	79.569	79.662	30.106	79.969
Proposed PDD-ET model	95.325	95.265	95.955	95.1225	95.925	19.325	88.00

Table 6. Performance comparison among all state-of-the-art models.

5.7. Impact and Application of the PDD-ET Model

The impact and application of our proposed PDD-ET model can be defined as follows:

- 1. Early and Accurate Diagnosis:
 - ML ensemble techniques can lead to more accurate and dependable detection of Parkinson's disease, including its initial phases when symptoms might be subtle.
 - Early diagnosis enables timely intervention and treatment, potentially slowing the advancement of the condition and enhancing the quality of life for patients.
- 2. Personalized Treatment:
 - Accurate diagnosis allows for customized treatment strategies tailored to the unique condition of each individual patient.
 - Healthcare professionals can prescribe targeted therapies and medications, reducing unnecessary side effects and optimizing treatment outcomes.
- 3. Reduced Misdiagnosis:
 - ML ensemble techniques can significantly decrease the rate of misdiagnosis, which is expected due to the complexity of Parkinson's symptoms.
 - This process reduces patient frustration and the risk of inappropriate treatments.
- 4. Improved Monitoring and Progression Tracking:
 - ML-based ensemble models can monitor patients' symptoms and disease progression continuously.
 - This process enables medical professionals to make informed adjustments to treatment plans as needed.
- 5. Advancing Medical Research:
 - Our designed customized big dataset contributes to a more comprehensive understanding of Parkinson's disease.
 - The dataset can be valuable for researchers investigating the disease's genetic, environmental, and clinical factors.
- 6. Facilitating Research Collaboration:
 - Sharing our customized big dataset and methodologies can foster collaboration among researchers, enabling them to advance the field of Parkinson's disease research collectively.
- 7. Enhancing Medical Expertise:
 - ML models can complement the expertise of medical professionals, providing them with an additional tool for accurate diagnosis.

- Medical professionals can focus more on patient care and treatment decisions.
- 8. Data-Driven Insights:
 - Analyzing our customized big dataset using ensemble techniques can reveal insights into the disease that might not be apparent through traditional methods.
 - These insights could lead to new hypotheses and avenues of research.
- 9. Public Health Impact:
 - Improved diagnosis and treatment can positively impact public health by reducing the burden of Parkinson's disease on healthcare systems and improving patients' overall well-being.
- 10. *Ethical Considerations:*
 - Our work raises awareness about the ethical considerations of using AI in healthcare, encouraging discussions on data privacy, patient consent, and responsible AI deployment.
- 5.8. Discussion

This section contains a two-way discussion of our proposed PDD-ET model.

Clinical Application: Integrating ML ensemble techniques and a customized big dataset for Parkinson's Disease (PD) detection holds significant potential for clinical application and patient care. Here are some potential clinical applications:

- 1. *Early Detection and Diagnosis:* The developed model can contribute to the early identification of Parkinson's disease, allowing for timely intervention and treatment planning. This has the potential to enhance patient results and quality of life.
- 2. *Personalized Treatment:* The PDD-ET model's accurate detection can lead to increased customization of treatment strategies uniquely adapted to the individual patient's condition and needs.
- 3. *Monitoring Disease Progression:* The PDD-ET model can monitor disease progression over time, assisting clinicians in adjusting treatment strategies as needed.
- 4. *Assisting Medical Professionals:* Clinicians can use the PDD-ET model's predictions as an additional diagnostic tool, helping them make more informed decisions in conjunction with their expertise.
- 5. *Telemedicine and Remote Monitoring:* The PDD-ET model can be integrated into telemedicine platforms, enabling remote monitoring of patients' PD status and providing healthcare professionals with valuable insights for remote consultations.
- 6. *Clinical Trials and Research:* The PDD-ET model can contribute to the recruitment and stratification of randomized participants, leading to more accurate research outcomes.

