Grenze International Journal of Engineering and Technology, June Issue

Early Detection of Parkinson's Disease using Voice and Motion Data

Amruta R. Sutar[1] and Nikita Shetty[2]
[1]Government College of Engineering, Karad, India
Email: amrutasutar8421@gmail.com
[2]Professor, Dept. of Information Technology/Government College of Engineering, Karad, India
Email: nikita.shetty@gcekarad.ac.in

***Abstract*— Parkinson's disease (PD) is a degenerative disorder of the nervous system that quite often results in speech, movement, and fine motor functions getting impaired and the gradual decline of an individual's quality of life. The importance of early detection of PD cannot be overemphasized as it helps to facilitate the right time medical intervention and slow down the disease progression. This research introduces a non, invasive, inexpensive, and easy, to, understand early detection system that uses data obtained from voice, motion, and handwriting to figure out neurological problems that the patient might not have identified but which are noticeable in the early stages of the disease. Parkinson's disease, related speech abnormality may be lessened pitch change, monotonous speech, and vocal tremor whereas motion problems are accompanied with tremors, slow movements, irregular gait, and loss of motor coordination. Analysis of handwriting additionally unveils micrographia and irregular stroke dynamics that are associated with fine motor dysfunction. This paper focuses on the extraction of acoustic features, for example, Mel, Frequency Cepstral Coefficients (MFCCs), as well as motion, related temporal features, and handwriting trajectory features for further analysis. These deep learning architectures, for example, Convolutional Neural Networks (CNNs) as well as Long Short, Term Memory (LSTM) networks, are used to understand the spatial and temporal patterns present in the multimodal data. The ultimate goal of the proposed system is to be able to provide a reliable early screening tool as well as to help the healthcare professionals in the early diagnosis of Parkinson's disease.**

***Index Terms*— Parkinson's Disease, Early Detection, Voice Analysis, Motion Data, Machine Learning, Deep Learning.**

I. INTRODUCTION

Parkinson disease (PD) is a progressive, chronic, neurodegenerative disorder, which leads to the breakdown in the motor system and gradually reduces the capacity of the person to do daily activities. It is the second prevalent neurodegenerative disease in the whole world and is likely to increase with the growing life expectancy. Diagnosis of the initial stages of Parkinson is usually delicate and could hardly be detected because some signs like a slight trembling, a slight alteration of the speech or a slight alteration of the body could be ignored or caused by old age. However, an early diagnosis is essential due to the possibility to slow the development of the disease and lower the level of symptoms, as well as increase the overall state of the life of the patients.

Grenze ID: 01.GIJET.12.2.70

The conventional methods of diagnosing depend greatly on the clinical assessments done by neurologists that is subjective and time consuming and cannot detect the disease in its initial phases. Over the past few years, sensor technology and artificial intelligence led to the establishment of non-invasive, data-driven methods to early detection of Parkinson. These include voice and motion data, which have proved to be very effective biomarkers. The muscles of the voice in Parkinsonism are frequently impaired resulting in vocal cues of a decreased range of variation of pitch range, low-pitched voice, monotonous speech, and lack of regular rhythm. In the same way, motor symptoms (vibrations, slowness, and stiffness, and gait imbalances) can be observed using motion sensors and wearable devices recording finer movements and formatting them.

Machine learning and deep learning applications are important in addressing these complex physiological indicators and determining patterns that can be used to diagnose early signs of Parkinson. Computational models are able to distinguish early-stage PD and healthy conditions with a high level of accuracy by extracting acoustic features of speech and kinematic features of motion data. This multimodal methodology can be used to treat the fact that it improves the reliability of detection and it is also a non-invasive, cost-effective diagnostic method that can be used in a continuous search. The proposed research will build upon the concept that an integrated system will be developed to process voice and motion data to diagnose Parkinson disease at the early stages of the development with the help of advanced feature extraction and classification algorithms. The suggested system can serve the purpose of an early screen, remote care, and enhanced clinical decision-making, which, in the end, will lead to improved patient outcomes and better disease management.

