One-dimensional Cellular Automata-based Encoding of Speech Features for Parkinson's Disease Classification

Hema Sudha Rani * and Senthil Athithan

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Guntur (Dt.), Andhra Pradesh, India
Email: hemasudhareddy@gmail.com (H.S.R.); senthilathithan@hotmail.com (S.A.)
*Corresponding author

***Abstract*—Parkinson's Disease (PD) transforms characteristic neuromotor abnormalities into measurable speech output changes. In particular, this paper presents a signal-processing framework for structured encoding of acoustic speech information prior to categorization based on One-Dimensional Cellular Automata (1DCA). Mel-Frequency Cepstral Coefficients (MFCCs), jitter, shimmer, and Harmonic-to-Noise Ratio (HNR) were extracted to form a 16-dimensional acoustic feature vector. Normalized features from the CA input set are binarized and represent an initialization on a 1D CA grid, where any input configuration of the 1D grid evolves under Rule 110 to perform a neighborhood-aware non-linear feature transformation that models local interactions among adjacent acoustic features and produces a CA-enhanced representation for Parkinson's disease classification. A feed-forward Deep Neural Network (DNN) is used to classify the CA-transformed representations. Experiments with the publicly available Italian Parkinson's Voice and Speech (IPVS) dataset showed that coding based on CA improves discriminative performance, compared to traditional feature-based baselines. The proposed framework achieved a mean five-fold cross-validation accuracy of 95.14% with a Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) of 0.9895. Sensitivity analysis and ablation experiments were conducted to examine the contribution of the cellular automata configuration to classification performance. Thus, these results demonstrate that the encoding via cellular automata features is a computationally efficient and interpretable augmentation of acoustic biomarkers for potential use in automated speech-based screening systems for Parkinson's disease.**

***Keywords*—Parkinson's disease, speech analysis, cellular automata, deep neural network, Mel-Frequency Cepstral Coefficients (MFCCs), jitter, shimmer, Harmonic-to-Noise Ratio (HNR)**

I. Introduction

Parkinson's Disease (PD) is one of the most prevalent neurodegenerative diseases and it is characterized by motor dysfunction, dementia, depression, and a variety of non-motor symptoms (e.g., cognitive decline, mood disturbances). PD also affects the speech production mechanism in addition to motor functions, resulting in changes across respiratory, phonatory, and prosodic aspects of speech. These changes may reduce vocal intensity, affect speech intonation, and decrease articulatory precision, which can be examined through acoustic analysis of voice recordings [1]. Clinically, the early and objective detection of PD is important as this allows earlier intervention and better monitoring of disease progression, facilitating improved therapy personalization; however, standard clinical examinations are often subjective, laborious and insensitive to subtle early-stage signs.

This study aims to support non-invasive, low-cost, and scalable screening for Parkinson's Disease (PD) using voice-based biomarkers based on voice analysis, a modality that can easily be monitored in remote/telemedicine settings with simple recordings of phonation tasks. Previous studies have identified acoustic, prosodic, and articulatory characteristics that can help distinguish individuals with Parkinson's Disease (PD) from healthy controls. Commonly investigated features include Mel-frequency Cepstral Coefficients (MFCCs), jitter, shimmer, and the Harmonic-to-Noise Ratio (HNR). These features have been used in machine-learning and deep-learning models for PD classification, with promising results reported in previous studies [2]. Although deep-learning approaches have been explored for modelling speech and language in Parkinson's disease, effectively capturing complex temporal dependencies and nonlinear relationships among acoustic speech features remains an important challenge.

We propose a framework that uses 1D-Cellular Automata (CA) feature encoding to improve conventional acoustic features before classification. This method models local neighborhood interactions through iterative Cellular Automata evolution and nonlinear patterns that are hard to represent with static vector encodings alone, by mapping the extracted acoustic feature vector into 1D CA

Manuscript received June 2, 2026; revised July 3, 2026; accepted August 4, 2026; published November 25, 2026.

cell states and evolving under Rule 110. The patterns by 1DCA are then classified with a Deep Neural Network (DNN), which provides greater discrimination between PD and healthy control recordings in a publicly available Italian Parkinson's Voice and Speech dataset (IPVS) [3].

II. Related Work

The early and common non-motor sign of PD is speech impairment, stemming from dopaminergic dysfunction involved in controlling respiration, phonation and articulatory planning. Clinical and acoustic descriptions indicate that PD speech is hypophonic, breathy, monotonous, with decreased loudness variability, imprecise consonants and deviant prosody; these traits can be measured or detected using appropriate signal-processing techniques. Speech-based assessment has been investigated as a non-invasive approach for remotely monitoring Parkinson's disease progression, using acoustic features extracted from speech recordings to estimate symptom severity [4, 5].

There is a sustained research interest in the effectiveness of speech as a remote assessment tool for voice-based PD diagnosis, and systematic reviews have identified its potential benefit for longitudinal tracking of disease progression [6]. Simultaneously, however, the reviews highlight that many studies tend to rely upon comparatively small, homogeneous corpora and diverse preprocessing pipelines that can restrict generalizability and complicate efforts to make direct comparisons of performance across methods.

A. Acoustic Features for PD Detection

Jitter, shimmer, and Harmonics-to-Noise Ratio (HNR) are acoustic measures used to characterize voice quality. Jitter and shimmer quantify variations in fundamental frequency and amplitude, respectively, while HNR reflects the relative proportion of harmonic components to noise in the voice signal. In Parkinson's Disease (PD), increased jitter and shimmer and reduced HNR may indicate less stable phonation. These acoustic characteristics can help describe voice-quality abnormalities, including breathiness and roughness that are relevant to speech assessment. Analysing these measures in the time and frequency domains can provide useful information for characterizing speech abnormalities associated with PD.

Cepstral representations are widely used in speech analysis because they capture the spectral patterns of speech signals. Mel-Frequency Cepstral Coefficients (MFCCs) are widely applied in speech processing to describe the distribution of spectral energy within short segments of an audio signal. In Parkinson's disease detection, these features can be combined with voice measurements, including jitter, shimmer, and the Harmonic-to-Noise Ratio (HNR). While these measurements characterize variations in vocal stability and the balance between harmonic and noise components, MFCCs provide complementary information about the acoustic structure of speech, supporting a more comprehensive analysis for classification.

The use of MFCCs as well as other types of acoustic features as input to conventional machine learning algorithms such as Support Vector Machines (SVM) and ensembles for classification have been evaluated. Results reported from these evaluations vary based upon the dataset, the way that the features are represented and the evaluation methodology used. Computation of evolutionary feature selection techniques was also performed for classification of Parkinson's disease using vocal features [7]. Machine-learning and Deep-learning approaches have also been investigated for Parkinson's disease detection using multivariate vocal data [8–10]. Deep neural network embeddings have also been investigated alongside interpretable speech features for automatic Parkinson's disease assessment across multilingual scenarios [11]. Voice-based feature extraction and classification have also been investigated for distinguishing Parkinson's disease patients from healthy individuals [12]. Research has established that the best measures of Parkinsonian dysphonia appear to be combinations of MFCCs with jitter, shimmer, HNR and non-linear dysphonia descriptors, such combinations regularly seen in more recent PD voice datasets.

B. Machine and Deep Learning Approaches

Traditional Parkinson's disease speech-classification pairs handcrafted acoustic features with machine-learning algorithms. Commonly used acoustic features, including Mel-Frequency Cepstral Coefficients (MFCCs), jitter, shimmer, and Harmonic-to-Noise Ratio (HNR), have been investigated for Parkinson's disease detection using machine-learning classifiers. Feature-based deep learning approaches have combined autoencoder-based feature extraction with deep neural network classifiers to detect Parkinson's disease from audio recordings [2]. Machine-learning algorithms are increasingly applied to track progression of Parkinson's disease using speech signals, with investigations including multiple types of speech signals. Recently, deep learning has been used for PD voice analysis, with Convolutional Neural Networks (CNNs), recurrent architectures such as Long Short-Term Memory (LSTM) and Bidirectional Long Short Term Memory (BiLSTM), and self-supervised speech models that learn discriminative representations directly from spectrograms or raw waveforms [13, 14]. Deep neural networks using speech spectrograms have also been investigated to distinguish Parkinson's disease from spinocerebellar degeneration, demonstrating the potential of speech-based models for differentiating between neurological disorders. Wav2Vec 1.0 embeddings have also been investigated for Parkinson's disease detection across multiple speech databases, demonstrating their potential for cross-database classification and speech-feature analysis [15].

Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have been investigated for Parkinson's disease detection using acoustic features such as Mel-frequency Cepstral Coefficients (MFCCs) and spectrogram representations. Their reported performance varies depending on the dataset, recording conditions, language, and evaluation

protocol. Recent studies have also explored fine-tuned self-supervised speech encoders.

C. Cellular Automata in Biomedical Signal Modelling

Cellular Automata (CA) are deterministic discrete dynamical systems in which cells evolve according to local update rules defined over a neighbourhood [16]. These systems can generate complex global behaviour from simple local interactions and have been applied to the modelling of biological processes. Cellular automata frameworks have been used to simulate infectious disease spread [17] and tumour growth dynamics [18, 19]. These applications demonstrate the potential of CA to represent local interactions and evolving spatial structures in biological systems. Such properties motivate further investigation of CA-based approaches for structured feature representation in biomedical signal analysis.

Recently, CA-based methods for edge detection and segmentation as well as noise suppression techniques have been proposed in a way that selection of proper rules improves structures while suppressing noise in the context of biomedical image and signal processing.

Studies on CA-driven medical image analysis show that CA rules can accentuate boundaries and local structures that conventional filters may blur, and CA–neural network architectures have been explored to combine interpretable local evolution with powerful statistical learning. These findings motivate the use of one-dimensional CA to transform acoustic feature vectors into structured nonlinear representations by iteratively updating neighboring feature states according to Cellular Automata rules. This neighborhood-aware transformation emphasizes local feature relationships that may be associated with Parkinsonian speech irregularities.

D. Motivation for 1D CA–DNN Models

Although there has been significant advancement in developing effective methods for classifying speech from individuals with Parkinson's Disease (PD), many current PD speech classification approaches have several limitations. For example, most of the existing approaches to PD speech classification utilize frame-based feature vector representations of speech signals which do not provide a complete representation of the nonlinear relationships between various acoustic speech parameters and Parkinson's disease. Furthermore, most of the deep learning approaches used for PD speech classification are viewed as black boxes, and therefore, it is difficult to determine what specific acoustic patterns contribute to the classification decision [20].

Using local updating rules, 1D CA provide an organized yet adaptable framework for modelling local neighborhood interactions among acoustic features. Various CA rules result in unique patterns that can be qualitatively linked to acoustic feature relationships. Prior applications of CA in biomedical signal and image analysis such as noise suppression [21] and tissue-growth simulation demonstrate their capacity to highlight local structures while suppressing irrelevant variability. This encourages incorporating a nonlinear feature transformation using a 1D CA layer prior to deep neural network classification to achieve improved nonlinear feature representation and robustness to noise and more interpretable intermediate representations in comparison with traditional feature-Deep Neural Network (DNN) pipelines.

