Parkinson's Disease Classification using Speech under Data-Constrained Regime

Mayurakshi Mukherji*, Hrithik Mhatre†, Vamshika Sutar† Nirmal Punjabi*, Ganesh Ramakrishnan‡,
*Koita Center for Digital Health, Indian Institute of Technology Bombay, India
†Department of Civil Engineering, Indian Institute of Technology Bombay, India
‡Department of Computer Science and Engineering, Indian Institute of Technology Bombay, India

***Abstract*—Speech-based detection of Parkinson's disease (PD) provides a promising pathway for early and noninvasive diagnosis using everyday devices such as smartphones and tablets. Motor impairment associated with PD affects the muscles responsible for controlling the vocal folds and articulators, resulting in measurable changes in speech production. This work presents a complete pipeline for PD detection using two types of speech samples: sustained /aa/ phonation and reading speech, collected from a dataset with a limited number of speakers.**

Acoustic features were extracted using the ComParE 2016 feature set, which includes both functional and low-level descriptor (LLD) features. For functional features, the mRMR algorithm was applied to select an optimal subset of predictors for both tasks, with each audio sample segmented into 3s clips and evaluated using traditional machine learning classifiers. For LLD features, 60ms frames were extracted from the /aa/ phonation and reading speech and used to train deep learning models.

Experimental results indicate that the reading-speech task provides more discriminative acoustic cues for PD classification compared to the sustained phonation task. Traditional machine learning models achieve the highest performance on reading speech, while deep learning models perform best on /aa/ phonation. Additionally, classifiers trained on handcrafted acoustic features were evaluated as baselines; however, the proposed pipeline outperformed these approaches.

***Index Terms*—Biomedical signal processing, speech analysis, Parkinson's disease, machine learning, deep learning**

I. Introduction

Parkinson's Disease (PD), a progressive neuro-degenerative condition, affects over 8.5 million people worldwide according to a 2019 global estimate [1]. It affects the motor functions due to the gradual decline of dopamine-producing neurons in the substantia nigra. This in turn leads to disruptions in the functioning of the basal ganglia, which plays a major role in controlling movement. The characteristic symptoms of PD include tremors, slowness of movement, rigidity, impaired balance and coordination, and altered speech and voicing [2]–[4]. Speech changes often manifest as monotonic speech (reduced variation in fundamental frequency), hypophonia (low volume), hypokinetic dysarthria (poor articulation) and dysphonia (voice disorders) [5], [6]. Since PD cannot be diagnosed by blood work, early diagnosis is often overlooked as many of the symptoms are discernible only in the advanced stages [7], [8]. Subtle changes in speech patterns may appear even before changes in the movement of the limbs, hence aiding in an early diagnosis and treatment plan. Furthermore, speech and voice signals can also be a useful tool for continuous monitoring of PD patients using wearable devices.

Studies have shown that voice biomarkers can be effective for early detection and monitoring of PD. [9] reports that features derived from fundamental frequency and the second formant when used as inputs to the machine learning model to distinguish PD patients from healthy controls, show statistical significance even when symptoms were minimal. This suggests that voice analysis could aid in early PD detection, even before obvious symptoms appear. Another study by [2] found that individuals with PD have significantly lower fundamental frequency (F0) variability in speech compared to controls, regardless of medication status, although F0 variability increases when patients are ON medication. Additionally, this variability has no correlation to the duration of PD, thus also indicating its usefulness for early detection. When using jitter and shimmer related features [10] reports a 70% accuracy with both types of features in separating PD voices and healthy voices with an increase in feature values for men and a decrease in feature values for women . This is further supported by [11]; pitch dynamics differ notably between male and female speakers with age and PD progression.

Although speech signals show promise as a diagnostic tool for PD, despite various past studies in this domain, it has not yet been clinically adopted due to concerns surrounding generalization across patients, speech types and mode of speech collection and explainability of speech features.

II. Related Work

In recent times, many works have been published on PD speech classification targeting various aspects of the pipeline such as feature extraction, feature selection, model selection and data augmentation.

OpenSMILE toolkit is popularly used to extract various audio feature sets such as AVEC, Emobase, IS-series features, ComParE 2016, etc. [12] evaluates the performance of 17 feature sets extracted using OpenSMILE and other toolkits. 88% accuracy was achieved for PD speech classification using the Emobase feature set and a Gaussian Naive Bayes model. Feature selection within each feature set was not explored. [13] achieves the highest intelligibility level assessment accuracy using a multiclass-SVM with a combination of selected glottal features and OpenSMILE features (6552 features before selection). It was trained and evaluated on the Universal Access

(UA) Speech database of dysrthric speech samples. Features selection was performed using Sequential Forward Feature Selection which adds each feature iteratively to a feature subset and checks the performance using a classifier. Although effective, this is an extremely computationally heavy and time-intensive process. Similarly, for UPDRS score estimation in a study by [14], a linear regression model with L2 regularization yielded a mean absolute error of 5.9 over a dataset of 168 PD and 21 HC. It was trained using the INTERSPEECH 2010 paralinguistic challenge feature set consisting of total 1582 features and extracted using the OpenSMILE toolkit. No feature selection was performed.

In a PD vs. HC speech classification task, [15] used a dataset of 754 extracted features from different types of speech samples of 188 PD patients and 64 HC and achieved an accuracy of 85.09% with an AdaBoost model. Their study involved feature selection using a LinearSVC with L1 penalty and augmentation using SMOTE. Yet another study by [16] employing a cascaded L1-SVM and DNN approach claims an accuracy of 97.5% with 10-fold CV using Praat derived feature set. Although effective, feature selection methods using classifier performance rely heavily on validation sets and may not be generalizable across different datasets. Whereas, feature selection methods such as mRMR rank the features at the pre-processing level. On using 1200 Audio-Visual Emotion recognition Challenge (AVEC) features selected and ranked using mRMR, gradient boosted decision tree achieved an overall accuracy of 86% [17]. It was trained and tested on the mPower sustained /aa/ phonation dataset of 4739 HC and 1087 PD patients. A 10-fold stratified CV strategy was used to evaluate ML models on various lengths of the feature set. [18] compared different feature selection methods and showed the best performance using the RELIEF feature selection method combined with an SVM classifier. They selected the 10 best features out of 132 dysphonia features and arrived at 99% classification accuracy over an imbalanced dataset of 33 PD and 10 HC by performing 10 CV 100 times.

III. Methodology

A. Dataset

In our work we have used the publicly available Italian PD Dataset by [19] which has voice samples from 28 PD patients and 22 elderly HC. Each speaker has 16 voice samples, 2 recordings of each vowel “a”, “e”, “i”, “o”, “u” in a sustained fashion, 2 recordings of the reading of a phonetically balanced passage, 1 recording of phonetically balanced phrases and another of phonetically balanced words, and repeated syllables “pa”, and “ta”.