II. Literature Review

The research of Shyamala and Navamani (2025) proposes a speech-based severity stage classifier of Parkinson's disease (PD), based on deep-learning that is capable of identifying Healthy, Mild, Moderate and Severe stage, thereby sealing this gap in the recent research, which is mainly binary detection of PD. They primarily introduce the Parkinson Speech Stage-Net (PSS-Net) an optimized pipeline that consists of Isolation Forest as an outlier removal classifier, KMeansSMOTE as a class balance classifier, PCA-based optimized dimensionality reduction, Random Forest feature selection and an enhanced deep neural network (OptiFusionNet) with batch normalization and dropout, that reaches an 89 percent accuracy, which is higher than that of classical ML and baseline DL classifiers. SHAP-based explainability is also brought to the model to recognize specialist vocal biomarkers like jitter and shimmer, MDVP parameters, DFA, and TQWT attributes to make clinical interpretability stronger. Limitations, however, are that it uses single UCI dataset, may overfit because of synthetic methods of oversampling, missing real-life noisy speech data validation and is not multimodal (UPDRS, gait, imaging) in clinical use [1].

Dimauro et al. (2017) introduce a trial that assesses the speech intelligibility of Parkinson diseases (PD) with an objective, automated method, that is, a Speech-to-Text (STT) module, which is implemented in their software called Voxtester, and it is intended to replace a subjective test, namely the UPDRS rating of speech. Work adds to order a speech-recording protocol, phonically balanced text and word analysis and systematic error rates of recognition differences among young healthy control, elderly control and PD patients such that Google STT misrecognition patterns increase steadily with the severity of the disease and thus can be used as an indirect indicator of speech impairment. Although the research does not provide any single classifier performance, it shows clearly defined patterns, including the increased instances of recognition errors in the patients of PD, the high level of correlation of systematic STT error rates across groups, and the significant distinction between the patients of PD and healthy speakers. Weaknesses are reliance on Google STT, systematic word-related mistakes, small sample size, reliance on Italian-only phonetic sources, and lack of independent STT system and clinical validation of early PD detection, which implies that there should be a more extensive set of data and multilingual testing in the future [2].

The article by Gunduz et al. (2019) discusses a deep-learning approach to the classification of Parkinson's disease (PD) relying on speech features in the form of sets of vocal features, where the two convolutional neural network (CNN) frameworks are proposed, namely, feature-level and model-level combinations, to learn the discriminative representations based on a combination of multiple speech-based features. The research introduces a new application of parallel layers of convolution to obtain deep features of various feature groups (TQWT, MFCC, Wavelet and baseline vocal features) showing that model-level combination can perform better when three groups of features are used. The highest result had an accuracy of 0.869, F-measure 0.917 and MCC of 0.632, which was better than the earlier models of SVMs and the previous research on the same dataset. Others are reliance on one dataset, low number of voices samples per subject and less externalisability as all the

data was recorded into a controlled clinical environment; furthermore, the model was not trained on real-life or noisy speech samples [3].

The core aspect of the present study by Shivangi et al. (2019) is the ability to combine two complimentary modules in an attempt to detect early signs of Parkinson disease (PD) the CNN-based VGFR Spectrogram Detector, which operates on gait signal (i.e. vertical ground force reaction), and the ANN-based Voice Impairment Classifier, which relies on vocal biomarkers (speech).The authors state that both the CNN-based VGFR Spectrogram Detector and the ANN-based Voice Impairment Classifier The CNN module has an accuracy of 88.17 per cent and the voice-based ANN classifier has an accuracy of 89.15 per cent meaning that both gait and voice feature are very discriminative. The research, however, has some limitations such as comparatively small and imbalanced data, possible overfitting through the small validation, absence of cross-dataset generalization testing, and limited coverage of the symptoms, as other PD indicators (e.g., olfactory loss, handwriting) are not included in the final model [4].

To eliminate redundancy and noise in the 754-feature UCI PD vocal dataset, Xiong and Lu (2020) offer a Parkinson physical disease (PD) classification model based on combinational Adaptive Gray Wolf Optimization (AGWO) in the selection of the most discriminative vocal features with a sparse autoencoder, and generate deep latent representations of speech data (high-dimensional). Their achievement consists of an integrated pipeline, which reduces a set of 754 features to 37 feature predictors of interest, compressor features with a sparse auto-encoder and using Linear Discriminant Analysis (LDA) to evaluate various supervised machine-learners, with the LDA performing best, having 95 percent correct classification, outperforming CFS, mRMR, RFE, and baseline full-feature models. The experiment has shown that bio-inspired feature selection combined with deep unsupervised representation learning has a strong positive effect on the accuracy of PD classification. The shortcomings of these methods, however, are that they use a single dataset that has limited variety, and that they are not tested with noisy or multilingual speech in the real world, that the scale of the used sample may be overfit, and that they are not compared to more modern deep learning models such as CNNs or RNN-based acoustic models [5].