Elementary Cellular Automata (ECA) consist of 256 transition rules, each exhibiting distinct evolution characteristics. Among these, Rules 30, 90, and 110 have been extensively investigated because of their well-established theoretical properties and complementary dynamical behaviors. Rule 30 generates highly irregular and pseudo-random evolution patterns, Rule 90 produces symmetric and self-similar structures, whereas Rule 110 exhibits complex dynamics and has been proven to be computationally universal [22]. Due to these properties Rules 30, 90, and 110 are commonly studied in cellular automata research because they exhibit different patterns of evolution. Thus, for this comparative evaluation within our study Rules 30, 90 and 110 were chosen as sample accessible elementary Cellular Automata rules.

Rule 30 generates irregular, pseudo-random patterns through successive cellular updates, as illustrated in Fig. 1. These patterns demonstrate the complex spatial structures that can emerge from a simple local update rule. In this study, Rule 30 was evaluated as a candidate for transforming acoustic feature vectors into structured representations for classification.

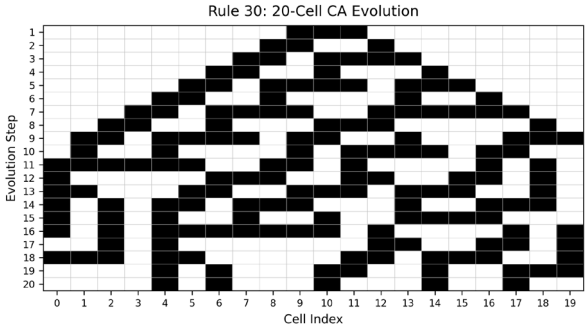

Fig. 1. Representation of Rule 30.

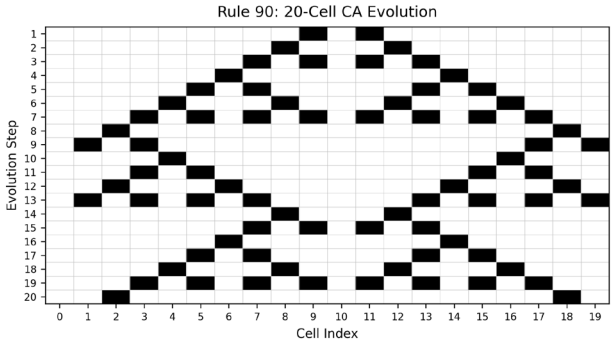

Fig. 2. Representation of Rule 90.

As illustrated in Fig. 2, Rule 90 generates symmetric and self-similar patterns through its cellular update process. These patterns provide a contrasting representation to the irregular patterns generated by Rule 30. Their usefulness for Parkinson's disease classification

is evaluated through the experimental comparison of the selected CA rules.

Fig. 3 shows that rule 110, evolves into complex yet ordered patterns that balance repetitiveness and variation, making it well-suited for generating structured nonlinear feature representations from acoustic speech features (e.g., prosodic flattening, intermittent instability).

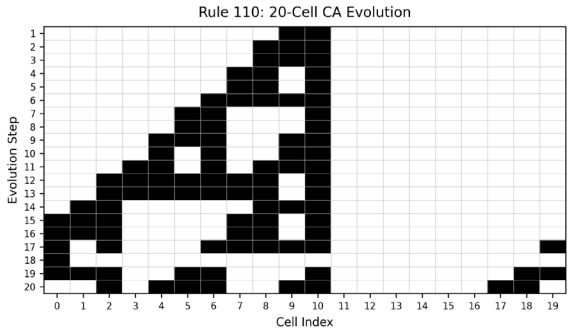

Fig. 3. Representation of Rule 110.

Among the three investigated rules, Rule 30 generates irregular, pseudo-random patterns, Rule 90 produces symmetric and self-similar patterns, and Rule 110 exhibits complex evolution with both structured and irregular characteristics. These differences provide a basis for comparing their effectiveness in transforming acoustic feature vectors. Accordingly, Rules 30, 90, and 110 were evaluated under the same experimental settings, and Rule 110 was selected based on their comparative classification performance. This normalized acoustic feature vector with 13 MFCC coefficients, jitter, shimmer and HNR is mapped on a one-dimensional CA grid where it is iteratively evolved using Rule 110 for the purpose of generating a nonlinear neighborhood-aware feature before classification by Deep Neural Networks.

While utilizing Cellular Automata in biomedical image analysis and computational biology have shown promise, it has yet to be used for speech-based deep learning detection of Parkinson's disease. Specifically, limited research has been conducted on the systematic study of CA rule selection/parameter/configuration to transform acoustically modeled speech features into discriminative representations. To address this limitation, the proposed 1D CA–Deep Neural Network (DNN) framework first encodes traditional acoustic features into structured Cellular Automata evolution patterns for classification, thus providing enhanced discriminability, robustness, and interpretability compared with conventional feature-based Deep Neural Network (DNN) methods.

Table I presents a comparative analysis of representative Parkinson's disease speech-classification methods and highlights the motivation for the proposed framework.

TABLE I. COMPARATIVE ANALYSIS OF EXISTING PARKINSON'S DISEASE DETECTION METHODS

S.No	Author & Year	Data set	Method	Strengths	Limitations
1	Tsanas *et al*. (2010) [4]	UCI Parkinson Dataset	Acoustic Features + Regression Models	Effective telemonitoring of PD progression	Limited nonlinear feature representation
2	Sakar *et al*. (2017) [5]	UCI Parkinson Dataset	Vocal Features + Machine Learning	Good classification accuracy	Sensitive to feature selection
3	Nagasubramanian and Sankayya (2021) [8]	Parkinson Voice Dataset	Deep Neural Network	Learns complex speech patterns	Requires larger training data
4	Suppa *et al*. (2022) [9]	Voice Recordings	Machine Learning Framework	High detection performance	Reduced interpretability
5	Dao *et al*. (2022) [7]	Parkinson Speech Dataset	Evolutionary Feature Selection + ML	Improved feature optimization	Increased computational cost
6	Khaskhoussy and Ayed (2022) [10]	PD Speech Dataset	Deep Learning-Based Diagnosis	Automatic feature extraction	High computational complexity
7	Valarmathi *et al*. (2025) [2]	Voice Dataset	Autoencoder + Neural Network	Enhanced feature representation	Black-box decision process
8	Çelik and Akbal (2025) [12]	Speech Dataset	Voice Feature Extraction + Classification	Robust speech-based detection	Limited neighborhood feature representation
9	Proposed Work	IPVS Dataset	MFCC + Jitter + Shimmer + HNR + 1D Cellular Automata + DNN	Captures structured cellular automata feature representation high accuracy (95.14%)	Requires validation on multi-language and real-world datasets

Note: Italian Parkinson's Voice and Speech (IPVS), Deep Neural Network (DNN), University of California, Irvine (UCI) Parkinsons Telemonitoring Dataset

III. METHODOLOGY

A. Dataset

For the experimental evaluation, we used a subset of the publicly available Italian Parkinson's Voice and Speech (IPVS) dataset [3], available through IEEE DataPort. The original corpus contains 831 recordings from 65 participants: 15 young healthy controls, 22 elderly healthy controls, and 28 individuals with Parkinson's Disease (PD). During the initial feature-extraction process, 200 recordings for which one or more of the 16 acoustic features could not be computed, resulting in undefined or NaN values, were excluded. The remaining 631 recordings, associated with 53 of the original 65 participant folders, were retained for analysis. Feature extraction was performed on the selected recordings, and the resulting feature set was verified to contain no missing values. The final subset comprised 250 recordings from Healthy Controls (HC) and 381 recordings from individuals with PD, associated with 53 participant folders (27 HC and 26 PD). These figures distinguish recording-level counts from participant-folder counts. Since the samples included in our analysis were chosen based on the

output of feature extraction techniques, rather than randomly selected from the larger corpus, the analyzed subset does not represent a truly random sampling of the entire corpus. Our subset has an estimated class distribution of 40% HC and 60% PD, which differs slightly from the 47% HC and 53% PD contained within the larger corpus. We have combined both the younger and older healthy control groups into one group of HC; however, we have maintained the assignment of PD labels to individuals as they were originally assigned by the dataset. To ensure accuracy, each of the 631 selected files was cross-referenced with the complete inventory of available files and we did not find any duplicate files or inconsistencies with regard to class label assignments. Additionally, we verified the class labels for all 631 selected recordings against their corresponding labels in the IPVS dataset and found them to be consistent.

The IPVS database contains four different types of spoken language tasks; text reading, reading words/phrases, repeating syllables, and sustaining vowel phonemes. Each recording has been processed using 13 Mel-frequency Cepstral Coefficients (MFCCs), jitter, shimmer, and Harmonics to Noise Ratio (HNR) to produce a 16-dimensional acoustic feature vector. The MFCCs represent spectral characteristics of speech while the other variables relate to the quality of the voice. All 16 features have then been provided as inputs into the cellular automata-based feature transformation and classification system developed by the researchers.

Performance of the models was assessed through a five-fold StratifiedGroupKFold cross-validation method; participant folders served as the grouping variable. A total of 631 recordings identified in the validated inventory were associated with 53 participant folders (27 Healthy Controls (HC) and 26 Parkinson's Disease (PD). Each recording from a given participant folder was included in the same fold. An audit of fold assignment confirmed there was no overlap in participant folder inclusion between the training set and test set. There were 7 to 13 participant folders and 84 to 156 recordings in each test fold.

Table II summarizes the participant-folder and recording distributions across the five test folds. Because participants contributed different numbers of recordings, the recording-level class proportions varied across folds. Accordingly, results are reported as mean across the five folds, together with pooled out-of-fold results.

TABLE II. FOLD COMPOSITION OF THE PARTICIPANT-GROUPED FIVE-FOLD CROSS VALIDATION

Fold	Participant folders (HC / PD)	Recordings (HC / PD)	Total Recordings
1	7 (3 / 4)	22 / 62	84
2	11 (4 / 7)	38 / 103	141
3	13 (8 / 5)	89 / 67	156
4	13 (6 / 7)	44 / 103	147
5	9 (6 / 3)	57 / 46	103
Total	**53 (27 / 26)**	**250 / 381**	**631**

B. *Preprocessing and Feature Extraction*

The raw speech recordings are preprocessed to improve audio clarity and allow feature extraction to be performed in a consistent manner. The audio recordings are resampled at 16 kHz before being processed through silence elimination, frequency-domain noise suppression, speech segmentation, and amplitude scaling which produces a consistent speech signal for further analysis. A 16-dimensional acoustic feature vector is extracted from the preprocessed recordings consisting of 13 Mel-frequency Cepstral Coefficients (MFCCs), jitter, shimmer and HNR. MFCCs quantify the acoustic characteristics associated with the vocal tract, while jitter, shimmer and HNR estimate variability in fundamental frequency and voice quality measures commonly linked to Parkinsonian speech. Recordings with undefined or missing feature values were excluded during initial data screening. Feature extraction was then performed on the selected 631 recordings, and the final feature set was verified to contain no missing values. All preprocessing steps were applied independently in each training fold to avoid data leakage since the evaluation uses participant-grouped five-fold StratifiedGroupKFold cross-validation. The minimum and maximum values used for min–max normalization were estimated from the training samples of each individual fold only and these values were subsequently used to transform the associated test samples. Each feature was scaled to the range [0, 1] before Cellular Automata thresholding. This fold-wise procedure prevents information from the test fold from entering the feature normalization.