As the recordings were collected in Bari, Italy, the phonetically balanced passage, phrases, and words are in Italian. The dataset was recorded in a quiet, echo-free room and a distance of 15 to 25 cm from the microphone was maintained. The sampling frequency of the available recordings is either 16 KHz or 44.1 KHz. None of the HC reported any speech or language disorders. The PD patients, 19 men and 10 women were aged 40-80 years and the elderly HC, 10 men and 12 women, were aged 60-77 years. The dataset also has voice samples of 15 young HC but they were not considered since the content of the recordings did not match with the PD patients.

For our analysis of PD vs. HC speech we only consider the sustained phonation of the vowel “a”, which we refer to as the /aa/ phonation and the phonetically balanced passages, which we refer to as the reading passage. In past PD speech studies, it has been widely acknowledged that sustained /aa/ phonation is effective due to it’s universal nature coupled with distinguishable vocal source characteristic. The 3-second /aa/ phonation requires minimal articulator movement and is therefore useful for examining features associated with vocal fold motion. We have also considered the reading passage to take into account features that are affected due to the altered movement of articulators and as the reading passage is of a longer duration and consists of various words, it captures the changes in the speech pattern in a more natural setting.

The recordings in the dataset were available at two different sampling rates, 16 kHz and 44.1 kHz. To maintain uniformity throughout the training process, all samples were down-sampled to 16 kHz with anti-aliasing. Files that were predominantly noise were manually excluded to ensure high-quality data.

B. Features Extraction and Selection

1) Feature Extraction: In all our experiments we use the ComPaRE_2016 feature set extracted using the OpenSMILE toolkit. It was introduced by [20] in the INTERSPEECH 2016 Computational Paralinguistics Challenge (ComParE). The main purpose of this feature set is to provide a comprehensive acoustic representation using a standardized fixed-length feature vector for computational paralinguistic tasks involving speaker state, trait, and affect-analysis. Some of the challenge tasks included snoring detection, cold detection, eating condition classification, native language identification, sleepiness and likability. It has also been used for disease classification such as voice-based Covid-19 detection, PD classification etc. [21], [22]. Although ComPaRE_2016 has been used for PD classification previously, we find that extensive discussion and evaluation for different types of feature selections, sound types, and models is missing in the literature.

The extracted feature set comprises 65 Low-Level Descriptors (LLDs), including both base descriptors and their first-order temporal derivatives (deltas). LLDs are time-series features computed at regular short-time intervals throughout the entire audio recording, capturing frame-level variations in acoustic properties such as prosody (e.g., loudness, fundamental frequency (F0), voicing probability), spectral characteristics (e.g., MFCCs 1–14, spectral flux, spectral energy, spectral roll-off), and voice quality (e.g., jitter, shimmer, harmonics-to-noise ratio, formant frequencies). These LLDs were extracted with a frame length of 60 ms and a frame shift of 10 ms, the default parameters of the OpenSMILE library. To obtain a fixed-dimensional representation for each audio sample, a set of statistical functionals is computed over the full duration of the LLD time series. These functionals include operations such as mean, standard deviation, percentiles, min, max, range,

skewness, and kurtosis, resulting in 6,373 functional features per recording.

As a baseline for performance evaluation, we extracted standard handcrafted acoustic features, widely used in prior studies namely [23], [24] and [25], to serve as benchmarks for comparison with the OpenSMILE ComParE feature set. These features were derived in two forms: point features, serving as a benchmark for OpenSMILE functional features, and time-series features, serving as a benchmark for OpenSMILE low-level descriptors (LLDs). In both cases, the same set of acoustic parameters was obtained, including 13 Mel-frequency cepstral coefficients (MFCCs, excluding the 0th coefficient), fundamental frequency (F0: mean, minimum, maximum) estimated within a 75–300 Hz range, jitter and shimmer (computed with pitch-period thresholds of 0.0001–0.02 s and amplitude factor limits of 1.3–1.6), short-term energy (mean squared magnitude of the Hilbert-transformed signal), zero-crossing rate (ZCR), and the first two formants (F1, F2) estimated using Burg's method with a maximum formant frequency of 5500 Hz and up to five formats. Point features were computed as statistical aggregates (mean) over the entire audio segment, yielding a single value per feature. Time-series features retained the frame-wise trajectories of the same parameters, using a frame length of 60 ms and a frame shift of 10 ms. For time-series features, all sequences were standardized to a fixed length by zero-padding or truncation to ensure consistent temporal resolution across samples.

2) Feature Selection: Given the high dimensionality of the functional feature set (6,373 features) relative to the dataset size, there exists a substantial risk of overfitting. To address this, feature selection techniques were employed. In contrast, the deep learning classifiers utilized the 65 low-level descriptor (LLD) features, which provide a more compact representation and are therefore less prone to overfitting. As a result, feature selection was not applied for these classifiers.

We employed the mRMR-MIQ algorithm (Minimum Redundancy Maximum Relevance with Mutual Information Quotient) for feature selection for ML Classifiers. This method optimally selects features by simultaneously maximizing their relevance to the target variable while minimizing redundancy among selected features. The feature importance is quantified through mutual information, which measures the statistical dependence between any two features A and B:

$$I(A;B) = \sum_{a\in A}\sum_{b\in B} p(a,b)\log\left(\frac{p(a,b)}{p(a)p(b)}\right)$$

The algorithm constructs the optimal feature set S through an iterative process that maximizes the Mutual Information Quotient:

$$\text{MIQ}(X_i) = \frac{I(X_i;Y)}{\sqrt{\frac{1}{|S|}\sum_{X_j\in S} I(X_i;X_j)}}$$

Here, the numerator $I(X_i;Y)$ represents the relevance of feature X_i to the target Y, while the denominator measures its average redundancy with respect to currently selected features in S. This balanced criterion ensures the selection of features that are both highly predictive and minimally correlated with each other.

To identify the optimal number of mRMR-selected features for our task, we conducted experiments across various Machine Learning classifiers using different feature set sizes. To prevent information leakage and ensure a fair evaluation, mRMR feature extraction was carried out solely on the data from the training–validation splits, while the held-out test set was left completely untouched.

C. Model Training

For model training and evaluation, the dataset was divided into training, validation, and test sets. The training–validation portion followed a 5-fold cross-validation scheme, where each fold contained recordings from three speakers with two audio files per speaker, resulting in six files per fold. The test set included recordings from four speakers with two audio files each, totaling eight files. The same five folds and fixed test set were used across all experiments to ensure comparability. Hyperparameter tuning was performed on the training–validation sets using a grid search strategy to optimize model performance. The reported results are averaged across all five folds.