Future Directions: Looking ahead, there are several directions in which the application of ML ensemble techniques and customized big dataset for PD detection can evolve:

- 1. *Improved Performance:* Further optimization and fine-tuning of ensemble models can enhance their accuracy and reliability in detecting early-stage PD.
- 2. *Incorporating Multi-Modal Data:* Integration of multiple data sources, such as imaging, genetic, and wearable sensor data, can provide a more comprehensive understanding of PD and improve detection accuracy.
- 3. *Longitudinal Monitoring:* Developing models that analyze changes in patient data over time can aid in predicting disease progression and treatment responses.
- 4. *Explainable AI:* Enhancing the interpretability of the model's predictions can increase its clinical acceptance by providing insights into the features driving the detection.
- 5. *Real-Time Monitoring:* Creating models that can operate in real time can enable continuous monitoring of PD symptoms, allowing for rapid adjustments to treatment plans.
- 6. *Global Deployment:* Scaling the model's deployment across different healthcare systems and regions can ensure a broader impact on PD detection and patient care.

- 7. *Collaboration with Clinicians:* Continued collaboration with medical professionals is essential for refining the model's clinical relevance, validation, and integration into clinical workflows.
- 8. *Ethical Considerations:* Addressing ethical concerns related to patient data privacy, model bias, and responsible AI usage is critical for the sustainable application of these techniques in clinical settings.
- 9. *Patient Empowerment:* Developing patient-friendly tools that leverage the model's predictions can empower individuals to monitor their PD status actively.
- 10. Expanding the model's capabilities to detect other neurological disorders can make it a versatile tool in clinical neurology.

6. Conclusions

The early identification of PD holds significant importance in gaining insights into its underlying causes. Through early PD detection, individuals afflicted with PD can initiate therapy and treatments at the nascent stages of the condition. This paper introduces an ensemble-based PDD-ET model, leveraging premotor features of PD to distinguish between individuals in good health and those afflicted by PD. The proposed PDD-ET model showcases an enhanced capability for PD detection compared to other models, achieving a remarkable accuracy level of 95.325%. This achievement can be predominantly attributed to the amalgamation of diverse ML and DL models. The experimental outcomes unequivocally establish the superiority of the PDD-ET model over the 14 compared ML and DL models.

7. Future Work

In the future, our domain of PDD using ML-based ELT could encompass various avenues for exploration and enhancement, such as *Dynamic Ensemble Adjustment*, *Imbalanced Data Handling*, and *Cross-Dataset Generalization*. Incorporating these aspects into upcoming research and development endeavors has the potential to drive advancements within the realm of Parkinson's disease identification, facilitated by ensemble learning. Ultimately, these advancements could improve early detection, treatment strategies, and patient outcomes. We also face a few challenges when the PDD-ET model comes to practical implementation in a clinical setting. Bridging the gap between advanced AI models and their practical clinical application is an important consideration. Here are some steps that can be taken to address this issue in the future:

- 1. *Interdisciplinary Collaboration:* Engage in collaborative efforts between data scientists, machine learning experts, and clinical professionals. This collaboration can help ensure that the model's development aligns with clinical realities and requirements.
- Simplified Reporting: Create an intermediate layer that translates the model's complex predictions into more understandable and actionable insights for clinicians. This layer could provide explanations for the model's decisions and present the results in a format that clinicians are familiar with.
- 3. *User-Friendly Interface:* Develop a user-friendly interface that simplifies the interaction with the model. This could involve a dashboard or application that presents the model's output in an easily interpretable manner.
- 4. *Clinical Guidelines:* Develop guidelines for clinicians on how to interpret the model's results and incorporate them into their decision-making process. This could include recommendations on when and how to use the model's predictions alongside traditional diagnostic methods.
- 5. *Training for Clinicians:* Provide training sessions for clinicians to understand the underlying concepts of the model and its practical application. This can help them gain confidence in utilizing the model effectively.
- 6. *Gradual Implementation:* Introduce the model in a phased manner, starting with specific use cases where the model's predictions can provide valuable insights. Gradually expand its usage as clinicians become more comfortable with its application.