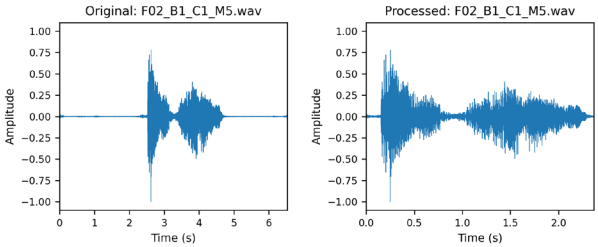

Fig. 4. Signal before and after preprocessing.

An example of the speech signal before and after pre-processing is shown in Fig. 4, which validates that we not only removed silence and background noise but also retained clinically relevant speech characteristics needed for acoustic feature computation.

C. *Feature Extraction Techniques*

Acoustic features are extracted from the preprocessed speech signal and represented numerically. These features are subsequently normalized and thresholded to initialize the Cellular Automata grid. These characteristics are used as input to identify Parkinson's disease early on. The speech feature extraction procedure involves dividing the audio signal into frames, using the window function to reduce spectral leakage.

1) *MFCC*

Mel-Frequency Cepstral Coefficients (MFCCs) are commonly employed to analyze speech signals by describing their acoustic properties. MFCC extraction involves applying the Fourier transform to a windowed speech frame to obtain its frequency spectrum, projecting

the power spectrum onto Mel-scale filter banks, calculating the logarithm values of the filter-bank energies, and performing a Discrete Cosine Transform (DCT) to obtain the cepstral coefficients. In this study, 13 MFCC coefficients are extracted to represent the acoustic characteristics of the speech signal. These coefficients capture information related to the spectral envelope and characteristics of the vocal tract properties. The lower-order coefficients generally represent broad spectral characteristics, while the higher-order coefficients capture finer variations in the spectral envelope. Together, the 13 coefficients provide a compact representation of speech characteristics for subsequent analysis and classification.

Computing MFCC with pre-processed raw audio signal

$$y[n] = x[n] - \alpha x[n-1] \quad (1)$$

In Eq. (1), $x[n]$ denotes the input speech signal at sample index n , $x[n-1]$ is the input signal at the preceding sample, α is the pre-emphasis coefficient and $y[n]$ is the resulting pre-emphasized signal. The equation represents a first-difference filter that emphasizes sample-to-sample changes in the signal. It is an output signal, and it is mainly used for noise removal and detecting the changes that are in the raw audio signal. Segmentation divides the continuous audio signal into short, overlapping frames within which the signal can be considered quasi-stationary. The frame duration is typically 20–40 ms.

$$N = 1 + \left\lfloor \frac{L-F}{S} \right\rfloor \quad (2)$$

In Eq. (2), N is total number of complete frames, L is the number of audio samples in the signal, and F is frame length and S is frame shift (number of samples between consecutive frames / hop size). The floor symbol means rounding down to the nearest integer.

Applying the Hamming window function to avoid spectral leakage of the audio signal.

$$W(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1 \quad (3)$$

As per Eq. (3), $W(n)$ is the window value at sample position n, and N is the total sample count with in frame (window length). A spectrogram illustrates how the frequency distribution changes in the frequency components of a speech signal across time. In this study, the speech signal is segmented into consecutive, partially overlapping segments, and the Fast Fourier Transform (FFT) operation is performed on each frame to obtain its frequency components. The obtained spectral magnitudes are used to describe the time-varying frequency content within the speech waveform.

$$C_m = DCT_m\left(S_{Mel,dB}\right), \quad m = 0,1,\ldots,12 \quad (4)$$

In Eq. (4) $S_{Mel,dB}$ denotes the log-scaled Mel power spectrogram expressed in decibels and DCT_m denotes the $m-th$ coefficient of the discrete cosine transform and C_m is the resulting MFCC coefficient. The MFCCs are obtained by applying the DCT to the log-scaled Mel spectral representation. In this study, 13 MFCC coefficients were extracted using the Librosa library and averaged over time to obtain a fixed-length feature representation for each speech recording.

In Fig. 5, the pipeline of MFCC extraction shows how the features are extracted with the help of the hamming window function after framing the signal, then applying the FFT, which converts the time domain into the frequency domain. The Mel filter bank applies perceptually motivated Mel-scale filtering to the frequency spectrum. Then the Discrete Cosine Transform is applied to obtain the MFCC, and finally it is represented in the form of a feature vector or matrix.

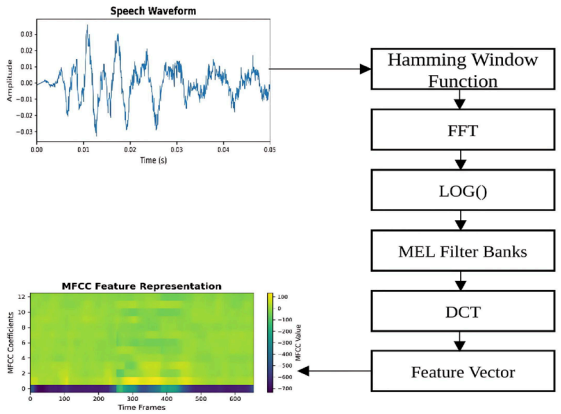

Fig. 5. Procedure for extracting MFCC from pre-processed signal.

2) Jitter

Jitter measures cycle-to-cycle variations in the fundamental period of a voiced speech signal. Jitter focuses on random changes in pitch and frequency. Jitter measures the variations in fundamental frequency (f0). Jitter shows the variations, particularly in sustained vowels, and mainly focuses on the stability of vocal fold vibrations, which are used primarily in detecting Parkinson's disease.

$$jitter_{local} = \frac{1}{N-1}\sum_{i=1}^{N-1}\frac{|T_i - T_{i+1}|}{(T_i + T_{i+1})/2} \quad (5)$$

As in Eq. (5), where T_i is the duration of the ith pitch period, and N is the Number of pitch periods, it measures how the pitch period changes between consecutive cycles. In Fig. 6, the waveform appeared after the extraction of jitter from the pre-processed speech sample.

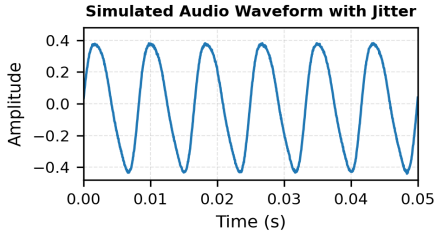

Fig. 6. Speech signal extracted from Jitter.

3) Shimmer

Shimmer measures cycle-to-cycle variations in the amplitude of a voiced speech signal and quantifies the relative differences between consecutive pitch-cycle amplitudes. In Parkinson's disease detection, Shimmer can help characterize voice irregularities associated with altered phonation. Higher Shimmer values may indicate increased amplitude perturbation in the voice. In Eq. (6), A_i and A_{i+1} represent the peak amplitudes of two consecutive pitch cycles, respectively. The Shimmer measure is expressed as a relative amplitude variation and is dimensionless in the formulation used here.

$$Shimmer_{local} = \frac{1}{N-1}\sum_{i=1}^{N-1}\frac{|A_{i+1}-A_i|}{(A_{i+1}+A_i)/2} \quad (6)$$

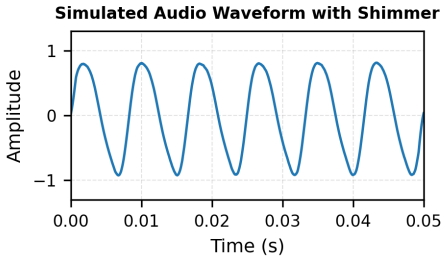

Fig. 7. Speech signal extracted from shimmer.

Fig. 7 illustrates a speech signal impacted by shimmer, where the waveform's amplitude fluctuates erratically over time. These differences point to decreased control over volume and instability in vocal fold vibration, two common signs of disordered speech.

4) Harmonic to noise ratio

Harmonic-to-Noise Ratio (HNR) measures the relative strength of periodic harmonic components compared with aperiodic noise in a voiced speech signal. It provides information about voice quality and phonatory stability. Lower HNR values may indicate increased noise or irregularity in phonation, whereas higher values generally indicate a more periodic voice signal. HNR is used as an acoustic feature for distinguishing speech characteristics between Parkinson's disease and healthy control recordings. The Harmonics-to-Noise Ratio (HNR) measures the relative strength of periodic harmonic components compared with aperiodic noise in a voice signal. In detecting Parkinson's disease, the HNR is mainly used to detect the noise irregularities. If HNR is high, the voice is stable; if HNR is low, the voice becomes unstable.

5) Representation of extracted speech features

The preprocessed speech recordings are represented as a 16-dimensional acoustic feature vector comprising 13 MFCCs together with jitter, shimmer, and HNR. Each recording corresponds to one row in the feature matrix, while each column represents an extracted acoustic feature. The class labels are encoded as 0 for Healthy Control (HC) and 1 for Parkinson's Disease (PD). The extracted acoustic features are combined into a single feature vector as follows

Feature Vector = [MFCC1, MFCC2, MFCC3…., MFCC13, jitter, shimmer, HNR]

This feature vector provides a compact numerical representation that combines spectral and voice-quality characteristics for Parkinson's disease classification. The extracted features preserve complementary information about vocal tract configuration, phonatory stability, and voice quality, making them suitable for subsequent Cellular Automata encoding and Deep Neural Network classification.

The full feature extraction pipeline is visualised in Fig. 8. Initially the input speech signal is preprocessed to filter noise and normalize the audio recording. Then, features including MFCC, jitter, shimmer and HNR are generated and indicate final 16-dimensional acoustic feature vectors of utterance used on a classification.

Extracted acoustic features have been associated with clinically relevant symptoms of Parkinson's disease. Jitter measures fundamental frequency cycle-to-cycle variations, meaning that it is an indicator of voice instability. Shimmer reflects amplitude perturbations (imperfect control over the vibration of vocal folds). HNR assesses the relative periodicity and aperiodicity of speech, such as weak or strong measures of breathiness and hoarseness in which lower HNR increases indicate increased breathiness or hoarseness, features seen within Parkinsonian Speech. MFCC features describe vocal tract spectral characteristics and are able to capture articulation and phonation changes associated with abnormal speech production. Collectively, these acoustic features provide complementary information that may help distinguish speech recordings from individuals with Parkinson's disease and healthy controls.