1) ML classifiers: The machine learning classifiers used in this study include Random Forest (RF), XGBoost (XGB), Support Vector Machine (SVM), and Logistic Regression (LR). These classifiers were trained and tested on functional features extracted from brief 3-second /aa/ phonation and reading speech sample, enabling evaluation with short-duration recordings. We present the results of feature selection using mRMR-MIQ. Standardization was first applied to the selected features, with the scaling parameters computed solely from the training data in each fold. The resulting scalers were then used to transform both the corresponding validation fold and the held-out test set, thereby preventing information leakage. To identify the best hyperparameters, we employed grid search, monitoring the binary cross-entropy loss for both training and validation sets across all hyperparameter combinations. The optimal configuration was selected when the validation loss was approximately equal to the training loss, with both being as small as possible. Separate classifier models were trained on the full set of 6,373 functional features, the optimal feature subsets identified through feature selection, and the handcrafted features, for both /aa/ phonation and reading passage tasks across all four classifier types.

2) DL models: To capture the temporal and spectral dynamics inherent within the LLDs of speech signals associated with PD, we explore deep learning models such as Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Convolutional LSTM (ConvLSTM), and a 1D Convolutional Neural Network followed by LSTM (CNN-LSTM). These models are chosen for their ability to model sequential data and learn both local and global temporal dependencies.

Both /aa/ phonation and reading speech audio files were divided into frames of 60 ms with a 10 ms overlap. Each frame was represented by 65 LLD features. Only complete segments were retained to maintain uniform input

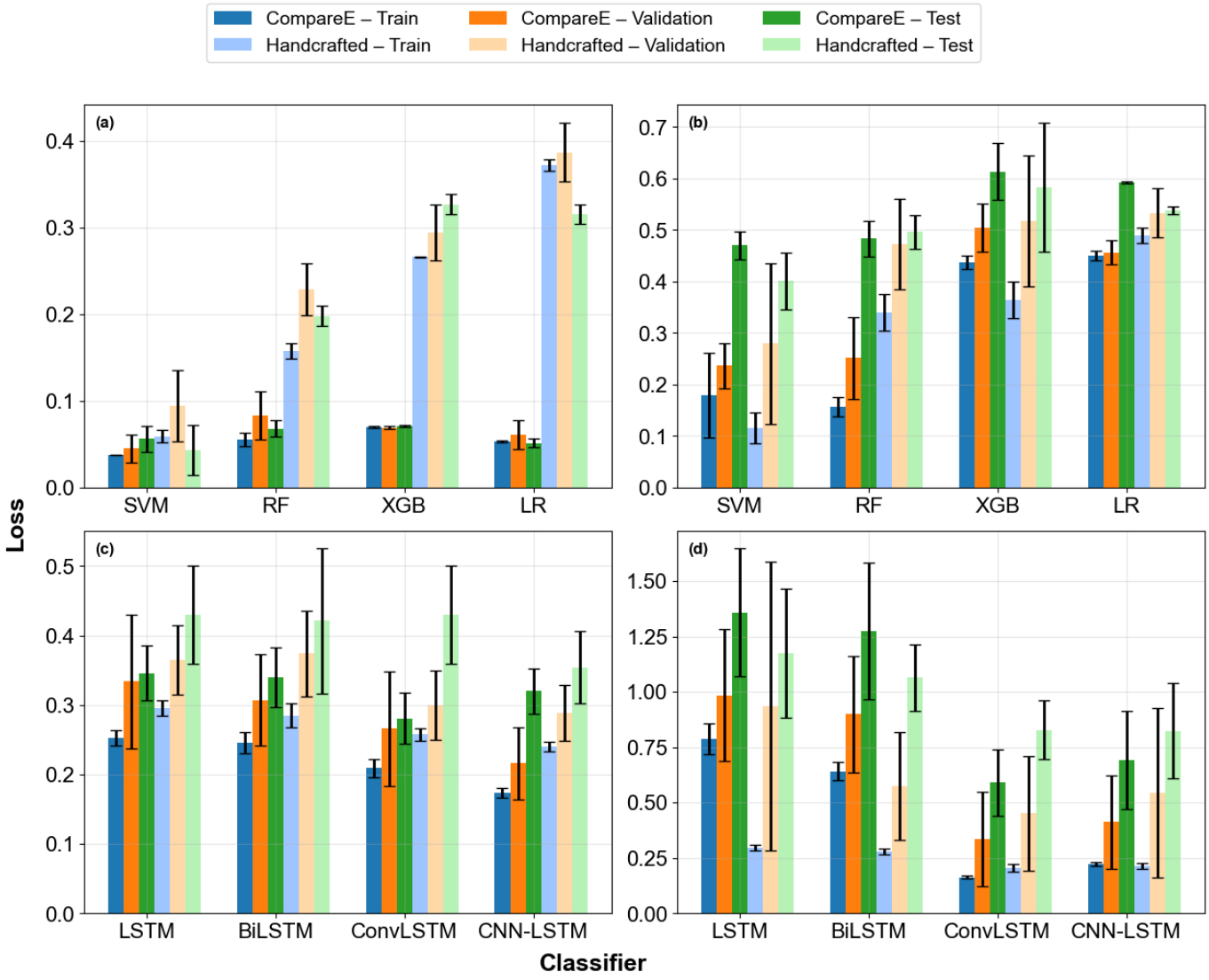

Fig. 1. Comparison of training, validation, and test losses for: (a) mRMR-selected 30 ComParE features combined with 22 handcrafted features evaluated on ML classifiers for the reading-speech task; (b) mRMR-selected 90 ComParE features combined with 22 handcrafted features evaluated on ML classifiers for the /aa/ phonation task; (c) 65 ComParE LLD features combined with 22 handcrafted LLD features evaluated on DL classifiers for the reading-speech task; and (d) 65 ComParE LLD features combined with 22 handcrafted LLD features evaluated on DL classifiers for the /aa/ phonation task.

dimensions. To ensure an unbiased learning process, the dataset was balanced such that both the HC and PD classes contributed an equal number of samples. Prior to training, all features were normalized to the $[0, 1]$ range using Min-Max scaling, with parameters estimated from the training data in each fold. These scalers were subsequently applied to the corresponding validation and held-out test sets to avoid information leakage. Separate deep learning classifier models were trained using the ComParE 65 LLD features and the standard handcrafted features for /aa/ phonation and reading passage task across all four classifier types. The four architectures considered in this study are detailed in the following sections.

LSTM: The Long Short-Term Memory (LSTM) network is capable of modeling long-range temporal dependencies within sequential data. Given input audio features in the form of Low-Level Descriptors (LLDs) extracted at regular time intervals (frame length = 60 ms, frame shift = 10 ms), we treat the entire audio as a time series $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$, where each $\mathbf{x}_t \in \mathbb{R}^{65}$ represents the 65-dimensional LLD vector at time t. The LSTM maintains a hidden state $\mathbf{h}_t$ and a memory cell $\mathbf{c}_t$, updated at each time step using:

$$
\begin{aligned}
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
\tilde{c}_t &= \text{LeakyReLU}(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}
$$

Here, σ denotes the sigmoid activation, and $\odot$ indicates element-wise multiplication. The forget gate f_t controls the

proportion of the previous memory c_{t-1} to retain, while the input gate i_t determines how much new candidate memory $\tilde{c}_t$ should be added. The output gate o_t decides how much of the updated cell state is exposed to the next layer or time step.

