Automatic Classification of Post-Stroke Dysarthric Speech in Modern Standard Arabic Using Support Vector Machines
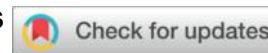
Khaled BAAZI [1], Siham BAHBOUH[3] , Mohamed AMMI[3] , Rachid BELLEMOU [1],

Abd-el-hak GASMI [2], Amina SAADEDINE[1]

[1]Centre de Recherche Scientifique et Technique pour le Développement de la Langue Arabe CRSTDLA,
Algiers, Algeria
[2]'Institut National de Recherche en Education INRE,
Algiers, Algeria
[3]'Université des sciences de la santé Dr Youcef El Khatib,
Algiers, Algeria

k.baazi@crstdla.dz
s_bahbouh@yahoo.fr
Ammimoh80yahoo.fr
brac23@yahoo.fr
gashak@proton.me
A.saadedine@crstdla.dz

Abstract

Stroke is among the leading causes of long-term disability worldwide, and the resulting speech and language impairments—such as aphasia and dysarthria—substantially degrade patients' quality of life. Despite significant advances in automatic speech analysis, the objective characterization and classification of post-stroke pathological speech remain insufficiently explored for Arabic, particularly within real clinical environments.

The objective of this study is to identify and classify speech disorders in stroke patients in Algerian hospital settings, relying on clinically grounded criteria. More specifically, the work focuses on the automatic classification of pathological speech (PS) produced by post-stroke speakers, with the aim of supporting diagnosis, assessment, and rehabilitation in hospital-based speech therapy practice.

To this end, a controlled corpus of logatomes with the structure [ACVC] was constructed. The consonantal slot [C] was instantiated by one of three Arabic plosives—[k], [q], and [t]—systematically combined with the three short vowels of Standard Arabic [a], [u], and [i]. The corpus was recorded by ten (10) post-stroke speakers and a control group of eighteen (18) healthy speakers, with no known history of speech production or perception. Clinical file analysis and direct patient evaluation were conducted to establish a typology of post-stroke language and speech disorders.

Prior to classification, a detailed acoustic analysis of the recorded speech was performed to extract features with established relevance for pathological speech characterization. These included fundamental frequency (F_0), the first three formants (F_1–F_3), overall energy (E_0), voice onset time (VOT), closure duration (silence corresponding to occlusion maintenance), consonant–vowel ([CV]) and vowel ([V]) durations, total word duration, cycle-to-cycle perturbation of F_0 (jitter), amplitude perturbation (shimmer), and the harmonics-to-noise ratio (HNR). These features

were subsequently used as input to Support Vector Machine (SVM) classifiers to discriminate pathological speech from normal speech (NS).

The results indicate that SVMs exhibit strong discriminative performance for the recognition and classification of post-stroke pathological speech, even in relatively small and clinically constrained datasets. The findings suggest that the proposed approach may contribute to the development of automatic diagnostic tools, expert systems capable of reliably identifying vocal anomalies, and decision-support systems for speech-language therapy education and clinical practice.

1. Introduction

Cerebrovascular accidents (CVA) constitute a major clinical pathology of the central nervous system (CNS). They are frequent, severe, and highly disabling conditions, widely recognized as a critical public health issue. CVAs represent the second leading cause of disability worldwide, the third leading cause of mortality, and the primary cause of morbidity in industrialized countries, where they also constitute the foremost cause of long-term disability. Indeed, approximately 75% of stroke survivors present residual impairments, and only 25% of individuals affected by stroke are able to resume full professional activity at their workplace [1].

In Algeria, although precise epidemiological statistics remain limited, the Algerian Society of Neurology estimates an annual national average of approximately 40,000 stroke cases [1].

Post-stroke speakers frequently experience a wide range of language and speech impairments. These disorders are heterogeneous in nature and may occur in isolation or in association with motor and/or sensory deficits. The type and severity of cognitive and linguistic impairments depend largely on the cerebral hemisphere affected. Dysarthria is observed in approximately 25% to 70% of post-stroke patients [2], with 42% of cases persisting beyond three months after onset [3].

Dysarthria—arguably the most acoustically tractable motor speech disorder—is typically characterized by slowed articulation (often 30–50% below normal speech rate [4]), increased jitter and shimmer reflecting phonatory instability [5], a reduced harmonic-to-noise ratio (HNR) indicative of turbulent airflow and incomplete vocal fold closure, as well as vowel centralization resulting from formant undershoot [6]. Additional dysarthric cues include pitch breaks, reduced intonational variability (monotony), and spectral narrowing, which collectively contribute to the disorder's acoustic signature [4].