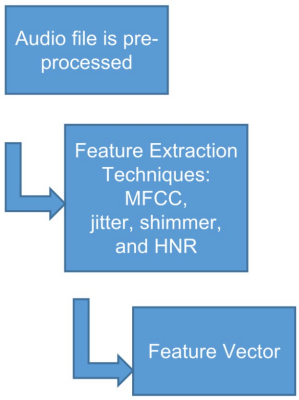

Fig. 8. Pipeline for feature extraction.

D. Feature Mapping and Threshold Selection for 1D Cellular Automata

The feature vector is 16-dimensional, which includes 13 MFCC coefficients, jitter, shimmer, and HNR, and that is first normalized to the range [0,1] by using the min-max normalization function. Each value is normalized with x_i^{norm} and, then binarised using a fixed threshold $\theta = 0.6$ to produce a binary vector $b \in \{0,1\}$

$$\begin{cases} 1, if\ x_i^{norm} \geq 0.6 \\ 0\ , otherwise \end{cases} \quad (7)$$

As shown in Eq. (7) x_i^{norm} represents the normalized value of the $i-th$ feature. Features with normalized values greater than or equal to 0.6 are assigned a binary value of 1, whereas features with normalized values below 0.6 are assigned a binary value of 0. Continuous acoustic features are converted into binary states by this thresholding process, which represents relatively high and low normalized feature value states. This resulting binary vector is then used as the starting point for Rule 110 Cellular Automata evolution, allowing for interaction between adjacent feature states through neighborhood-based interactions, and providing a structured representation for subsequent Deep Neural Network (DNN) classification.

The resulting 16-dimensional binary feature vector is mapped to a one-dimensional Cellular Automata (1D CA) grid of length$L = 20$. The first 16 cells are initialized with the binary states corresponding to the extracted acoustic features, while the remaining four cells are assigned zero to complete the 20-cell grid. The resulting initial configuration is then evolved using Rule 110 for $T = 20$ iterations. The state update of each cell is defined by Eq. (8).

$$S_i^{t+1} = f_{Rule110}(S_{i-1,}^t S_i^t, S_{i+1}^t) \quad (8)$$

Here, S_i^{t+1} denotes the state of the ith cell at iteration $t+1$, whileS_{i-1}^t,S_i^t and S_{i+1}^t represent the states of the left neighboring cell, the current cell, and the right neighboring cell, respectively, at iteration t . The function $f_{Rule}(.)$ applies the deterministic Rule-110 transition rule to update each cell based on the states of its three-cell neighborhood. Periodic boundary conditions are used in the one-dimensional Cellular Automaton. Under this condition, the grid wraps around at its boundaries: the last cell is treated as the left neighbor of the first cell, and the first cell is treated as the right neighbor of the last cell. Accordingly, Rule 110 updates the cells using their left, current, and right neighboring states, including at the boundaries. Both the first and last cells are updated using Rule 110 at every evolution step, with their neighboring states determined by the periodic boundary condition. By iteratively applying this update rule for 20 evolution steps, the proposed Cellular Automata performs a neighborhood-aware nonlinear feature transformation, producing a structured representation that preserves local relationships among adjacent acoustic speech features for subsequent classification. The proposed Cellular Automata does not generate new information from the original acoustic features. Instead, Rule 110 applies a deterministic, neighbourhood-based transformation in which each cell state is updated according to its neighbouring states. This process produces a CA-enhanced representation of the original acoustic features for subsequent classification.

The proposed framework operates on recording-level acoustic feature vectors extracted from each speech recording. Accordingly, the proposed Cellular Automata performs an iterative neighborhood-based nonlinear transformation of the extracted acoustic features rather than explicitly modelling the temporal evolution of the original speech waveform.

1) Feature-wise CA state aggregation and CA feature aggregation

After Rule-110 evolution, the 1D Cellular Automata generates a 20×20 CA state evolution matrix containing 400 binary cell values. Instead of directly using the entire evolution matrix for classification, feature-wise temporal pooling is applied to obtain a compact representation. The first 16 CA cells correspond to the mapped acoustic speech features (13 MFCC coefficients, jitter, shimmer, and HNR), while the remaining 4 cells are zero-padding cells introduced only to preserve neighborhood continuity during CA evolution. For each feature cell, the activation values are averaged over the 20 evolution steps as

$$\bar{S}_i = \frac{1}{T}\sum_{t=1}^{T} S_i^{(t)},\ T = 20$$

where $S_i^{(t)}$ is the state of cell i at evolution step (t), T=20 represents the total number of evolution steps and $\bar{S}_i$ is the average activation of cell i across those steps. The average activation values of the first 16 cells are retained as the CA-enhanced feature vector, while the four zero-padding cells are discarded. This operation reduces the 20×20 evolution matrix (400 cell values) to a 16-dimensional CA-enhanced feature vector. The four padding cells are discarded after pooling because they do not represent acoustic speech features. The resulting 16-dimensional feature vector is then used as the input to the Deep Neural Network.

2) Parameters

Based on the results of the sensitivity analysis, the threshold (θ = 0.6), (L = 20) and (T = 20) were chosen. There was a variation in the average classification accuracy for each of the different threshold evaluated. The highest average classification accuracy occurred with (θ = 0.6) with an average classification accuracy of approximately 95.18 percent; however, the lower classification accuracies were achieved with all of the other threshold evaluated. These findings indicate that the selection of the threshold can affect how features are represented using cellular automata and ultimately the classification performance. Therefore, (θ = 0.6) was selected as the preferred threshold from those evaluated.

3) Overall workflow of the proposed CA encoding framework

Fig. 9 shows the whole process of using the proposed one-dimensional Cellular Automata (1DCA) encoding method. A 16-dimensional normalized acoustic feature vector is converted into a 20-cell 1DCA grid. This is accomplished by placing the first 16 cells as the acoustic features and then adding four more cells of zeros. The CA is evolved using Rule-110 for 20 steps to produce a

20×20 state matrix. Feature-wise CA state aggregation is then applied to the first 16 feature cells to obtain a compact 16-dimensional CA-enhanced feature vector, which is subsequently used as the input to the Deep Neural Network (DNN) classifier for Parkinson's disease classification.

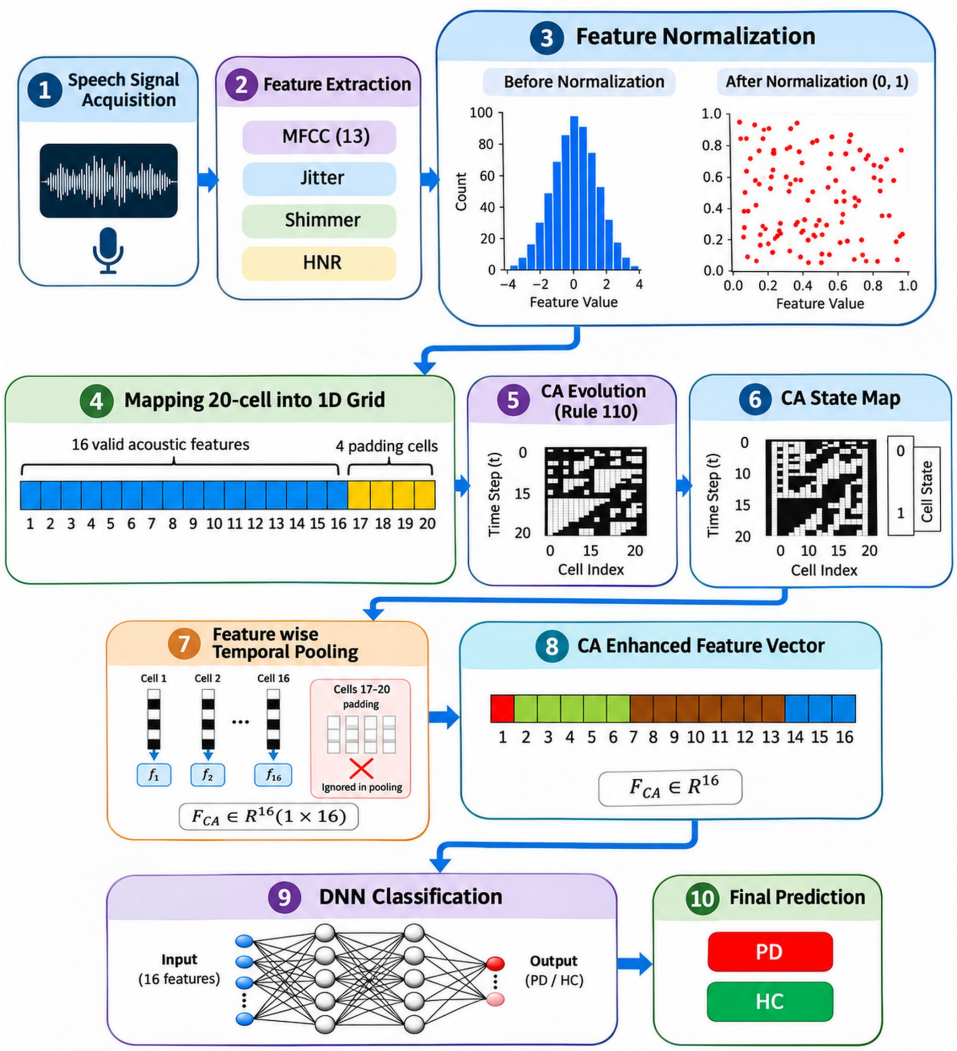

Fig. 9. Workflow of the proposed 1D cellular automata encoding framework.

4) Sensitivity to CA parameters

In Fig. 10, there is an evaluation of the sensitivity of the Cellular Automata (CA) to the threshold value. The vertical axis shows the mean classification accuracy (%), and each of the tested threshold values was evaluated using the same five fixed folds and the same evaluation process used in the primary study. There are three primary cellular automata parameters that can affect performance: the threshold value, the grid size, and the number of evolutionary cycles. As shown in Fig. 10, the mean accuracy is 92.54% at $\theta = 0.3$ and decreases to 89.78% at $\theta = 0.4$. It then increases to 94.01% at $\theta = 0.5$ and reaches its highest value of 95.18% at $\theta = 0.6$. When the threshold is increased to $\theta = 0.7$, the mean accuracy decreases to 90.09%. These results indicate that threshold selection influences the CA-based feature representation and the resulting classification performance. The main experimental evaluation yielded a mean classification accuracy of 95.14%, whereas the threshold sensitivity analysis yielded 95.18% at θ = 0.6, a difference of 0.04 percentage points. Both evaluations used the same five fixed folds and evaluation procedure; however, the deep-learning model was retrained for the sensitivity analysis, so the two values represent results from separate training runs. Therefore, θ = 0.6 was selected as the preferred threshold because it achieved the highest mean accuracy among the evaluated threshold settings.