The candidate memory $\tilde{c}_t$ uses a LeakyReLU activation with negative slope α, which improves gradient flow compared to the traditional $\tanh$. The final memory c_t and hidden state h_t are updated using gating mechanisms that combine past and current information.

The matrices $W_* \in \mathbb{R}^{d_h \times d_x}$ apply linear transformations to the input x_t, while $U_* \in \mathbb{R}^{d_h \times d_h}$ apply linear transformations to the previous hidden state h_{t-1}. Both W and U are learned during training and are unique for each gate.

The final hidden state h_T is passed to a dense output layer with an activation function (e.g., softmax for classification). A dropout layer is used after the LSTM to reduce overfitting. The model is trained using the Adam optimizer with learning rate η, and Categorical Crossentropy loss with label smoothing ϵ to improve calibration of predicted class probabilities.

Bidirectional LSTM: To incorporate both past and future temporal context, we use Bidirectional LSTMs (BiLSTMs). The input LLD sequence is processed simultaneously in forward and reverse directions. At each time step t, the hidden states from the forward and backward passes, $\overrightarrow{\mathbf{h}}_t$ and $\overleftarrow{\mathbf{h}}_t$, are concatenated:

$$\mathbf{h}_t^{\text{bi}} = [\overrightarrow{\mathbf{h}}_t; \overleftarrow{\mathbf{h}}_t].$$

This enriched temporal representation is passed through a LeakyReLU activation followed by a dense output layer. BiLSTMs are especially effective in capturing anticipatory and contextual cues from speech, which are indicative of motor impairments in Parkinson's Disease.

CNN-LSTM: The CNN-LSTM model leverages convolutional layers to capture short-term local dependencies in the temporal Low-Level Descriptor (LLD) signal before passing the output to an LSTM layer for sequential modeling. The input sequence is reshaped into a 3D tensor of shape $(\text{nseq}, \text{tper\_seq}, 65)$, where nseq is the number of segments and tper_seq is the number of time steps per segment. A 1D convolution is then applied along the temporal axis:

$$z_t = \text{ReLU}(\text{Conv1D}(x_{t:t+k-1}))$$

Here, $x_{t:t+k-1}$ represents a sliding window of size k over the input sequence starting at time step t, and Conv1D refers to a one-dimensional convolutional filter that learns local patterns in the temporal context. The activation function used is ReLU (Rectified Linear Unit), applied element-wise to introduce non-linearity into the model. This operation is typically followed by max pooling to downsample the feature maps and flattening to convert the 3D tensor into a 2D format suitable for the LSTM. The resulting high-level feature sequence is then fed into the LSTM layer to capture long-term dependencies. Finally, the output of the LSTM is passed through a dense (fully connected) layer with a task-specific activation function. This hybrid architecture effectively combines local temporal feature extraction with global sequence modeling.

ConvLSTM: ConvLSTM is designed to learn both spatial and temporal correlations simultaneously, typically used when the input has an additional spatial structure. Here, the 1D LLD sequence is reshaped into a pseudo-2D format of shape $(n_{\text{seq}}, 1, t_{\text{per\_seq}}, 65)$, enabling the use of 2D convolutions over time and feature dimensions. The hidden state $\mathbf{H}_t$ and cell state $\mathbf{C}_t$ in ConvLSTM are updated as:

$$\begin{aligned}
F_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + b_f) \\
I_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + b_i) \\
\tilde{C}_t &= \text{LeakyReLU}(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\
C_t &= F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \\
O_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + b_o) \\
H_t &= O_t \odot \text{LeakyReLU}(C_t)
\end{aligned}$$

where $*$ denotes convolution. In ConvLSTM, X_t denotes the current input, and H_{t-1} is the previous hidden state. The learnable weight matrices W and U are applied to X_t and H_{t-1}, respectively, via convolution operations, and b terms are the corresponding biases. The forget gate F_t determines how much of the previous cell state C_{t-1} is retained. The input gate I_t, along with the candidate cell state $\tilde{C}_t$, modulated through a LeakyReLU activation (as specified in the implementation), introduces new information. The updated cell state C_t integrates both retained and incoming content. The output gate O_t controls the information passed from C_t to the hidden state H_t, which also uses LeakyReLU. These operations collectively capture spatiotemporal patterns using convolutional transformations at each step. The final ConvLSTM output is flattened and passed through a dense layer for classification. This model is effective in capturing fine-grained spatiotemporal patterns in LLD sequences.

IV. Results and Discussion

A. Features Selection

1) Selecting the ideal number of mRMR Features: For the reading speech task, as shown in Table I, increasing the number of mRMR-selected ComParE features from 15 to 30 enabled all models to achieve perfect metrics, with low training and validation losses. Performance saturated when the number of features was increased beyond 30, up to 45. Therefore, 30 features were considered optimal for this task.

For the phonation task, Table II shows that increasing the number of mRMR-selected features from 15 to 30 and then 45 resulted in noticeable instability across the feature sets. While the 30-feature subset consistently produced the highest metrics, performance fluctuated at 15 and 45 features. Training and validation losses also did not follow a monotonic trend, indicating that convergence was not achieved until 45 features.

Extending the analysis further, Table III shows that model performance begins to stabilize as the number of features increases to 90. Minimal variation is observed between 90

TABLE I
COMPARISON OF 15, 30, AND 45 MRMR-SELECTED COMPARE FUNCTIONAL FEATURES FOR READING SPEECH

Feature No.	Model	Accuracy	ROC-AUC	Precision	Recall	Train Loss	Validation Loss
15	SVM	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.30 ± 0.04	0.36 ± 0.02
	RF	0.98 ± 0.03	1.00 ± 0.00	0.97 ± 0.06	1.00 ± 0.00	0.07 ± 0.01	0.10 ± 0.03
	XGB	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.05 ± 0.00	0.05 ± 0.01
	LR	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.09 ± 0.00	0.11 ± 0.03
30	SVM	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.04 ± 0.00	0.05 ± 0.02
	RF	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.05 ± 0.01	0.08 ± 0.03
	XGB	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.07 ± 0.00	0.07 ± 0.00
	LR	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.05 ± 0.00	0.06 ± 0.02
45	SVM	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.04 ± 0.00	0.04 ± 0.02
	RF	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.06 ± 0.01	0.09 ± 0.03
	XGB	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.05 ± 0.00	0.05 ± 0.01
	LR	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.04 ± 0.00	0.05 ± 0.01