Wang et al. (2020) present an early-stage Parkinson's disease (PD) detection framework that integrates deep learning with machine-learning methods using premotor biomarkers such as REM sleep behavior disorder (RBDSQ), olfactory loss (UPSIT), cerebrospinal fluid (CSF) markers, and dopaminergic SPECT imaging features, based on the PPMI dataset of 401 early-stage PD patients and 183 healthy controls.Their major contribution is the development of a feed-forward deep neural network (two hidden layers) and a comprehensive comparison against twelve ML models, demonstrating that their ensemble deep learning model (DEEP EN) achieves the highest performance with 96.45\% accuracy, outperforming boosting, SVM, RF, logistic regression, and discriminant analysis. They also provide feature-importance analysis showing that putamen and caudate SBR imaging markers and UPSIT scores are the strongest predictors for early PD detection. However, limitations include reliance on a relatively small and imbalanced dataset, limited interpretability of the deep model, no inclusion of longitudinal progression data, and constraints of using clinical biomarkers rather than more easily collected signals like speech or gait, which may affect scalability and generalizability [6].

When Parkinson disease (PD) patients are put under levodopa medication, Pah et al. (2021) examine three sustained phonemes, including /a/, /o/, and /m/ and train out phonatory data, including the jitter, shimmer, pitch, and harmonics to noise ratio of segmented voice samples. The primary value of the contribution is the demonstration of statistically significant increase in the levodopa effect primarily on the nasal phoneme /m/ and more so on the time and amplitude perturbation measures, although /o/ and /a/ also indicate significant differentiation between the PD and healthy control group. Through Support Vector Machines, (SVM) the paper also shows that by joining all the three phonemes, the maximum classification rate is achieved in that they obtained the highest AUC of 0.81 in differentiating between the PD-off and the PD-on, and 0.90 in distinguishing PD-off and controls, thus the multi-phoneme analysis is important in the detection of the two disorders, dysarthria and medication response. Limitations are the comparatively small sample size (24 PD, 22 controls), none of the two genders are modeled, none of the accents or linguistic diversity is explored, and there is no follow-up within the same subject so one cannot assess the level of intra subject consistency [7].

Quan, Ren, and Luo (2021) suggested a deep learning model to identify and detect Parkinson disease based on dynamically extracted speech features. The study is based on time-varying characteristics of speech rather than being confined to traditional and static acoustic features derived in sustained vowel phonation. The authors used the deep neural net architectures that discover discriminative speech-representation patterns automatically by the dynamics of speech representations and this showed better classification results than the traditional machine-learning techniques. The conducted experimental results on test voice sample parks of the clinical interests of the Parkinsonism disease proven that mixes of dynamic speech characteristics showed a stronger effect through deep

learning models and raised detection accuracy which indicates the potential of the temporal speech analysis to detect early causes of the Parkinsonism disease [8].

Arora et al. (2021) offer a speech assessment system on the smartphone that can analyze the symptoms of Parkinson diseases (PD) and REM Sleep Behaviour Disorder (RBD) as a proposal of an available digital biomarker method based on voice recordings that can be remotely prepared. The primary idea is to take the acoustic, prosodic, and nonlinear features of sustained phonation and speech performances recorded using smartphones, which allows one to characterize the symptoms early and without the use of clinical-grade devices. Their significant contribution is showing that low-cost and smartphone quality data of speech can effectively distinguish between PD, RBD, and healthy controls, where machine-learning models are strong in their discriminative performance which is reported with accuracy of up to 80-90 percent depending on feature set and task. Scalability of telemedicine applications and the possibility of real-time monitoring of symptoms (continuous, nonstop) are also mentioned in the study. Yet, some of its limitations are environmental noise variability, audio quality based on the device, small cohort size compared to the level of screening the population, and the possibility to withstand changes in the stages of the diseases and heterogeneous population to validate the findings [9].

Khedimi et al. (2025) suggested a single deep learning ensemble model to identify Parkinson's disease and predict motor severity through voice. The approach combines several deep learning models to obtain complementary voice characteristics to be able to accurately classify diseases and estimate motor severity. Obtained experimental outcomes show that ensemble method has a superior performance over single models, which proves that it is an effective non-invasive procedure to test Parkinsonism disease [10].