Fig. 11 presents the sensitivity analysis of the number of Cellular Automata (CA) evolution steps. The mean classification accuracies obtained at 20, 30, 40, 50, and 60 evolution steps were 95.54%, 93.26%, 88.92%, 85.27%, and 92.11%, respectively. Classification accuracy decreased as the number of evolution steps increased from 20 to 50, before recovering at 60 steps. The highest accuracy of 95.54% was observed at 20 steps among the evaluated configurations. The main experimental evaluation yielded a mean classification accuracy of 95.14%, whereas the separate CA evolution-step sensitivity analysis yielded an accuracy of 95.54% at 20 steps. The evaluation of each of these approaches involved

the same five fixed folds and evaluation procedure. However, for the sensitivity analysis, the deep-learning model was retrained. This resulted in the two values representing different training runs. The difference between the two values is 0.40 percentage points. Based on the classification performance that was observed and the reduced computational expense that results from having fewer iterations, 20 CA evolution steps were chosen as the most suitable configuration.

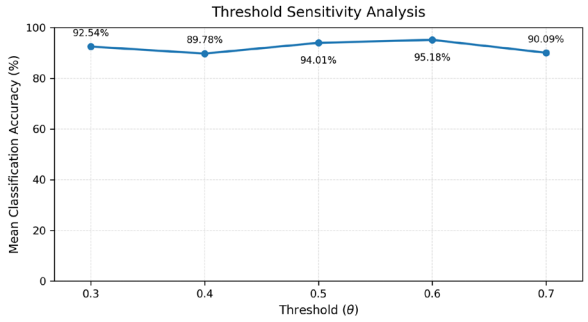

Fig. 10. Sensitivity analysis of the CA threshold parameter.

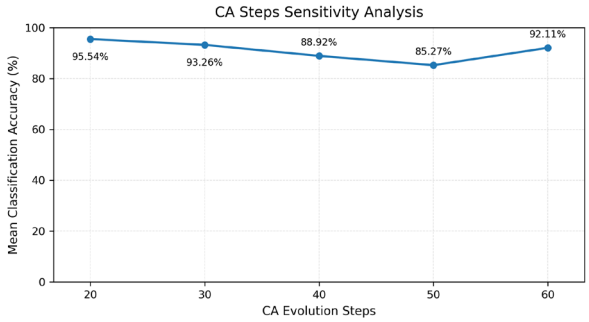

Fig. 11. Sensitivity analysis of the number of CA evolution steps.

The sensitivity analysis on Cellular Automata (CA) grid size is shown in Fig. 12. It includes evaluation of four different configurations: 20, 30, 40, and 50. The y-axis shows the average classification accuracy (%); all four configurations were evaluated under the same five fixed folds and evaluation procedure as the main experiment. The corresponding average classification accuracy was 95.12%, 93.36%, 90.86%, and 89.06%, respectively. There was a trend toward lower classification accuracy as grid size increased. This was most evident when comparing grid size 30 to grid size 40, whereas the decline from grid size 40 to grid size 50 was less pronounced. Grid size 20 resulted in the best classification accuracy among the evaluated configurations. In fact, the main experimental evaluation resulted in an average classification accuracy of 95.14%, whereas the grid-size sensitivity analysis produced an average classification accuracy of 95.12% with a grid size of 20. This indicates that the classification accuracy produced by the grid-size sensitivity analysis is slightly below that produced by the main experimental evaluation by 0.02 percentage points. While both evaluations used the same five fixed folds and evaluation procedure, the sensitivity analysis involved retraining the DNN. Figs. 10–12 display average classification accuracy (%) across the same five participant-grouped folds as used in the main evaluation. Each sensitivity analysis evaluated the specific CA parameters under the same fold assignments and evaluation procedure as the main evaluation. The DNN was retrained for each of the specific sensitivity analyses. Sensitivity analysis results are displayed separately from the main experiment's average classification accuracy of 95.14%.

The CA parameters (θ, number of evolution steps, and grid size) were selected using the same five participant-grouped folds on which performance is reported; the reported accuracy may therefore be slightly optimistic, and a nested cross-validation or an independent test set would be required for an unbiased estimate.

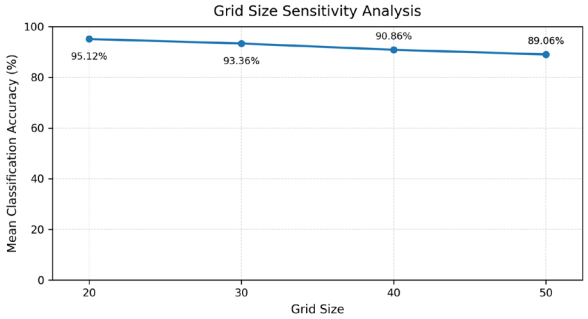

Fig. 12. Sensitivity analysis of CA grid size.

5) Evolution of 1D cellular automata patterns for healthy and parkinsonian speech features

Fig. 13 shows the evolution of 1D CA patterns produced from speech feature vectors for people with PD and HC using three distinct CA rules—Rule 30, Rule 90, and Rule 110.

The vertical axis in each subplot shows the evolution of the CA over subsequent time steps, while the horizontal axis shows the speech feature cells. Black and white pixels, which represent active (1) and inactive (0) states, respectively, are used to show binary cell states. In contrast to HC, PD samples show denser and more disordered structures, demonstrating more variability in diseased speech. Rule 30 results in extremely chaotic and irregular patterns.

Fig. 13 illustrates the evolved CA patterns obtained from speech feature representations of Parkinson's Disease (PD) and Healthy Control (HC) samples using Rules 30, 90, and 110. The black and white cells represent binary states of 1 and 0, respectively. Differences in the visual patterns suggest that CA evolution can produce distinct representations of the extracted acoustic features. However, these visual differences alone do not establish specific clinical symptoms or a causal relationship with Parkinson's disease. The classification experiments are therefore used to assess the usefulness of the CA-transformed representations for distinguishing PD from HC samples.

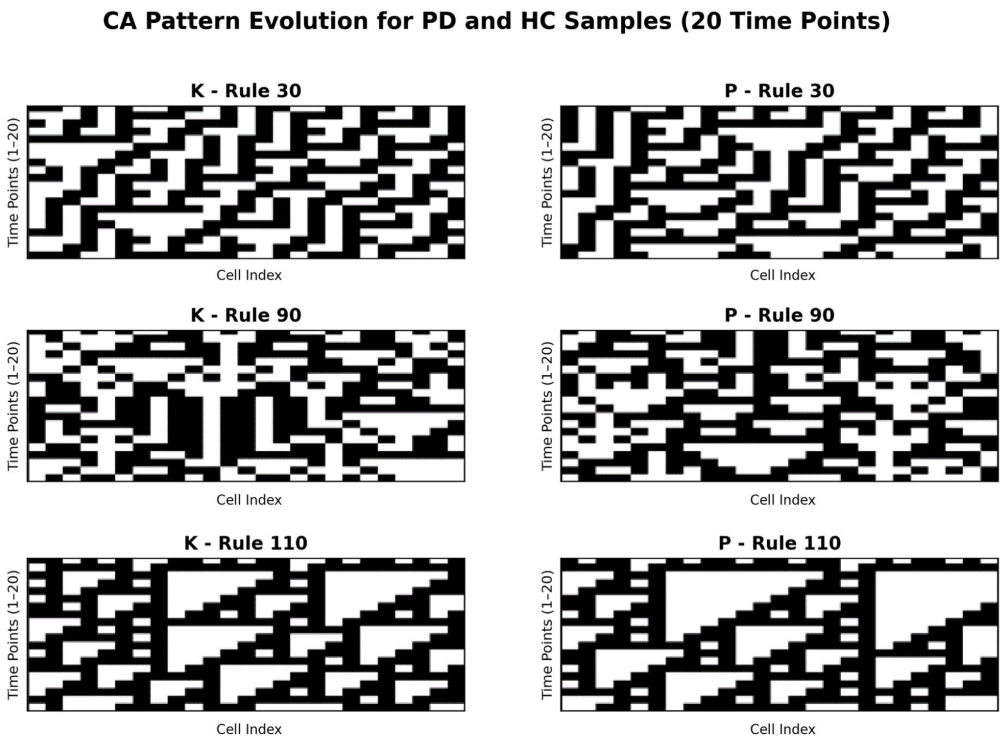

Fig. 13. 1DCA evolution patterns for healthy and Parkinson's disease.

6) Ablation study on CA rule selection

To justify the selection of the Cellular Automata (CA) rule, an ablation study was conducted using three representative Elementary Cellular Automata (ECA) rules: Rule 30, Rule 90, and Rule 110. These rules were selected because they are among the most extensively studied ECA rules and exhibit distinctly different evolution behaviors. Rule 30 produces chaotic and pseudo-random patterns, Rule 90 generates regular self-similar structures, whereas Rule 110 exhibits complex and computational universality. The same Deep Neural Network (DNN) architecture and identical experimental settings (θ = 0.6, grid size = 20, and 20 iterations) were used for all three rules. The corresponding classification performance is presented in Table III.

TABLE III. PERFORMANCE COMPARISON OF DIFFERENT CA RULES ON THE ITALIAN PARKINSON'S VOICE AND SPEECH (IPVS) DATASET

1D CA Rule	Accuracy (%)	Recall (%)
Rule 30	89.2	88.9
Rule 90	91.7	91.1
Rule 110	**95.14**	**94.37**

In terms of classifying correctly, Rule 110 outperformed the other two tested cellular automata rules as shown in Table III. Rule 110 had an accuracy rate of 95.14% and a recall rate of 94.37%. Rule 30 had an accuracy rate of 89.2% and Rule 90 had an accuracy rate of 91.7%. The same five participant-grouped StratifiedGroupKFold cross-validation folds were used to assess the performance of each of the three CA rules. Additionally, the same evaluation protocols were applied to each of the three CA rules. There was no variation in the DNN architecture or the experimental settings; only the CA rule was changed. Accuracy and recall were reported as the average performance across all five folds. The above findings indicate that Rule 110 produced improved classification results over both Rules 30 and 90, as demonstrated by our evaluation of these rules under the specific experimental conditions. Rule 30 is characterized by chaotic and/or pseudo-randomly evolving patterns, while Rule 90 produces regular, self-similar patterns. Rule 110 produces even more complex patterns of evolution. The variations we observed in the classification metrics demonstrate that the selection of a particular CA rule will affect how well the resulting feature representation can be utilized. Our findings do not demonstrate that Rule 110 suppresses irrelevant variation or preserves relevant clinical relationships between acoustic features. There are a total of 256 elementary Cellular Automata (CA) rules, but only three particular ones were examined in this study: Rules 30, 90, and 110- chosen as they represent three well-characterised but complementary evolution behaviours. Other elementary cellular automaton rules, such as Rule 182 [23], were not evaluated because a comprehensive comparison of all 256 possible rules was beyond the scope of this study. Evaluating additional rules will be considered in future work. Final Framework Selection: Based on the comparative results in Table III, Rule 110 was selected for the proposed framework because it achieved the highest accuracy and recall, among the three evaluated CA rules. This selection is based on the observed classification metrics under the experimental settings used in this study

E. Deep Neural Network Classifier

For binary PD/HC classification, the 16-dimensional CA-enhanced feature vector obtained through feature-wise temporal pooling is used as the input to the Deep Neural

Network. Although the Cellular Automata generates a 20×20 CA evolution matrix, the matrix is first transformed into a compact 16-dimensional representation using the pooling strategy described in Section III-D before classification. The network architecture includes an input layer of 16 neurons, a first hidden layer with 64 ReLU activated neurons, followed by a dropout layer of rate 0.3, second hidden layer with 32 ReLU activated neurons, and the second hidden layer is followed by a dropout layer with a rate selected from 0.1 to 0.4 and a sigmoid output neuron that estimates the probability of Parkinson's disease.