TABLE II
COMPARISON OF 15, 30, AND 45 MRMR-SELECTED COMPARE FUNCTIONAL FEATURES FOR PHONATION

Feature No.	Model	Accuracy	ROC-AUC	Precision	Recall	Train Loss	Validation Loss
15	SVM	0.77 ± 0.12	0.22 ± 0.23	0.89 ± 0.13	0.63 ± 0.24	0.71 ± 0.02	0.71 ± 0.02
	RF	0.85 ± 0.11	0.94 ± 0.08	0.88 ± 0.15	0.83 ± 0.15	0.24 ± 0.03	0.38 ± 0.12
	XGB	0.85 ± 0.10	0.93 ± 0.06	0.89 ± 0.09	0.80 ± 0.12	0.43 ± 0.01	0.53 ± 0.04
	LR	0.92 ± 0.05	0.96 ± 0.06	0.97 ± 0.06	0.87 ± 0.12	0.32 ± 0.01	0.39 ± 0.07
30	SVM	0.93 ± 0.06	1.00 ± 0.00	0.97 ± 0.06	0.90 ± 0.13	0.22 ± 0.08	0.26 ± 0.06
	RF	0.95 ± 0.07	0.99 ± 0.01	0.97 ± 0.06	0.93 ± 0.13	0.20 ± 0.02	0.28 ± 0.07
	XGB	0.87 ± 0.10	0.98 ± 0.03	0.87 ± 0.11	0.87 ± 0.13	0.33 ± 0.01	0.42 ± 0.05
	LR	0.93 ± 0.06	1.00 ± 0.00	0.97 ± 0.06	0.90 ± 0.13	0.30 ± 0.01	0.32 ± 0.05
45	SVM	0.82 ± 0.14	0.82 ± 0.15	0.84 ± 0.18	0.83 ± 0.17	0.35 ± 0.02	0.46 ± 0.18
	RF	0.80 ± 0.10	0.90 ± 0.06	0.87 ± 0.13	0.73 ± 0.22	0.46 ± 0.02	0.56 ± 0.03
	XGB	0.82 ± 0.14	0.82 ± 0.15	0.84 ± 0.18	0.83 ± 0.17	0.35 ± 0.02	0.46 ± 0.18
	LR	0.80 ± 0.17	0.87 ± 0.17	0.79 ± 0.18	0.80 ± 0.27	0.60 ± 0.01	0.62 ± 0.04

TABLE III
COMPARISON OF 75, 90, AND 120 MRMR-SELECTED COMPARE FUNCTIONAL FEATURES FOR PHONATION

Feature No.	Model	Accuracy	ROC-AUC	Precision	Recall	Train Loss	Validation Loss
75	SVM	0.92 ± 0.06	1.00 ± 0.00	0.94 ± 0.08	0.90 ± 0.15	0.20 ± 0.09	0.26 ± 0.05
	RF	0.88 ± 0.07	0.99 ± 0.01	0.92 ± 0.12	0.87 ± 0.14	0.17 ± 0.02	0.27 ± 0.08
	XGB	0.85 ± 0.12	0.96 ± 0.06	0.88 ± 0.11	0.80 ± 0.18	0.44 ± 0.01	0.50 ± 0.06
	LR	0.92 ± 0.06	1.00 ± 0.00	0.94 ± 0.08	0.90 ± 0.15	0.48 ± 0.01	0.48 ± 0.02
90	SVM	0.92 ± 0.06	1.00 ± 0.00	0.94 ± 0.08	0.90 ± 0.15	0.18 ± 0.08	0.24 ± 0.04
	RF	0.93 ± 0.08	0.99 ± 0.01	0.97 ± 0.07	0.90 ± 0.13	0.16 ± 0.02	0.25 ± 0.08
	XGB	0.88 ± 0.07	0.97 ± 0.03	0.90 ± 0.08	0.87 ± 0.07	0.44 ± 0.01	0.50 ± 0.05
	LR	0.92 ± 0.05	1.00 ± 0.00	0.94 ± 0.07	0.90 ± 0.13	0.45 ± 0.01	0.46 ± 0.02
120	SVM	0.92 ± 0.06	1.00 ± 0.00	0.94 ± 0.08	0.90 ± 0.15	0.17 ± 0.08	0.23 ± 0.05
	RF	0.93 ± 0.04	0.99 ± 0.02	0.97 ± 0.06	0.90 ± 0.09	0.05 ± 0.00	0.18 ± 0.07
	XGB	0.88 ± 0.07	0.97 ± 0.05	0.90 ± 0.09	0.87 ± 0.07	0.43 ± 0.01	0.50 ± 0.05
	LR	0.92 ± 0.06	1.00 ± 0.00	0.94 ± 0.08	0.90 ± 0.15	0.44 ± 0.01	0.44 ± 0.03

and 120 features, suggesting convergence in predictive performance, with training and validation losses remaining low and stable.

B. Reading Passage

1) ML vs. DL models on ComParE_2016 features: Table IV shows that the RF, XGB, and LR classifiers achieved perfect scores (1.00) across all metrics, with SVM performing slightly lower. These results indicate that the mRMR-selected ComParE features offer strong discriminative capability for this task. The training, validation, and test losses remained uniformly low, confirming clear class boundaries and stable generalization.

Table V reports that the CNN–LSTM model attained the best performance among the deep learning architectures, reaching an accuracy of 0.96 and an ROC–AUC of 0.99. This highlights the advantage of combining convolutional feature

TABLE IV
COMPARISON OF CLASSIFIERS ON READING SPEECH TEST SET USING 30 MRMR COMPARE-SELECTED AND 22 HANDCRAFTED FUNCTIONAL FEATURES.

Dataset	Metric	SVM	RF	XGB	LR
	Accuracy	0.99 ± 0.03	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	ROC-AUC	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	Precision	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
30 mRMR-ComParE features	Recall	0.98 ± 0.05	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	Accuracy	0.99 ± 0.03	1.00 ± 0.00	0.94 ± 0.00	1.00 ± 0.00
	ROC-AUC	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.01	1.00 ± 0.00
	Precision	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
22 Handcrafted features	Recall	0.98 ± 0.05	1.00 ± 0.00	0.88 ± 0.00	1.00 ± 0.00

TABLE V
COMPARISON OF DEEP LEARNING CLASSIFIERS ON READING SPEECH TEST SET USING 65 COMPARE AND 22 HANDCRAFTED LLD FEATURES.

Dataset	Metric	Vanilla	BiDirectional	CNN-LSTM	ConvLSTM
	Accuracy	0.94 ± 0.02	0.94 ± 0.02	0.96 ± 0.01	0.95 ± 0.02
	ROC-AUC	0.98 ± 0.01	0.98 ± 0.01	0.99 ± 0.01	0.99 ± 0.01
	Precision	0.97 ± 0.01	0.97 ± 0.02	0.99 ± 0.01	0.98 ± 0.01
65 ComParE 2016 LLD features	Recall	0.91 ± 0.03	0.90 ± 0.03	0.93 ± 0.03	0.91 ± 0.04
	Accuracy	0.90 ± 0.03	0.90 ± 0.04	0.90 ± 0.03	0.90 ± 0.03
	ROC-AUC	0.96 ± 0.03	0.96 ± 0.02	0.97 ± 0.01	0.96 ± 0.03
	Precision	0.96 ± 0.02	0.95 ± 0.03	0.96 ± 0.02	0.96 ± 0.02
22 Handcrafted LLD features	Recall	0.83 ± 0.05	0.85 ± 0.06	0.84 ± 0.04	0.83 ± 0.05

TABLE VI
COMPARISON OF CLASSIFIERS ON /AA/PHONATION TEST SET USING 90 MRMR COMPARE-SELECTED AND 22 HANDCRAFTED FUNCTIONAL FEATURES.