Govindu and Palwe (2023) introduce a machine-learning-based framework on the early diagnosis of Parkinson's disease (PD) based on non-invasive speech biomarkers to help identify the disease at an early stage, relying on MDVP vowel-phonation records, to facilitate the process of telemedicine. Their significant input is in assessing the three methodological strategies namely full 22-feature training, PCA-based dimensionality reduction into 5 key attributes, and class-balancing with the use of up-sampling and the four ML classifiers (SVM, Random Forest, KNN, Logistic Regression). It is found that the Random Forest classifier most accurately forecasts the dataset at 91.83 percent and 0.95 sensitivity whereas SVM correctly forecasts the dataset at 91.75 \% following PCA and KNN correctly forecasts the same at 91.83 percent. The authors utilize the significance of jitter, shimmer, MDVP properties, PPE, RPDE, and DFA to predict PD as well as show that telemedicine may utilize speech as a viable option in early screening. Limitations however are few as the dataset size is tiny and skewed (195 total recordings in total), demographic diversity is small, micro-generalization to the real-world of mobile audio, overfitting due to up-sampling and lack of external validation or deep-learn benchmarking, which lowers clinical generalization [11].

The study conducted by Goel et al. (2023) suggests an early PD diagnosis model, based on voice data in the UCI Parkinson dataset, prioritizing voice acoustic markers (jitter, shimmer, RAP, NHR, and HNR) to detect the patterns of dysphonia related to PD. The principal contribution is the creation of a Feedforward Neural Network (FNN) with numerous dense layers, dropout regularization, and Adam optimization, and preprocessing measures, such as correlation-based feature selection, inverting features in negative correlations, and minmax scaling. The model has a performance of 97.43 per cent accuracy which compares to XGBoost and beats SVM when it comes to the same dataset. The paper shows how deep neural networks can efficiently acquire discriminative vocal patterns that are based on rather limited datasets. Nevertheless, it is limited by small sample (31 subjects, 197 recordings), no evaluation of noisy speech or in real-time, possible overfitting of high-dimensional layers, a complete lack of cross-dataset validation, and use of only sustained vowel phonation, which makes it difficult to generalize to spontaneous speech and clinical use in the real world [12].

Murari et al. (2024) describe an artificial intelligence, based method for identifying Parkinsons disease through a small, optimized Convolutional Neural Network (CNN) that was trained on the sustained vowel phonations from the UCI Parkinsons Disease Voice Dataset. Their study highlights the usage of fundamental frequency (F0), jitter, and shimmer as the main acoustic biomarkers. The authors' contribution is a minimal CNN structure with feature, axis convolutions, Bayesian hyperparameter optimization, thorough feature, ablation analysis, and robustness checking with synthetic and noise, perturbed validation data, thus, indicating strong interpretability and reproducibility. The suggested model has an accuracy of 96.85 %, along with high precision, sensitivity, and F1, score, thus, it outperforms a number of traditional machine, learning baselines while still being efficient and clinically applicable. Nevertheless, the authors acknowledge that their work is limited by the small and acoustically controlled dataset, the absence of formant features, the lack of external validation, and the possibility of demographic bias, which together suggest that a larger multi, corpus study and more extensive speech tasks would be required for their next research [13].

Hou et al. (2025) propose an audio-based deep learning system for early Parkinson's disease (PD) detection, introducing a multimodal framework that fuses raw acoustic biomarkers—such as jitter, shimmer, MFCCs, F0, and spectral features—with temporal deep representations extracted using their end-to-end Time-CNN model, trained on 173 Mandarin-speaking participants (131 early PD, 42 healthy controls). Their main contribution lies in combining handcrafted clinical speech features with deep temporal–spectral features from log-Mel spectrograms, outperforming existing methods such as TLENet and TWIESN and demonstrating that hybrid feature fusion captures early dysarthric cues more effectively. The proposed Time-CNN achieved 0.78 accuracy and 0.831 F1-score, showing strong sensitivity to early-stage vocal impairments, although limitations include a relatively small and demographically narrow dataset, recordings collected only in controlled acoustic environments, potential device-specific biases, and limited external validation, which may affect generalizability to real-world or multilingual scenarios [14].