The Adam optimizer with a learning rate of 0.001 and binary cross-entropy loss was used to train the model for 20 epochs. To lessen the imbalance between PD and HC samples, class weights were applied during training and a batch size of 32 was used. Its lightweight architecture and computational efficiency make it a promising candidate for future speech-based Parkinson's disease screening systems. However, additional validation on larger multi-center clinical datasets is required before real-world deployment. The feature vector transformed using the Cellular Automata (CA) approach is subsequently fed into a Deep Neural Network (DNN) for Parkinson's Disease (PD) classification. The DNN learns the relationship between the transformed feature representations and the corresponding class labels through multiple hidden layers, enabling the classification of speech samples into the respective categories.

Input vector features can be written as

$$X = [x_1, x_2, x_3 \ldots\ldots x_{16}] \tag{9}$$

In Eq. (9), X denotes the 16-dimensional CA-transformed feature vector (one averaged CA activation per original feature: 13 MFCCs, jitter, shimmer and HNR).

The output of the first hidden layer is computed as

$$H_1 = RELU(W_1 X + b_1) \tag{10}$$

In Eq. (10) W_1 represents the weight matrix of the first hidden layer, b_1 denotes the bias vector, and H_1 is the hidden layer output. The ReLU activation function is applied element-wise to the weighted input and bias. The rectified Linear Unit (ReLU) activation function is defined as

$$RELU(z) = \max(0, z) \tag{11}$$

In Eq. (11) z denoted the weighted sum of inputs, ReLU introduces nonlinearity into the network and improves learning efficiency.

The final output probability of the Parkinson's disease is obtained using the sigmoid activation function

$$\hat{y} = \frac{1}{1+e^{-(W_0 H_2 + b_0)}} \tag{12}$$

In Eq. (12) $\hat{y}$ represents the predicted probability of Parkinson's disease w_0 denotes the output layer weight vector, b_0 is the output bias, and H_2 represents the output of the second hidden layer,

The network parameters are optimized using the binary cross-entropy loss function given by

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{13}$$

In Eq. (13) L denotes the average binary cross-entropy loss N represents the total number of training samples included in the loss calculation, y_i is the actual class label (0 or 1), and $\hat{y}_i$ is the predicted probability of the positive class for the $i-th$ sample. The objective of the optimization process is to minimize the loss function and improve classification accuracy.

Algorithm: 1DCA Enhanced Speech Feature Extraction and Parkinson's disease Classification

Input: Speech Signal $x(t)$

Output: Predicted class label $\hat{y} \in \{PD, HC\}$

1. Preprocess the speech signal and extract acoustic features (MFCC, Jitter, Shimmer, HNR):
2. $F = [MFCC_1, .., MFCC_{13}, Jitter, Shimmer, HNR]$
3. Normalize the extracted feature vector

$$F_{n,i} = \frac{F_i - F_{i,min}}{F_{i,max} - F_{i,min}}$$

4. Convert the normalized feature vector into binary Cellular Automata states using the threshold θ=0.6:

$$C_i^0 \begin{cases} 1, & F_{n,i} \geq \theta \\ 0, & Otherwise \end{cases}$$

5. Apply 1D Cellular Automata evolution using Rule-110 for *T=20* iterations:

$$C_i^{t+1} = f_{Rule110}(C_{i-1}^t, C_i^t, C_{i+1}^t)$$

6. Generate evolved CA feature representation:

$$C = \{C^1, C^2, \ldots, C^T\}$$

7. Apply feature-wise temporal pooling to obtain the 16-dimensional CA-enhanced feature vector:

$$F_{CA} = Pool(C)$$

$$F_{CA} \in R^{16}$$

8. Compute the Parkinson's disease probability using the sigmoid activation function.

$$\hat{y} = \frac{1}{1 + e^{-(w_0 H_2 + b_0)}}$$

9. Output predicted class label:
   - Parkinson's Disease (PD) or
   - Healthy Control (HC)

The proposed framework was evaluated using five-fold StratifiedGroupKFold cross-validation with participant-level grouping and stratification to maintain class representation across folds. All recordings belonging to the same participant were assigned exclusively to a single fold, ensuring that no participant appeared in both the training and testing sets within any fold. This evaluation protocol was designed to assess the model's ability to generalize to participants whose recordings were excluded from the corresponding training fold. The evaluation metrics included accuracy, recall (sensitivity), specificity, balanced accuracy, the Area under the

Receiver Operating Characteristic-Area Under the Curve (ROC-AUC), and the confusion matrix. The evaluation metrics were averaged across the five folds to provide an overall summary of the model's performance. In addition, 95% Wilson confidence intervals were calculated for the proposed model's pooled out-of-fold accuracy, sensitivity, and specificity using the corresponding confusion-matrix counts. These confidence intervals quantify uncertainty in the observed recording-level proportions and should be interpreted in the context of the participant-grouped evaluation protocol, as they do not explicitly account for dependence between recordings from the same participant.

IV. Results and Analysis

All experiments reported in this section were conducted using the selected subset of 631 recordings from the Italian Parkinson's Voice and Speech (IPVS) dataset. Model performance was evaluated using participant-grouped five-fold StratifiedGroupKFold cross-validation, as described in Section III. All recordings belonging to the same participant were assigned to a single fold, ensuring that no participant appeared in both the training and test sets within any fold. The following sections present the baseline comparisons, classification performance, feature importance, computational efficiency, and clinical relevance of the proposed one-dimensional Cellular Automata (1D CA)-based framework.

A. Baseline Comparison

Recent developments in the analysis of speech in PD have made use of advanced deep learning architectures, such as transformer-based encoders and self-supervised speech models (e.g., Wav2Vec2), which learn directly from raw waveforms or spectrograms to achieve state-of-the-art accuracy. However, these models frequently have poor interpretability, high computational resource requirements, and enormous data sets—factors that can prevent clinical application in settings with limited resources.

By comparing established baseline classifiers, including Support Vector Machines (SVM), Gaussian Mixture Models (GMM), and Convolutional Neural Networks (CNN), we assess their classification performance relative to the proposed ALL+1DCA+DNN framework. The baseline classifiers were evaluated on the Italian Parkinson's Voice and Speech (IPVS) dataset using the pre-processing and evaluation protocol described in Section III. The input features differed across the baseline classifiers: SVM used jitter and shimmer, whereas GMM and CNN used MFCCs. The DNN without CA used all 16 acoustic features. The results are summarized in Table IV.

TABLE IV. Baseline Comparison of Models

Features	Classifier	Accuracy
Jitter, Shimmer	SVM	91.4%
MFCC	GMM	89.0%
MFCC	CNN	92.3%
All 16 features	DNN (without CA)	90.01%
Proposed	ALL+1DCA+DNN	95.14%

Note: Italian Parkinson's Voice and Speech (IPVS) Cellular automata (CA)

B. Comparison of Model Configurations

Table V presents the classification accuracies of the four evaluated model configurations under the participant-grouped five-fold cross-validation protocol described in Section III. The MFCC+DNN baseline achieved an accuracy of 85.34%, while incorporating the proposed one-dimensional Cellular Automata (1D CA) component into the MFCC-based classifier resulted in an accuracy of 87.15%. Using all acoustic features without 1D CA achieved an accuracy of 90.01%, whereas the proposed ALL+1DCA+DNN framework achieved an accuracy of 95.14%. The results demonstrate higher observed classification accuracy for the configurations incorporating the proposed 1D CA component. However, these findings are descriptive and should not be interpreted as evidence of statistically significant differences between the evaluated configurations.

TABLE V. Classification Accuracy of the Model Configurations

S.NO	Model	Accuracy
0	MFCC+DNN	85.34%
1	MFCC+1DCA+DNN	87.15%
2	ALL+DNN	90.01%
3	ALL+1DCA+DNN	95.14%

Note: Deep Neural Network (DNN)

The improvement provided by the proposed ALL+1DCA+DNN framework is due to the use of Rule-110 cellular automata to preserve neighborhood relationships between neighboring acoustic features and create non-linear interaction among those features. The Rule-110 cellular automata transformation allows for a more discriminative representation of MFCC, jitter, shimmer, and HNR features that allow the DNN classifier to more effectively differentiate between the speech of individuals with Parkinson's disease and that of healthy controls.

TABLE VI. Wilson 95% Confidence Intervals for the Proposed Model

Metric	Correct predictions / Total	Observed Value (%)	95% Wilson confidence interval (%)
Accuracy	601/631	95.25	93.29–96.65
Sensitivity (PD)	363/381	95.28	92.66–96.99
Specificity (HC)	238/250	95.20	91.80–97.23

Table VI provides the 95% Wilson confidence intervals for the proposed ALL+1DCA+DNN framework as determined by combining all of the predictions obtained from the various out-of-fold evaluations. The classification accuracy achieved was 95.25% with a 95% Wilson confidence interval ranging from 93.29% to 96.65%. For the detection of PD recordings, the sensitivity of the model was 95.28% and for the detection of HC recordings, the specificity of the model was 95.20%. While these confidence intervals provide a measure of the uncertainty surrounding the proportion of correctly classified recordings, they do not provide evidence that

one model configuration performs better than another, nor do they fully account for the correlation that exists among recordings obtained from the same individual.

Table VII presents the mean performance metrics obtained across the five folds of StratifiedGroupKFold cross-validation. The proposed framework achieved a mean accuracy of 95.14% and recall of 94.37%. The mean specificity and balanced accuracy were 95.63% and 95.00%, respectively, while the mean ROC-AUC was 0.9895. These results summarize the model's average fold-level performance. The pooled accuracy of 95.25%, calculated from the aggregated confusion matrix across all 631 recordings, differs slightly from the mean fold-level accuracy because it is calculated from the combined out-of-fold predictions rather than by averaging the five fold-level accuracy values.

TABLE VII. PERFORMANCE METRICS

S.NO	Metric	MEAN
0	Accuracy	95.14%
1	Recall	94.37%
2	Specificity	95.63%
3	ROC-AUC	0.9895
4	Balanced Accuracy	95.00%

Note: Receiver Operating Characteristic-Area Under the Curve (ROC-AUC)

The cross-validation results and confidence intervals provide information about the performance of the proposed CA-enhanced model on the selected dataset. These results are specific to the evaluated dataset and evaluation protocol and are not directly comparable with accuracies reported on other datasets.