Dataset	Metric	SVM	RF	XGB	LR
	Accuracy	0.74 ± 0.02	0.77 ± 0.02	0.75 ± 0.13	0.75 ± 0.00
	ROC-AUC	0.85 ± 0.01	0.88 ± 0.03	0.82 ± 0.10	0.77 ± 0.00
	Precision	0.76 ± 0.03	0.83 ± 0.03	0.82 ± 0.17	0.78 ± 0.00
90 mRMR ComParE features	Recall	0.70 ± 0.00	0.68 ± 0.10	0.66 ± 0.08	0.70 ± 0.00
	Accuracy	0.80 ± 0.05	0.80 ± 0.05	0.75 ± 0.06	0.78 ± 0.05
	ROC-AUC	0.93 ± 0.05	0.93 ± 0.05	0.76 ± 0.14	0.92 ± 0.05
	Precision	0.91 ± 0.11	0.91 ± 0.11	0.66 ± 0.22	1.00 ± 0.00
22 Handcrafted features	Recall	0.68 ± 0.07	0.68 ± 0.07	0.70 ± 0.19	0.60 ± 0.00

extraction with temporal modeling.

Overall, the ML classifiers surpassed the DL models on the reading-speech dataset. Their comparable loss values across all phases indicate stable learning without notable underfitting or overfitting. By contrast, the DL models exhibited higher losses, likely due to their larger parameter space and the difficulty of optimization under limited data, whereas the simpler ML models remained more robust.

2) ML vs. DL models on Handcrafted features: As shown in Table IV, the RF and LR classifiers achieved perfect scores (1.00), outperforming XGB and SVM. The higher loss values observed across models indicate reduced class separability when using handcrafted features.

Table V shows that the CNN–LSTM achieved the best DL performance with an accuracy of 0.90 and an ROC–AUC of 0.97; however, the losses across all phases remained elevated, reflecting weaker generalization.

Overall, the ML models outperformed the DL models on handcrafted features, with RF demonstrating the most consistent performance.

C. Sustained Phonation

1) ML vs. DL models on ComParE_2016 features: For the /aa/ phonation task, RF attained the highest performance with an accuracy of 0.77 and an ROC–AUC of 0.88, followed by XGB (0.75) and SVM (0.74), as shown in Table VI. Logistic

TABLE VII
COMPARISON OF DEEP LEARNING CLASSIFIERS ON /AA/ PHONATION TEST SET USING 65 COMPARE AND 22 HANDCRAFTED LLD FEATURES.

Dataset	Metric	LSTM	BiLSTM	CNN-LSTM	ConvLSTM
65 ComParE 2016 LLD features	Accuracy	0.75 ± 0.08	0.82 ± 0.06	0.83 ± 0.06	0.83 ± 0.06
	ROC-AUC	0.79 ± 0.08	0.81 ± 0.08	0.87 ± 0.06	0.87 ± 0.07
	Precision	0.77 ± 0.05	0.82 ± 0.03	0.85 ± 0.04	0.84 ± 0.04
	Recall	0.71 ± 0.12	0.81 ± 0.12	0.82 ± 0.10	0.80 ± 0.10
22 Handcrafted LLD features	Accuracy	0.70 ± 0.06	0.73 ± 0.03	0.71 ± 0.04	0.70 ± 0.06
	ROC-AUC	0.74 ± 0.45	0.78 ± 0.06	0.77 ± 0.07	0.80 ± 0.10
	Precision	0.78 ± 0.50	0.82 ± 0.05	0.78 ± 0.03	0.79 ± 0.08
	Recall	0.56 ± 0.32	0.59 ± 0.07	0.58 ± 0.09	0.54 ± 0.09

TABLE VIII
COMPARISON OF PARKINSON'S DISEASE DETECTION STUDIES USING SPEECH AND ML/DL MODELS

Study	Dataset	Model	Features	Sound Type	Train/Val/Test	Acc.	AUC	Prec.	Recall	F1	Loss
Milosz Dudek, Daria Hemmerling et al.	21 PD, 36 HC	CNN-LSTM	SHAP on OpenSMILE ComParE LLDs + F1–F5 formants	Phonation	5-fold CV	–	–	0.96	0.95	0.96	–
Mohammad A. Hossain, Francesco Amenta	188 PD, 64 HC (SMOTE)	SVC + AdaBoost	755 acoustic, spectral, cepstral features	Phonation	10-fold CV	1.00	–	0.92	–	0.91	–
Anu Iyer, Aaron Kemp et al.	40 PD, 41 HC	Inception V3	Acoustic, cepstral features + spectrograms	Phonation	Train/Test	–	0.97	–	–	–	–
Liaqat Ali, Ce Zhu et al.	20 PD, 20 HC	ANN	[2] selected 3 out of 26 time-frequency features	Vowels, numbers, words	LOSO CV	0.97	–	–	1.00	–	–
Antonio Suppa, Giovanni Costantini et al.	115 PD, 108 HC	SVM	CFS + IGAE on ComParE functionals	Phonation	10-fold CV	0.79	0.87	–	0.82	–	–
Mehmet B. Er, Esme Isik et al.	PC-GITA Spanish Dataset	ResNet101 + LSTM	ResNet-101 deep features	Monologue, vowels, reading	10-fold CV	0.98	–	–	0.96	0.95	–
Amjad Rehman, Tanzila Saba et al.	23 PD, 8 HC	LSTM + GRU	19 acoustic features	Phonation	Train/Test	0.98	–	1.00	0.95	0.97	–
Proposed Method	**20 PD, 20 HC**	**XGB**	**mRMR-MIQ on 30 ComParE functionals**	**Reading Passage**	**5-fold CV**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.07**

Regression yielded similar accuracy (0.75) but the lowest ROC–AUC (0.77). The consistently higher test losses across models indicate limited generalization.

Among the DL classifiers, the CNN–LSTM achieved the best results, with an accuracy of 0.83 and an ROC–AUC of 0.87, as shown in Table VII. However, its validation and test losses remained substantially elevated, indicating overfitting.

Overall, DL classifiers outperformed ML models in terms of accuracy and ROC–AUC, but both approaches exhibited high loss values. These trends suggest limited discriminative strength in the /aa/ phonation task compared to reading speech.