Lim et al. (2025) proposed a cross-language speech-based model for detecting Parkinson's disease using advanced neural network techniques. The core concept of the study is that speech impairments caused by Parkinson's disease appear consistently across different languages, allowing machine learning models to generalize better if trained on multilingual datasets. The authors contributed a novel cross-language framework that integrates acoustic features, prosodic variations, and deep speech embeddings to improve robustness across English, Mandarin, and Malay speakers. The experimental results demonstrated strong classification performance, achieving an accuracy of approximately 92–94 %, outperforming single-language models. Despite these promising results, the study has notable limitations, including a relatively small multilingual dataset, limited representation of diverse dialects, and potential bias due to uneven sample distribution among languages. They also acknowledge that real-world deployment may require additional adaptation to account for environmental noise, mobile recording variability, and individual voice differences [15].

A. Research Gaps

Existing speech-based Parkinson's disease detection works have a number of significant shortcomings, namely, they are all based on small and demographically limited subsets, have not been tested in real-world or noisy speech, or have overly focused on sustained phonation of the vowel, as opposed to natural connected speech. Majority of the available methods are mostly concerned with fixed acoustic characteristics and are not effective in capturing dynamic time series especially those that are associated with early stages of Parkinsonism where vocal collections are mild. In addition, cross-corpus validation and multilingual validation is very seldom used, which lessens the generalizability of the models. Despite the rising adoption of deep learning methods, most models are not explainable, making them unacceptable to clinicians, whereas the application of small datasets together with oversampling strategies tends to cause overfitting. Also, the combination of complementary modalities like gait, handwriting has not been studied to an extent and longitudinal or disease-progress analysis is not always done, which restricts the clinical usefulness and practical applicability of existing systems.

B. Problem Statement

Precise and prompt diagnosis of Parkinson's Disease is still a problem because of similarity with other diseases, absence of biomarkers, and subjective diagnosis. The current diagnostic systems cannot be scaled or exploited at a continuous level of monitoring. Automated, non-invasive, and multimodal (ML-based) strategies, which are able to early identify PD based on voice, motion, or handwriting data are needed.

III. PROPOSED SYSTEM

As shown in Fig. 1, the proposed system will seek to realize early detection of the Parkinson disease (PD) by using a multimodal deep-learning-based system that interventions voice, motion, and handwriting data. The technological process starts with information capture, i.e. speech recordings (e.g. sustained vowels, short spoken phrases, etc.), motion measurements and handwriting (e.g. spiral drawings). These inputs are first pre-processed, to filter noise, normalize, filter, and segment data to offer clean and similar information across every modality. Then the modality-specific features are obtained: the acoustic features (MFCCs, jitter, shimmer, fundamental frequency, and spectrograms of voice data), the kinematic features (tremor frequency, acceleration, gait variability of motion data), and the fine motor ones (stroke dynamics, pressure, and drawing irregularities of handwrite data). Deep spectro-temporal representations are then combined with both handcrafted features and used to predict disease patterns of complex temporal and spatial structures using deep learning models such as CNN, LSTM or hybrid CNN-LSTM. The trained model judges people as Parkinson's disease and healthy, with severity estimation possible. Optimization techniques that are used to improve robustness and generalization are

dropout, batch normalization and learning-rate learning. The resulting output is presented in the form of an explainable class or a risk score that is non-invasive, economically feasible, and clinically feasible to make efficient early screening of Parkinson in patients.