- 7. *Feedback Loop:* Establish a feedback loop where clinicians can provide input on the model's performance, usability, and areas for improvement. This iterative process can lead to a model that better aligns with clinical needs.
- 8. *Real-World Testing:* Conduct pilot studies or simulations within a controlled clinical environment to observe how the model's predictions integrate into the workflow and impact decision making.
- 9. *Validation Studies:* Conduct validation studies that compare the model's predictions with established diagnostic methods. This can help establish the model's clinical validity and reliability.
- 10. *Ethical and Regulatory Considerations:* Ensure compliance with ethical guidelines and regulatory requirements for medical devices and AI applications in healthcare.
- 11. *Patient Involvement:* Involve patients in the implementation process. Their feedback can provide insights into the practical implications of using the model in clinical care.

By taking these steps, the transition from a conceptual model to practical implementation in a PD clinic can become more feasible. The aim is to ensure that the model's advanced capabilities are harnessed to enhance clinical decision making and patient care effectively.

Author Contributions: K.C. and R.P.K.: writing manuscript, developing software, experiment design, performing the experiments. A.B., S.S. and S.M.: initial idea, writing manuscript, revision of the manuscript, experiment design, data analysis. A.L. and K.R.: resource management, data analysis and revision of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Prashanth, R.; Roy, S.D. Early detection of Parkinson's disease through patient questionnaire and predictive modelling. *Int. J. Med. Informat.* **2018**, *119*, 75–87. [CrossRef] [PubMed]
- Singh, N.; Pillay, V.; Choonara, Y.E. Advances in the treatment of Parkinson's disease. *Prog. Neurobiol.* 2007, 81, 29–44. [CrossRef] [PubMed]
- Gunduz, H. Deep learning-based Parkinson's disease classification using vocal feature sets. *IEEE Access* 2018, 7, 115540–115551. [CrossRef]
- Tsanas, A.; Little, M.A.; McSharry, P.E.; Ramig, L.O. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans. Biomed. Eng.* 2010, 57, 884–893. [CrossRef] [PubMed]
- Lahmiri, S.; Shmuel, A. Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. Biomed. Signal Process. Control 2019, 49, 427–433. [CrossRef]
- 6. Braga, D.; Madureira, A.M.; Coelho, L.; Ajith, R. Automatic detection of Parkinson's disease based on acoustic analysis of speech. *Eng. Appl. Artif. Intell.* **2019**, *77*, 148–158. [CrossRef]
- Zhao, A.; Qi, L.; Li, J.; Dong, J.; Yu, H. A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data. *Neurocomputing* 2018, 315, 1–8. [CrossRef]
- 8. Bilgin, S. The impact of feature extraction for the classification of amyotrophic lateral sclerosis among neurodegenerative diseases and healthy subjects. *Biomed. Signal Process. Control* **2017**, *31*, 288–294. [CrossRef]
- 9. Silveira-Moriyama, L.; Petrie, A.; Williams, D.R.; Evans, A.; Katzenschlager, R.; Barbosa, E.R.; Lees, A.J. The use of a color coded probability scale to interpret smell tests in suspected parkinsonism. *Mov. Disord.* **2009**, *24*, 1144–1153. [CrossRef]
- Valenza, G.; Orsolini, S.; Diciotti, S.; Citi, L.; Scilingo, E.P.; Guerrisi, M.; Danti, S.; Lucetti, C.; Tessa, C.; Barbieri, R.; Toschi, N. Assessment of spontaneous cardiovascular oscillations in Parkinson's disease. *Biomed. Signal Process. Control* 2016, 26, 80–89. [CrossRef]
- 11. Illner, V.; Sovka, P.; Rusz, J. Validation of freely available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease. *Biomed. Signal Process. Control* **2020**, *58*, 101831. [CrossRef]
- 12. Solana-Lavalle, G.; Galán-Hernández, J.-C.; Rosas-Romero, R. Automatic Parkinson disease detection at early stages as a prediagnosis tool by using classifiers and a small set of vocal features. *Biocybern. Biomed. Eng.* **2016**, *40*, 505–516. [CrossRef]