2) ML vs. DL models on Handcrafted features: Among the ML classifiers, SVM delivered the best results, achieving a test accuracy of 0.80 and an ROC–AUC of 0.93, as shown in Table VI.

Within the DL group, BiLSTM achieved the highest accuracy (0.73), but its validation and test losses were substantially higher, reducing the reliability of the result, as shown in Table VII. Similar trends were observed across the other DL architectures, indicating a pronounced degree of overfitting.

Overall, the ML classifiers trained on handcrafted features achieved higher accuracy and ROC–AUC values than the DL models. Higher loss values suggest lower generalization in /aa/ sustained phonation data.

V. CONCLUSION

For Parkinson's disease (PD) classification in a data-constrained setting, machine learning (ML) classifiers using mRMR-selected ComParE features from the reading-speech dataset achieved the best performance. This highlights the superior representational and discriminative power of reading-speech features for PD detection. On reading speech, deep learning (DL) classifiers did not match the performance of ML classifiers, and ComParE features consistently outperformed handcrafted ones.

In contrast, /aa/ phonation provided weaker feature representations, as reflected in the overall performance of both ML and DL classifiers. Nevertheless, DL models captured relevant patterns in phonation more effectively than ML models when trained on ComParE features, while ML classifiers performed better with handcrafted features. Overall, DL models trained on ComParE features achieved the best results for the /aa/ sustained phonation task.

XGB applied to reading-speech data with mRMR-selected ComParE features was identified as the best-performing model, achieving perfect accuracy with minimal divergence across training, validation, and test losses, and demonstrating clear improvements over prior work (Table VIII). The mRMR–ComParE pipeline isolated the most informative acoustic cues, enabling reliable separation of Parkinsonian and healthy speech in a data-constrained setting. In contrast to earlier studies that depended on complex deep learning architectures or narrow phonation-based inputs, our use of reading passages coupled with lightweight ML models yielded richer speech representations for PD classification.

In contrast to earlier studies that depended on complex deep learning architectures or narrow phonation-based inputs, our use of reading passages coupled with lightweight ML classifiers enhanced through optimal feature selection and hyperparameter tuning yielded high performance for PD classification.

REFERENCES

[1] J. W. Langston, "Parkinson's disease: current and future challenges," *Neurotoxicology*, vol. 23, no. 4-5, pp. 443–450, 2002.

[2] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Systems with Applications*, vol. 137, pp. 22–28, 2019.

[3] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of neurology, neurosurgery & psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[4] A. Khorasani and M. R. Daliri, "Hmm for classification of parkinson's disease based on the raw gait data," *Journal of medical systems*, vol. 38, pp. 1–6, 2014.

[5] N. Singh, V. Pillay, and Y. E. Choonara, "Advances in the treatment of parkinson's disease," *Progress in neurobiology*, vol. 81, no. 1, pp. 29–44, 2007.

[6] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a parkinson speech dataset with multiple types of sound recordings," *IEEE journal of biomedical and health informatics*, vol. 17, no. 4, pp. 828–834, 2013.

[7] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 4–21, 2016.

[8] M. Little, P. McSharry, E. Hunter, and J. Spielman, "Lo ramig suitability of dysphonia measurements for telemonitoring of parkinson's disease., 2009, 56," *DOI: https://doi. org/10.1109/TBME*, pp. 1015–1022, 2008.

[9] J. M. Tracy, Y. Özkanca, D. C. Atkins, and R. H. Ghomi, "Investigating voice as a biomarker: deep phenotyping methods for early detection of parkinson's disease," *Journal of biomedical informatics*, vol. 104, p. 103362, 2020.

[10] H. Azadi, M.-R. Akbarzadeh-T, A. Shoeibi, H. R. Kobravi *et al.*, "Evaluating the effect of parkinson's disease on jitter and shimmer speech features," *Advanced Biomedical Research*, vol. 10, no. 1, p. 54, 2021.

[11] S. Skodda, W. Grönheit, and U. Schlegel, "Intonation and speech rate in parkinson's disease: General and dynamic aspects and responsiveness to levodopa admission," *Journal of Voice*, vol. 25, no. 4, pp. e199–e205, 2011.

[12] L. Ali, Z. He, W. Cao, H. T. Rauf, Y. Imrana, and M. B. Bin Heyat, "Mmdd-ensemble: A multimodal data–driven ensemble approach for parkinson's disease detection," *Frontiers in Neuroscience*, vol. 15, p. 754058, 2021.

[13] N. Narendra and P. Alku, "Automatic assessment of intelligibility in speakers with dysarthria from coded telephone speech using glottal features," *Computer Speech & Language*, vol. 65, p. 101117, 2021.

[14] A. Bayestehtashk, M. Asgari, I. Shafran, and J. McNames, "Fully automated assessment of the severity of parkinson's disease from speech," *Computer speech & language*, vol. 29, no. 1, pp. 172–185, 2015.

[15] M. A. Hossain and F. Amenta, "Machine learning-based classification of parkinson's disease patients using speech biomarkers," *Journal of Parkinson's Disease*, vol. 14, no. 1, pp. 95–109, 2024.

[16] L. Ali, A. Javeed, A. Noor, H. T. Rauf, S. Kadry, and A. H. Gandomi, "Parkinson's disease detection based on features refinement through l1 regularized svm and deep neural network," *Scientific Reports*, vol. 14, no. 1, p. 1333, 2024.

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

[18] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE transactions on biomedical engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[19] 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. 22 199–22 208, 2017.

[20] B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, and K. Evanini, "The interspeech 2016 computational paralinguistics challenge: deception, sincerity and native language," 2016.

[21] I. Södergren, M. P. Nodeh, P. C. Chhipa, K. Nikolaidou, and G. Kovács, "Detecting covid-19 from audio recording of coughs using random forests and support vector machines." in *Interspeech*, 2021, pp. 916–920.

[22] A. Suppa, G. Costantini, F. Asci, P. Di Leo, M. S. Al-Wardat, G. Di Lazzaro, S. Scalise, A. Pisani, and G. Saggio, "Voice in parkinson's disease: a machine learning study," *Frontiers in neurology*, vol. 13, p. 831428, 2022.

[23] A. Iyer, A. Kemp, Y. Rahmatallah, L. Pillai, A. Glover, F. Prior, L. Larson-Prior, and T. Virmani, "A machine learning method to process voice samples for identification of parkinson's disease," *Scientific reports*, vol. 13, no. 1, p. 20615, 2023.

[24] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[25] L. Zahid, M. Maqsood, S. S. Farooq, F. Aadil, I. Mehmood, M. Fiaz, and S. K. Jung, "Detection of speech impairments in parkinson disease using handcrafted feature-based model on spanish speech corpus," in *International Workshop on Frontiers of Computer Vision (IW-FCV 2020)*, ser. Communications in Computer and Information Science, vol. 1212. Springer Nature Singapore, 2020, pp. 54–65.