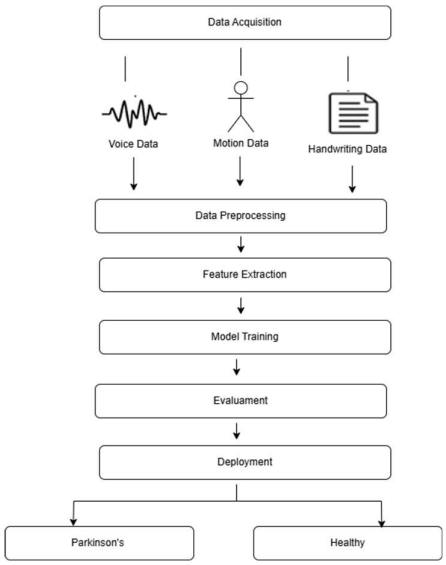

Figure 1. Proposed System Architecture

A. Dataset Description

The proposed system uses three publicly available datasets, which are voice, motion, and handwriting modalities to be used in Parkinson disease detection. To analyze the speech, the Parkinson's Disease Voice Dataset will be utilized and it can be found in the UCI machine learning repository and it includes 195 sustained vowel phonations of 31 subjects (23 with Parkinson and 8 healthy controls) as well as 22 clinically relevant acoustic features are used to analyze speech; jitter, shimmer, fundamental frequency, harmonics to noise ratio and nonlinear features which are used to detect the Parkinsonian dysarthria. The CARE-PD gait dataset consists of clinically recorded movement and gait data (derived on the basis of RGB videos and 3D pose representations) that is used to analyze motion, and the gait and postural abnormalities are considered as the motor symptoms. Assessment based on handwriting uses the Parkinson Disease Spiral Drawings Dataset of UCI, comprising of digitized spiral drawings, which measured fine motor impairments, tremor and lack of coordination. The combination of all these datasets allows thorough multimodal processing towards effective detection of Parkinson's disease.

IV. Mathematical Formulas and Algorithms Used

A. Notation and Dataset: Let the dataset be

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N} \quad (1)$$

where $x_i \in \mathbb{R}^d$is the d-dimensional feature vector extracted from the i-th voice recording and $y_i \in \{0,1\}$is the label defined as

$$y_i = \begin{cases} 1, & \textit{if subject } i \textit{ has Parkinson's disease (PD)} \\ 0, & \textit{if subject } i \textit{ is healthy (HC)} \end{cases} \quad (2)$$

B. Preprocessing and Feature Map

Let P(·)denote preprocessing (DC removal, denoising, framing, normalization).
Let $\phi(\cdot)$be the feature-extraction map producing MDVP features (jitter, shimmer, F0, HNR, RPDE, DFA, PPE, etc.).
The input feature vector is

$$\mathbf{x}_i = \phi(\mathcal{P}(\text{audio}_i)) \in \mathbb{R}^d \quad (3)$$

A standard normalization step (zero mean, unit variance) is

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j}, j = 1, \dots, d \quad (4)$$

where $\mu_j$and $\sigma_j$are the mean and standard deviation computed from the training set.

C. Classifier Model Let

$$f_{\boldsymbol{\theta}}: \mathbb{R}^d \to [0,1]$$

be the probabilistic classifier (e.g., neural network with sigmoid output) with parameters $\boldsymbol{\theta}$.
The predicted probability of Parkinson's disease for sample $i$is

$$\hat{p}_i = f_{\boldsymbol{\theta}}(\tilde{\mathbf{x}}_i) \quad (5)$$

D. Decision Rule (PD vs Healthy Detection)

Using threshold $\tau \in (0,1)$(typically $\tau = 0.5$):

$$\hat{y}_i = \begin{cases} 1(\text{Parkinson's}), & \text{if } \hat{p}_i \geq \tau \\ 0(\text{Healthy}), & \text{if } \hat{p}_i < \tau \end{cases} \quad (6)$$

This is the exact step where the system declares PD or Healthy.

E. Training Objective

The model is trained using regularized binary cross-entropy:

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{N}\sum_{i=1}^{N}[\, y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i)] + \lambda \mathcal{R}(\boldsymbol{\theta}) \quad (7)$$

where $\mathcal{R}(\boldsymbol{\theta})$is a regularizer (e.g., $\| \boldsymbol{\theta} \|_2^2$).

F. Evaluation Metrics

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\hat{y}_i = y_i) \quad (8)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (12)$$

G. Threshold Selection

$$\tau^\star = \arg\max_{\tau} (\text{Sensitivity}(\tau) + \text{Specificity}(\tau) - 1) \quad (13)$$

Novelty of research

The originality of the study consists in the fact that a unified framework of deep-learning-enabled Parkinson's disease (PD) treatment is developed, which incorporates voice, motion, and handwriting recognition strategies to obtain reliable and timely diagnosis. In comparison to the literature, where researchers can present only one modality or use purely the task-independent features of the handcrafted, the suggested system is integrated with the handcrafted biomarkers based on acoustic and spectro-temporal speech and on the motion dynamics and handwriting motor patterns. To do voice analysis, time-domain and frequency-domain features are obtained and converted to dynamic spectrograms, that are learned with a hybrid CNN-LSTM model to capture finer temporal abnormalities linked to Parkinsonian dysarthria. The motion and handwriting features also reveal tremors, bradykinesia, gait instabilities, and the fine motor deficits, and hence fully evaluate the motor symptoms. Moreover, the framework also includes effective pre-processing, noise-cleaning, feature balancing, and data

balancing mechanisms so as to handle dataset imbalance and over-fitting as the previously discussed work mostly does not consider these issues. The lightweight and multi-modal architecture suggested can be applied to the real-life scenario in the mobile and tele-medicine system and offer a cost-effective, non-invasive and clinically feasible solution of the early-screening approach. All in all, the innovation of the system is in the multimodal integration of voice, motion, and handwriting characteristics, the temporal modeling of the disease trends, and the possibility to deploy the new system in a real-world setting, as the system is closer to clinical use and more precise than the current single-modality ones.