- 13. El Maachi, I.; Bilodeau, G.-A.; Bouachir, W. Deep 1D-convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Syst. Appl.* **2020**, *143*, 113075. [CrossRef]
- 14. Gupta, U.; Bansal, H.; Joshi, D. An improved sex-specific and age dependent classification model for Parkinson's diagnosis using hand writing measurement. *Comput. Methods Programs Biomed.* **2020**, *189*, 105305 [CrossRef]
- Wagner, A.; Fixler, N.; Resheff, Y.S. A wavelet-based approach to monitoring Parkinson's disease symptoms. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 5980–5984.
- Arroyo-Gallego, T.; Ledesma-Carbayo, M.J.; Sanchez-Ferro, A.; Butterworth, I.; Mendoza, C.S.; Matarazzo, M.; Montero, P.; Lopez-Blanco, R.; Puertas-Martin, V.; Trincado, R.; et al. Detection of motor impairment in Parkinson's disease via mobile touchscreen typing. *IEEE Trans. Biomed. Eng.* 2017, 64, 1994–2002. [CrossRef]
- 17. Dinov, I.D.; Heavner, B.; Tang, M.; Glusman, G.; Chard, K.; Darcy, M.; Madduri, R.; Pa, J.; Spino, C.; Kesselman, C.; et al. Predictive big data analytics: A paper of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. *PLoS ONE* **2016**, *11*, e0157077 [CrossRef]
- 18. Drotár, P.; Mekyska, J.; Rektorová, I.; Masarová, L.; Smékal, Z.; Faundez-Zanuy, M. Decision support framework for Parkinson's disease based on novel handwriting markers. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2015**, *23*, 508–516. [CrossRef]
- 19. Drotár, P.; Mekyska, J.; Rektorová, I.; Masarová, L.; Smékal, Z.; Faundez-Zanuy, M. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artif. Intell. Med.* **2016**, *67*, 39–46. [CrossRef]
- 20. Harrou, F.; Sun, Y.; Hering, A.S.; Madakyaru, M. Statistical Process Monitoring Using Advanced Data-Driven and Deep Learning Approaches: Theory and Practical Applications; Elsevier: Amsterdam, The Netherlands 2020.
- Xing, W.; Bei, Y. Medical Health Big Data Classification Based on KNN Classification Algorithm. *IEEE Access* 2020, 8, 28808–28819. [CrossRef]
- Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020, 577, 706–710. [CrossRef]
- Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* 2019, 115, 213–237. [CrossRef]
- Little, M.A.; McSharry, P.E.; Hunter, E.J.; Spielman, J.; Ramig, L.O. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Biomed. Eng.* 2009, 56, 1015–1022. [CrossRef] [PubMed]
- Das, R. A comparison of multiple classification methods for diagnosis of parkinson disease. *Expert Syst. Appl.* 2010, 37, 1568–1572.
 [CrossRef]
- Ashour, A.S.; El-Attar, A.; Dey, N.; El-Kader, H.A.; El-Naby, M.M.A. Long short term memory based patient-dependent model for FOG detection in Parkinson's disease. *Pattern Recognit. Lett.* 2020, 131, 23–29. [CrossRef]
- Govindu, A.; Palwe, S. Early detection of Parkinson's disease using machine learning. *Procedia Computer Science* 2023, 218, 249–261. [CrossRef]
- 28. Prashanth, R.; Roy, S.D.; Mandal, P.K.; Ghosh, S. High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. *Int. J. Med. Informat.* **2016**, *90*, 13–21. [CrossRef]
- Roh, S.-B.; Oh, S.-K.; Pedrycz, W.; Fu, Z. Dynamically Generated Hierarchical Neural Networks Designed With the Aid of Multiple Support Vector Regressors and PNN Architecture with Probabilistic Selection. *IEEE Trans. Neural Netw. Learn. Syst.* 2022, 33, 1385–1399. [CrossRef]
- Khatamino, P.; Canturk, I.; Ozyilmaz, L. A deep learning-CNN based system for medical diagnosis: An application on Parkinson's disease handwriting drawings. In Proceedings of the 2018 6th International Conference on Control Engineering & Information Technology (CEIT), Istanbul, Turkey, 25–27 October 2018; pp. 1–6.
- 31. Chakraborty, R.; Hasija, Y. Predicting MicroRNA Sequence Using CNN and LSTM Stacked in Seq2Seq Architecture. *IEEE/Acm Trans. Comput. Biol. Bioinform.* 2020, 17, 2183–2188. [CrossRef]
- Parziale, A.; Della, C.A.; Senatore, R.; Marcelli, A. A decision tree for automatic diagnosis of Parkinson's disease from offline drawing samples: Experiments and findings. In Proceedings of the Image Analysis and Processing, Trento, Italy, 9–13 September 2019; Springer: Cham, Switzerland, 2019; pp. 196–206.
- Moetesum, M.; Siddiqi, I.; Javed, F.; Masroor, U. Dynamic Handwriting Analysis for Parkinson's Disease Identification using C-BiGRU Model. In Proceedings of the 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR), Dortmund, Germany, 8–10 September 2020; pp. 115–120.
- 34. Nõmm, S.; Zarembo, S.; Medijainen, K.; Taba, P.; Toomela, A. Deep CNN based classification of the archimedes spiral drawing tests to support diagnostics of the Parkinson's disease. *IFAC-PapersOnLine* **2020**, *53*, 260–264. [CrossRef]
- Johri, A.; Tripathi, A. Parkinson disease detection using deep neural networks. In Proceedings of the 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, 8–10 August 2019; pp. 1–4.
- Folador, J.P.; Santos, M.C.S.; Luiz, L.M.D.; Souza, L.A.P.D.; Vieira, M.F.; Pereira, A.A.; Andrade, A.D.O. On the use of histograms of oriented gradients for tremor detection from sinusoidal and spiral handwritten drawings of people with Parkinson's disease. *Med. Biol. Eng. Comput.* 2021, 59, 195–214. [CrossRef]
- Parisi, L.; Neagu, D.; Ma, R.; Campean, F. Quantum ReLU activation for convolutional neural networks to improve diagnosis of Parkinson's disease and COVID-19. *Exp. Syst. Appl.* 2022, 187, 115892. [CrossRef]