C. Feature Importance and Interpretability

To better understand the contribution of the extracted acoustic features to the classification process, permutation feature importance analysis was performed. Fig. 14 illustrates the relative importance of the CA-enhanced acoustic features and their contribution to the Deep Neural Network (DNN) predictions. This analysis gives an understanding of the classification performance of the proposed framework influenced by individual acoustic features.

The analysis further reveals that MFCC features, especially MFCC_1 are the most important among all other features and can capture spectral distortions caused by Parkinsonian speech. Other acoustic measures, such as jitter, shimmer, and HNR, which reflects vocal instability, amplitude perturbations and increased noise components that are typical of Parkinson's speech also contribute to the classification. These various complementary acoustic characteristics in tandem result in better discernments between speech produced by patients with PD and healthy controls.

The proposed CA-enhanced feature representation captures local relationships among acoustic features through neighbourhood-based nonlinear transformations, providing structured inputs for the DNN classifier. The importance scores indicate which acoustic features most influenced the model's predictions on this dataset; they are not evidence of a clinical mechanism.

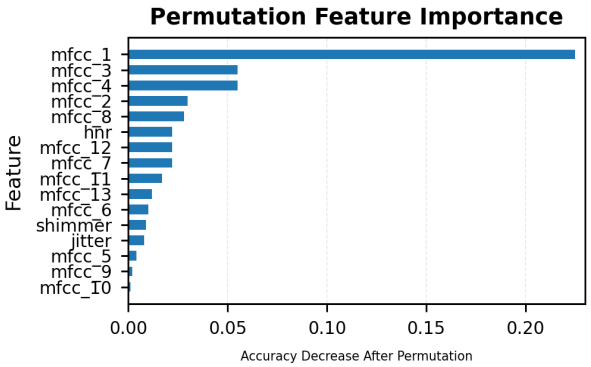

Fig. 14. Feature importance.

D. Computational Efficiency

In order to assess the effectiveness of our proposed framework from a computational standpoint, we have recorded the elapsed time required for each stage: 1D CA transformation, DNN training, DNN inference, and permutation feature importance analysis. The relationship between these stages is shown in Fig. 15. All experiments were performed on a standard computer system with an Intel Core i7 processor and 16GB of memory. No Graphics Processing Unit (GPU) acceleration was utilized.

Fig. 15 shows that the 1D CA transform required 0.21 s; the DNN prediction took 0.27 s; DNN training (20 sequential epochs of gradient-based optimization) required 17.19 s; and the permutation feature importance analysis required 2.85 s. Among these stages, DNN training accounted for the longest processing time, whereas CA transformation and DNN prediction required substantially less time. The training cost is incurred before model deployment, while prediction is performed during inference. These results indicate that the proposed 1D CA–DNN framework has relatively low computational overhead during feature transformation and prediction. The framework may therefore have potential for speech-based Parkinson's disease screening, including telemedicine and mobile health applications, although further validation under real-world recording conditions is necessary before clinical use.

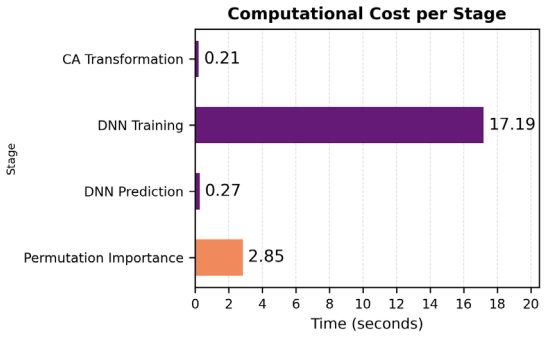

Fig. 15. Computational efficiency.

E. Clinical Relevance

The overall classification accuracy of the proposed 1D Cellular Automata-Deep Neural Network (DNN) achieved a mean ROC-AUC of 0.9895 and an accuracy of 95.14% in distinguishing PD from HC recordings on the evaluated dataset. Clinically, sensitivity and specificity are more insightful than overall accuracy because they show how sensitive the model is when predicting PD versus when falsely alarming.

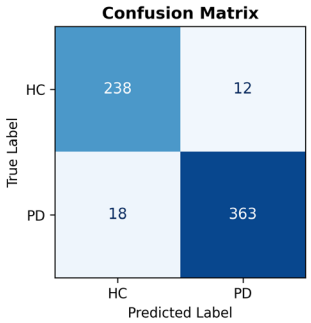

Fig. 16. Confusion matrix.

As illustrated in Fig. 16, the pooled confusion matrix summarizes the classification outcomes for all 631 recordings, comprising 381 Parkinson's Disease (PD) recordings and 250 Healthy Control (HC) recordings. The proposed model correctly classified 363 PD recordings, while 18 were misclassified as HC, resulting in a pooled sensitivity of 95.28%. Of the 250 HC recordings, 238 were correctly classified and 12 were misclassified as PD, yielding a pooled specificity of 95.20%. Based on these combined predictions, the pooled accuracy was 95.25%. The pooled accuracy reported in Table VI is different from the average accuracy (fold-level) reported in Table VII. The pooled accuracy is based upon the combined confusion matrix data for all of the five cross-validation folds. On the other hand, the average accuracy (fold-level) is determined by averaging the accuracy of each of the five cross-validation folds.

The confusion matrix clearly shows that the proposed 1D Cellular Automata–Deep Neural Network (DNN) solution achieves a relatively balanced classification with high sensitivity, specificity, and low misclassification rate. These results suggest that this framework could be a valuable computer-aided decision-support system for Parkinson's disease screening. Nonetheless, prior to clinical use, rigorous external validation of the model on multi-centre, linguistically diverse and real-world noisy speech datasets is required for robustness, generalizability and clinical reliability.

V. Discussion

The integration of one-dimensional Cellular Automata (1D CA) with acoustic features and a Deep Neural Network (DNN) classifier achieved a mean accuracy of 95.14% across the five folds of participant-grouped cross-validation. The results show improved classification accuracy when the 1D CA component is incorporated into both the MFCC-based and all-feature configurations. This improvement suggests that CA-based feature transformation may capture useful local interactions among feature components; however, further investigation is required to establish their specific contribution to Parkinson's Disease (PD)-related speech characteristics. Compared with the corresponding configurations without 1D CA, the proposed framework demonstrated improved classification accuracy under the evaluation protocol used in this study. The framework uses a feed-forward DNN classifier following CA-based feature transformation. Sensitivity and specificity are relevant in terms of how a clinical screening tool performs since incorrect negative and positive classification can impact further evaluation by clinicians. However, these laboratory-based findings alone cannot determine if a clinical tool is effective.

Fig. 17 presents the average Receiver Operating Characteristic (ROC) curve of the proposed model over the five folds of the cross-validation process which was based on grouping participants. Table VII reports the average ROC Area Under the Curve (ROC AUC) was 0.9895. The average ROC curve was generated by interpolating the true-positive rates at each fold of the cross-validation process with the false-positive rate points that were used for all folds, and then averaging the interpolated values across the five folds. The combined confusion matrix, which combined the predictions from all 631 recordings, demonstrated a sensitivity of 95.28% and a specificity of 95.20%. These findings demonstrate that there is strong discrimination between recordings of Parkinson's Disease (PD) and recordings of Healthy Control (HC) within the data used to evaluate this methodology. These findings suggest that the proposed methodology should be investigated further as a possible screening/triage tool for speech analysis, possibly through telemedicine/remote monitoring or in low-resource environments. It should be noted that the findings presented above are based upon a subset of the Italian Parkinson's Voice and Speech (IPVS) dataset; therefore, it is possible that the methodology's performance would vary in real-world scenarios due to factors such as background noise, variations in recording equipment, variations in language, variability among populations, and variations in severity of the disorder.

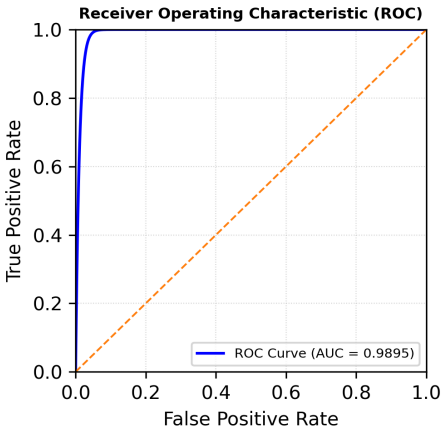

Fig. 17. Characterization of Receiver Operating Characteristic-Area Under the Curve (ROC-AUC).

A. Clinical Interpretation of Speech-based Biomarkers

The higher classification accuracy observed with speech-derived biomarkers supports their potential use for Parkinson's disease identification, subject to clinical validation. Jitter and shimmer changes indicate impaired neuromuscular control of the vocal folds, while a low harmonic-to-noise ratio describes breathiness and phonation breaks that are frequently present in Parkinsonian speech. This study aims to include those biomarkers into a 1D-cellular automata simulation framework so as to emphasize subtle pathological alterations in speech which make identification of patients with Parkinson's disease from healthy controls with higher observed accuracy.

B. Generalization and Robustness

The evaluation was participant-grouped, so recordings of the same participant never appeared in both the training and test folds. The consistent performance obtained under this protocol suggests that the framework learns discriminative acoustic patterns rather than relying on recordings of the same participants appearing in the training set.

C. Limitations

The proposed framework was evaluated using a selected subset of the publicly available Italian Parkinson's Voice and Speech (IPVS) dataset. Of the 831 recordings in the original corpus, 200 were excluded because feature extraction returned undefined (NaN) values. Since these exclusions may be related to recording characteristics, the analysed subset may not fully represent the original corpus. The class distribution shifted from approximately 47% Healthy Controls (HC) and 53% Parkinson's Disease (PD) in the original corpus to approximately 40% HC and 60% PD in the analysed subset. Differences in age and speech-task composition may also have influenced classification performance. All 631 analysed recordings and their HC/PD labels were checked against the original IPVS files and supplied annotations; however, the diagnoses were not independently verified against clinical records.

The evaluation of model performance was based on 5-fold StratifiedGroupKFold Cross-Validation, which ensured that all recordings for each participant were placed into the same fold. Only 53 participant folders were provided, with between 7 and 13 folders per test fold, therefore reducing the reliability of performance estimates at the fold level. As a result, no standard deviations for fold performance are presented; however, the 95% Wilson confidence intervals based upon the combined out-of-fold predictions represent the level of certainty regarding model performance. It is important to note that performing cross-validation within a single dataset does not demonstrate that a model will be applicable to other independent populations, languages, or recording environments.

There are other limitations to this study, including the fact that the binarization method employed within the Cellular Automata (CA) encoding is based on fixed threshold. As such, it cannot capture the full range of continuous variations in speech characteristics among different speakers, disease severities, and/or recording environments. In addition to exploring the potential of utilizing adaptive or probabilistic CA encoding methods, future studies will evaluate this methodology with larger, multilingual, and multi-center data sets, as well as independent cohort validation.