VI. APPENDIX

The final mRMR-selected features for the reading speech dataset are listed in Table IX, while the corresponding features for the /aa/ phonation dataset are provided in Table X.

TABLE IX
SELECTED 30 MRMR COMPARE FUNCTIONAL FEATURES FOR READING SPEECH

pcm_fftMag_spectralRollOff25.0_sma_quartile1	mfcc_sma[2]_pctlrange0-1	audSpec_Rfilt_sma[24]_lpc4
audSpec_Rfilt_sma_de[16]_minPos	pcm_zcr_sma_quartile1	mfcc_sma[2]_amean
audSpec_Rfilt_sma[1]_minSegLen	mfcc_sma[2]_flatness	mfcc_sma[8]_quartile3
audSpec_Rfilt_sma_de[16]_minSegLen	pcm_zcr_sma_percentile1.0	pcm_fftMag_spectralRollOff25.0_sma_rqmean
pcm_fftMag_spectralRollOff90.0_sma_de_minSegLen	logHNR_sma_posamean	mfcc_sma[2]_lpc0
pcm_zcr_sma_quartile2	pcm_fftMag_spectralRollOff25.0_sma_risetime	pcm_fftMag_psySharpness_sma_percentile1.0
pcm_fftMag_spectralRollOff90.0_sma_percentile1.0	audSpec_Rfilt_sma[5]_stddev	audspec_lengthL1norm_sma_percentile1.0
pcm_fftMag_spectralCentroid_sma_de_peakRangeAbs	pcm_fftMag_spectralRollOff50.0_sma_percentile1.0	pcm_fftMag_spectralRollOff50.0_sma_quartile1
mfcc_sma[6]_range	audspecRasta_lengthL1norm_sma_qregc3	mfcc_sma[14]_percentile99.0
mfcc_sma[5]_peakMeanAbs	audSpec_Rfilt_sma_de[10]_stddevRisingSlope	audSpec_Rfilt_sma_de[5]_meanRisingSlope

TABLE X
SELECTED 90 MRMR COMPARE FUNCTIONAL FEATURES FOR /AA/ PHONATION

logHNR_sma_upleveltime90	mfcc_sma[10]_iqr1-3	mfcc_sma[1]_upleveltime75
audSpec_Rfilt_sma[3]_minPos	pcm_fftMag_spectralRollOff25.0_sma_de_leftctime	pcm_fftMag_spectralRollOff25.0_sma_peakRangeRel
pcm_fftMag_spectralCentroid_sma_de_minPos	pcm_fftMag_fband1000-4000_sma_lpc4	mfcc_sma[14]_peakRangeAbs
audSpec_Rfilt_sma_de[7]_iqr2-3	logHNR_sma_quartile1	logHNR_sma_upleveltime75
mfcc_sma_de[3]_peakMeanRel	logHNR_sma_de_upleveltime75	mfcc_sma[10]_pctlrange0-1
pcm_fftMag_spectralFlux_sma_upleveltime25	mfcc_sma[1]_lpc2	mfcc_sma[14]_maxPos
audSpec_Rfilt_sma[0]_minRangeRel	audSpec_Rfilt_sma[1]_minRangeRel	mfcc_sma_de[11]_peakMeanMeanDist
pcm_zcr_sma_de_peakMeanRel	pcm_fftMag_spectralRollOff25.0_sma_upleveltime90	logHNR_sma_qregc2
mfcc_sma[7]_percentile1.0	mfcc_sma[5]_flatness	pcm_fftMag_fband250-650_sma_de_flatness
pcm_fftMag_spectralEntropy_sma_de_iqr1-3	logHNR_sma_centroid	pcm_fftMag_fband1000-4000_sma_kurtosis
mfcc_sma_de[5]_peakMeanMeanDist	mfcc_sma[3]_pctlrange0-1	pcm_fftMag_spectralRollOff25.0_sma_risetime
mfcc_sma[5]_quartile1	mfcc_sma_de[14]_percentile1.0	audSpec_Rfilt_sma_de[7]_iqr1-2
pcm_fftMag_spectralHarmonicity_sma_lpc0	mfcc_sma[7]_range	audSpec_Rfilt_sma[15]_upleveltime50
mfcc_sma[3]_qregc3	pcm_fftMag_spectralVariance_sma_de_peakMeanRel	mfcc_sma[11]_lpgain
audSpec_Rfilt_sma[1]_upleveltime50	pcm_RMSenergy_sma_qregc3	mfcc_sma[5]_percentile1.0
audSpec_Rfilt_sma[24]_kurtosis	mfcc_sma_de[11]_iqr1-2	audSpec_Rfilt_sma[12]_iqr2-3
shimmerLocal_sma_de_upleveltime50	audSpec_Rfilt_sma[0]_upleveltime75	pcm_fftMag_spectralRollOff25.0_sma_de_minRangeRel
audSpec_Rfilt_sma[3]_percentile1.0	mfcc_sma[9]_peakRangeAbs	audSpec_Rfilt_sma[24]_upleveltime50
mfcc_sma_de[13]_posamean	pcm_fftMag_spectralEntropy_sma_de_minRangeRel	audSpec_Rfilt_sma[17]_skewness
mfcc_sma_de[1]_rqmean	pcm_fftMag_spectralRollOff90.0_sma_de_maxPos	audSpec_Rfilt_sma_de[10]_iqr1-3
logHNR_sma_stddev	audSpec_Rfilt_sma_de[17]_iqr1-2	logHNR_sma_flatness
mfcc_sma[5]_quartile3	audSpec_Rfilt_sma_de[7]_quartile1	pcm_fftMag_spectralHarmonicity_sma_centroid
audSpec_Rfilt_sma[0]_upleveltime50	pcm_fftMag_spectralRollOff75.0_sma_upleveltime50	mfcc_sma[11]_peakRangeAbs
pcm_zcr_sma_de_lpgain	mfcc_sma[8]_pctlrange0-1	mfcc_sma_de[9]_peakMeanAbs
pcm_fftMag_spectralRollOff25.0_sma_peakMeanRel	pcm_fftMag_fband250-650_sma_flatness	mfcc_sma_de[1]_iqr1-2
mfcc_sma[10]_stddev	pcm_zcr_sma_meanFallingSlope	pcm_fftMag_spectralHarmonicity_sma_de_flatness
audSpec_Rfilt_sma[2]_iqr1-2	pcm_zcr_sma_lpc2	audSpec_Rfilt_sma[16]_upleveltime25
audSpec_Rfilt_sma_de[8]_quartile1	audspec_lengthL1norm_sma_qregc2	mfcc_sma[12]_peakRangeAbs
pcm_fftMag_spectralRollOff90.0_sma_de_minPos	audSpec_Rfilt_sma[3]_iqr1-2	audSpec_Rfilt_sma_de[10]_upleveltime50
mfcc_sma[11]_lpc0	audSpec_Rfilt_sma[0]_upleveltime25	pcm_fftMag_spectralRollOff25.0_sma_quartile2