- Available online: https://www.kaggle.com/datasets/debasisdotcom/parkinson-disease-detection (accessed on 15 January 2023).
- 39. Available online: https://en.wikipedia.org/wiki/Ensemble_learning (accessed on 15 January 2023).
- 40. Gu, X.; Angelov, P.P. Multiclass Fuzzily Weighted Adaptive-Boosting-Based Self-Organizing Fuzzy Inference Ensemble Systems for Classification. *IEEE Trans. Fuzzy Syst.* 2022, *30*, 3722–3735. [CrossRef]
- 41. Tie, J.; Lei, X.; Pan, Y. Metabolite-disease association prediction algorithm combining DeepWalk and random forest. *Tsinghua Sci. Technol.* **2022**, *27*, 58–67. [CrossRef]
- Jun, K.; Lee, D.-W.; Lee, K.; Lee, S.; Kim, M.S. Feature Extraction Using an RNN Autoencoder for Skeleton-Based Abnormal Gait Recognition. *IEEE Access*2020, *8*, 19196–19207. [CrossRef]
- Fang, Z. Improved KNN algorithm with information entropy for the diagnosis of Parkinson's disease. In Proceedings of the 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE), Guilin, China, 25–27 February 2022; pp. 98–101.
- 44. Wang, W.; Lee, J.; Harrou, F.; Sun, Y. Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning. *IEEE Access* 2020, *8*, 147635–147646. [CrossRef]
- 45. Available online: https://chat.openai.com/ (accessed on 25 August 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.