Finally, the CA parameters were selected through the use of the same five participant grouped cross-validation folds that were used in the performance evaluation; thus, there is potential for optimistic bias. The DNN was retrained after every sensitivity analysis was performed. At the selected parameter values, accuracy ranged between 95.12% and 95.54% for all of the reported analyses. Therefore, there was variability among different experimental runs. Thus, the reported mean five-fold cross-validation accuracy of 95.14% can be considered an estimate of the observed performance of the evaluation protocol used in this study and not necessarily an unbiased estimate of generalization performance.

D. Linking CA Patterns to Clinical Speech Symptoms

A CA-based feature encoding provides a discriminative representation for PD classification and an interpretable visualization of clinically relevant speech abnormalities. Fig. 13 shows the development of Cellular Automata (CA) patterns from Rules 30, 90, and 110 generated from HC and PD speech samples. These differences in patterns have qualitative associations with established clinical features of Parkinsonian speech. However, PD speech generates more fragmented and irregular CA evolution patterns than healthy speech under Rule 110. The disrupted structures correlate well with the increased cycle-to-cycle variations in vocal fold vibration reflected by higher jitter and shimmer values (common indicators of vocal tremor and neuromotor instability in Parkinson's disease).

Healthy speech tends to produce CA structures that are much more repetitive and regular, especially under Rule 90, demonstrating stable rhythmicity. On the other hand, PD speech may exhibit reduced loudness and pitch variability, which are recognized characteristics of Parkinsonian speech. In the present study, the observed differences in cellular automata patterns represent model-generated feature patterns and should be interpreted as computational representations rather than direct clinical evidence of hypophonia or monotone speech.

The proposed framework applies complex-systems analysis through 1D CA evolution to examine local relationships among acoustic features extracted from speech recordings. The resulting patterns provide a structured representation of the interactions among MFCCs, jitter, shimmer, and Harmonic-to-Noise Ratio (HNR). These features help characterize differences in spectral properties, phonatory stability, and voice quality between PD and HC speech samples. Through a process of neighborhood-based evolution, the CA transformation reorganizes the extracted feature states to create an improved representation that can then be used by the DNN for classification.

The overall CA framework converts conventional acoustic characteristics into organized development

patterns that may be visualized and examined with classification outcomes. These patterns offer another way of exploring localized characteristic relationships which relate to PD classification, in comparison to only the traditional characteristic representations. While the visualizations enhance the understanding of the computational characteristic change, additional validation of the patterns through the utilization of separate speech data sets and clinical assessments will be necessary prior to relating the patterns to particular clinical symptoms or employing them in clinical decision support systems.

E. Real-world Deployment Considerations

The proposed framework provides good results for controlled recording conditions; however, there are many practical problems that could occur when using this system in actual telemedicine or home monitoring applications. Possible causes for instability of the system include: ambient noise, variation in the quality of the microphone used, and/or variations in the recording environment. All of these factors could create issues with the model's performance due to changes in acoustic characteristics of the voice including: jitter, shimmer, and/or HNR. Additionally, there may be limitations to how well the learned CA patterns can be applied to different populations of patients. These limitations may be related to differences in the stages of the disease, differences among speakers (age, gender, accent, language), and/or differences among groups of speakers.

In order to address these issues, future releases of the system will likely include some type of channel invariant feature normalization algorithm, an adaptive noise suppression method, and a multi condition training set on simulated microphones with various signal to noise ratio levels.

In addition, adapting the domain through domain adaptation methods and/or establishing custom threshold within the CA binarization process can potentially allow the model to adjust to varying speaker characteristics and/or recording setups with minimal retraining required. Although there are other advantages to speech-based screening such as being low-cost and non-invasive, they remain two of the most significant reasons why this method remains a viable option for large scale Parkinson's disease monitoring. The suggested CA Deep Neural Network (DNN) pipeline might be incorporated into mobile health apps with the right robustness improvements, offering a useful instrument for early triage and longitudinal tracking in a variety of clinical and remote contexts.

F. Clinical Applicability

The proposed speech-based framework offers a non-invasive and cost-effective approach for supporting the early screening of Parkinson's disease. By leveraging easily obtainable speech recordings, the method has potential applicability in remote monitoring and clinical decision-support systems, complementing conventional neurological assessments.

VI. Conclusion

This study proposes a speech-based Parkinson's Disease (PD) detection framework that integrates extracted acoustic features with one-dimensional Cellular Automata (1D CA) and a Deep Neural Network (DNN) classifier. The CA conversion transforms continuous acoustic features into a binary pattern structure, which allows us to represent the relationship of adjacent feature states. The resulting CA based feature representations are used by the DNN to differentiate between healthy controls and people with PD. Experimental evaluation on the Italian Parkinson's Voice and Speech (IPVS) dataset yielded a classification accuracy of 95.14%, with additional performance assessment using sensitivity, specificity, and the area under the receiver operating characteristic curve (ROC-AUC). Future work will evaluate the framework on larger, independent datasets containing speech recordings from different languages and recording conditions to assess its generalizability. Further research will investigate two-dimensional Cellular Automata for modelling time-frequency representations, improve the interpretability of model predictions for clinical users, and examine additional CA rules for representing speech-feature patterns.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

H.S.R conducted the research, S.A developed the proposed methodology, H.S.R contributed to data preprocessing, S.A performed the experimental analysis, H.S.R wrote the original manuscript, S.A contributed to model evaluation, S.A revised the manuscript. Both authors had approved the final version.

Acknowledgement

The Italian Parkinson's Voice and Speech (IPVS) dataset, which is openly accessible, was used as the data source for the experimental assessment of the suggested framework for diagnosing Parkinson's disease.

References

[1] D. Escobar-Grisales, C. D. Ríos-Urrego, and J. R. Orozco-Arroyave, "Deep learning and artificial intelligence applied to model speech and language in Parkinson's disease," *Diagnostics*, vol. 13, no. 13, 2163, 2023. doi: 10.3390/diagnostics13132163

[2] P. Valarmathi, Y. Suganya, K. R. Saranya, and S. S. Priya, "Enhancing Parkinson disease detection through feature based deep learning with autoencoders and neural networks," *Sci. Rep.*, vol. 15, no. 1, 8624, 2025. doi: 10.1038/s41598-025-88293-w

[3] G. Dimauro and F. Girardi, "Italian Parkinson's Voice and Speech," *IEEE DataPort*, 2019. doi: 10.21227/aw6b-tg17

[4] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, 2010. doi 10.1109/TBME.2009.2036000

[5] B. E. Sakar, G. Serbes, and C. O. Sakar, "Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease," *PloS One*, vol. 12, no. 8, e0182428, 2017.
[6] Q. C. Ngo, M. A. Motin, N. D. Pah, P. Drotár, P. Kempster, and D. Kumar, "Computerized analysis of speech and voice for Parkinson's disease: A systematic review," *Computer Methods and Programs in Biomedicine*, vol. 226, 107133, 2022. doi: 10.1016/j.cmpb.2022.107133
[7] S. V. T. Dao *et al.*, "An analysis of vocal features for Parkinson's disease classification using evolutionary algorithms," *Diagnostics*, vol. 12, no. 8, 1980, 2022. doi:10.3390/diagnostics12081980G
[8] Nagasubramanian and M. Sankayya, "Multi-variate vocal data analysis for detection of Parkinson's disease using deep learning," *Neural Computing and Applications*, vol. 33, no. 10, pp. 4849–4864, 2021. doi:10.1007/s00521-020-05233-7
[9] A. Suppa *et al*., "Voice in Parkinson's disease: A machine learning study," *Frontiers in Neurology*, vol. 13, 831428, 2022. doi: 10.3389/fneur.2022.831428
[10] R. Khaskhoussy and Y. B. Ayed, "Speech processing for early Parkinson's disease diagnosis: Machine learning and deep learning-based approach," *Social Network Analysis and Mining*, vol. 12, no. 1, 73, 2022. doi: 10.1007/s13278-022-00905-9
[11] A. Favaro *et al*., "Interpretable Speech Features vs. DNN Embeddings: What to Use in the Automatic Assessment of Parkinson's disease in Multi-Lingual Scenarios," *Computers in Biology and Medicine*, vol. 166, 107559, 2023. doi: 10.1016/j.compbiomed.2023.107559
[12] B. Çelik and A. Akbal, "Distinguishing Parkinson's Patients Using Voice-Based Feature Extraction and Classification," arXiv preprint, arXiv: 2501.14390, 2025. doi: 10.48550/arXiv.2501.14390
[13] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12449–12460, 2020.
[14] H. S. Malekroodi, B.-I. Lee, and M. Yi, "Voice-Based Detection of Parkinson's Disease Using Machine and Deep Learning Approaches: A Systematic Review," *Bioengineering*, vol. 12, no. 11, 1279, 2025. doi: 10.3390/bioengineering12111279
[15] O. Klempíř and R. Krupička, "Analyzing wav2vec 1.0 embeddings for cross-database Parkinson's disease detection and speech features extraction," *Sensors*, vol. 24, no. 17, 5520, 2024. doi: 10.3390/s24175520
[16] S. Wolfram, *A New Kind of Science*, Champaign, IL, USA: Wolfram Media, 2002.
[17] B. Pfeifer, K. Kugler, M. M. Tejada, C. Baumgartner, M. Seger, M. Osl, T. Netzer, M. Handler, A. Dander, M. Wurz, A. Graber, and B. Tilg, "A cellular automaton framework for infectious disease spread simulation," *Open Medical Informatics Journal*, vol. 2, pp. 70–81, 2008. doi:10.2174/1874431100802010070
[18] A. R. Kansal, S. Torquato, G. R. Harsh IV, E. A. Chiocca, and T. S. Deisboeck, "Cellular automaton of idealized brain tumor growth dynamics," *Biosystems*, vol. 55, no. 1–3, pp. 119–127, 2000. doi:10.1016/S0303-2647(99)00089-1
[19] J. Moreira and A. Deutsch, "Cellular automaton models of tumor development: A critical review," *Advances in Complex Systems*, vol. 5, pp. 247–267, 2002. doi: 10.1142/S0219525902000572
[20] J. Li *et al*., "Image encoding and fusion of multi-modal data enhance depression diagnosis in Parkinson's disease patients," *IEEE Transactions on Affective Computing*, vol. 16, no. 1, pp. 145–160, Jan.–Mar. 2025. doi: 10.1109/TAFFC.2024.3418415
[21] P. L. Rosin, "Training cellular automata for image processing," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 2076–2087, Jul. 2006. doi: 10.1109/TIP.2006.877040
[22] M. Cook, "Universality in Elementary Cellular Automata," *Complex Systems*, vol. 15, no. 1, pp. 1–40, 2004. doi: 10.25088/ComplexSystems.15.1.1
[23] E. W. Weisstein. Rule 182. *MathWorld—A Wolfram Resource*. [Online]. Available: https://mathworld.wolfram.com/Rule182.html

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).