V. CONCLUSIONS

Machine, learning and deep, learning approaches have been very promising to detect Parkinsons Disease early and in a non, invasive way. Recent developments indicate that hybrid architectures and attention, based deep models are consistently better than traditional methods as they can capture complex and subtle patterns in patient data. Besides, various modalities like voice recordings, handwriting dynamics, and wearable sensor signals provide different but complementary views of motor and speech impairments. Combining these multimodal features in a single diagnostic framework can not only increase the robustness, accuracy, and clinical reliability, but also pave the way for more efficient early screening and continuous monitoring of Parkinson's Disease.

REFERENCES

[1] K. Shyamala and T. M. Navamani, "Design of an Efficient Prediction Model for Early Parkinson's Disease Diagnosis," IEEE Access, 2024, doi:10.1109/ACCESS.2024.3421302.

[2] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi, "Assessment of Speech Intelligibility in Parkinson's Disease Using a Speech-To-Text System," IEEE Access, vol. 5, pp. 22199–22208, 2017, doi:10.1109/ACCESS.2017.2762475.

[3] H. Gündüz, "Deep Learning Based Parkinson's Disease Classification Using Vocal Feature Sets," IEEE Access, vol. 7, pp. 115540–115551, 2019, doi:10.1109/ACCESS.2019.2936564.

[4] Shivangi, A. Johri, and A. Tripathi, "Parkinson Disease Detection Using Deep Neural Networks," Proc. IEEE, 2019.

[5] Y. Xiong and Y. Lu, "Deep feature extraction from the vocal vectors using sparse autoencoders for Parkinson's classification," IEEE Access, vol. 8, pp. 27821–27830, 2020, doi:10.1109/ACCESS.2020.2968177.

[6] J. Wang et al., "Early-stage Parkinson's disease detection using machine learning and deep learning approaches," IEEE Access, vol. 8, pp. 147635–147646, 2020, doi:10.1109/ACCESS.2020.3016681.

[7] N. D. Pah, R. K. Tripathy, and R. B. Pachori, "Detecting effect of levodopa in Parkinson's patients using sustained phonemes," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 1783–1791, 2021, doi:10.1109/TNSRE.2021.3092673.

[8] M. Khedimi et al., "A unified deep learning ensemble framework for voice-based Parkinson's disease detection and motor severity prediction," Bioengineering, vol. 12, no. 7, p. 699, 2025, doi:10.3390/bioengineering12070699.

[9] A. Govindu and S. Palwe, "Early detection of Parkinson's disease using machine learning," Procedia Comput. Sci., vol. 218, pp. 249–261, 2023, doi:10.1016/j.procs.2023.01.007.

[10] M. Goel et al., "Early detection of Parkinson disease using voice data," Int. J. Recent Innov. Trends Comput. Commun., vol. 11, no. 11s, pp. 580–585, 2023.

[11] T. B. Murari et al., "An approach for Parkinson's disease detection based on artificial intelligence techniques," IEEE Access, vol. 11, 2024.

[12] W. Hou et al., "Audio-based neural network to classify patients with early Parkinson's disease," IEEE Access, vol. 13, pp. 154283–154294, 2025, doi:10.1109/ACCESS.2025.3604877.

[13] W. S. Lim et al., "A cross-language speech model for detection of Parkinson's disease," J. Neural Transmiss., vol. 132, pp. 579–590, 2025.

[14] H. A. Mohammadigilani et al., "Enhanced LSTM by attention mechanism for early detection of Parkinson's disease through voice signals," arXiv:2502.08672, 2025.

[15] S. Naimi, W. Bouachir, and G.-A. Bilodeau, "1D-Convolutional Transformer for Parkinson disease diagnosis from gait," arXiv:2311.03177, 2023.