While research on pathological speech produced by post-stroke speakers is substantial, with clinically effective applications such as the VOICE system achieving diagnostic accuracies of up to 84% [7], investigations targeting the Arabic language remain remarkably scarce.

In this study, we explore the impact of this pathology on the production of phonetic categories in Arabic, with a particular focus on stop consonants. We examine the automatic evaluation and classification of pathological speech using a set of acoustic biomarkers, including

voice onset time (VOT), harmonic-to-noise ratio (HNR), jitter, shimmer, and spectro-temporal representations, combined with state-of-the-art machine learning and deep learning strategies.

Through an acoustic analysis of speech segment production in post-stroke speakers, the measurement of the selected acoustic parameters enables inter-speaker comparison and longitudinal monitoring of speech production in individual patients. Our objective is to identify thresholds distinguishing pathological speech from normal speech (NS). For classification, we employ Support Vector Machines (SVMs), and we provide a detailed description of the classification framework along with its performance criteria. The obtained results are subsequently subjected to interpretation and discussion, with implications for both clinical assessment and automatic speech processing.

2. Stroke (AVC)

The World Health Organization (WHO) defines stroke as "the rapid development of clinical signs of focal (or global) disturbance of cerebral function, lasting more than 24 hours or leading to death, with no apparent cause other than vascular origin" [8].

Other definitions emphasize the intrinsic complexity of stroke, viewing it as the acute complication of a chronic vascular pathology, resulting either from an interruption of cerebral blood flow within a specific vascular territory (cerebral infarction) or from the rupture of a blood vessel (cerebral or subarachnoid hemorrhage). These vascular alterations often evolve silently over many years—or even decades—before their abrupt clinical manifestation. The resulting deprivation of oxygen leads to neuronal destruction within the affected brain regions. Epidemiological studies indicate that approximately 85% of strokes are ischemic in origin, while the remaining 15% are attributable to cerebral hemorrhage [9].

The clinical manifestations of stroke vary according to both the location and the extent of cerebral tissue damage. Among the most prominent sequelae are speech and articulation disorders, which may occur in isolation or in conjunction with motor and/or sensory impairments. Strokes affecting the left cerebral hemisphere are particularly associated with disorders of speech production and language processing, including deficits in articulation, fluency, and linguistic formulation [10].

At the same time, lesions involving the right hemisphere tend to produce prosodic flattening, characterized by a reduced pitch range [11], diminished intensity variation [12], and disturbances in stress assignment, which collectively attenuate the emotional and pragmatic contours of speech. These prosodic impairments, though sometimes less immediately apparent than segmental deficits, significantly compromise communicative effectiveness and listener perception.

Given the multifaceted nature of post-stroke speech disorders—encompassing segmental, suprasegmental, motor, and cognitive dimensions—early intervention by a speech-language pathologist constitutes a critical determinant of functional recovery. Early and targeted

rehabilitation not only facilitates the restoration of intelligible speech but also enhances overall communicative competence, thereby improving patients' autonomy and quality of life.

3. Standard Arabic Phonemes

Modern Standard Arabic (MSA), consistent with the broader typological patterns of Semitic languages, is characterized by a predominantly consonantal phonological architecture. Its segmental inventory typically comprises 28 consonants and a symmetrical system of six vowels, bifurcated into three short phonemes and their three long counterparts [13]. Within this inventory, approximately twelve segments are regarded as typologically marked—most notably the "emphatic" (pharyngealized) and uvular consonants. These segments exhibit a distinctive "retracted tongue root" [RTR] or pharyngeal feature, necessitating a secondary constriction in the posterior vocal tract that complicates their articulatory gesture [14].

The evidence indicates that such sounds—specifically the emphatic coronal obstruents —impose significant physiological demands, requiring precise neuromuscular coordination between the motor cortex and the brainstem [13, 14]. Consequently, these complex segments may be particularly vulnerable to neuromotor disruption, such as that observed in apraxia of speech or dysarthria, where impaired planning or execution of motor sequences often results in distorted or incomplete pharyngeal articulations [15].

This phonological specificity appears to motivate a language-aware paradigm for pathological speech analysis. Rather than assuming the direct translatability of models validated on Western languages, this work seeks to ground automatic classification in the specific phonetic and acoustic ecology of Arabic. By addressing the unique coarticulatory effects and spectral properties of these marked segments, we may improve both the scientific validity of diagnostic frameworks and their subsequent clinical applicability in Arabic-speaking populations.

4. Application of Support Vector Machines to the Recognition and Classification of Pathological Speech

Support Vector Machines (SVMs), also referred to as large-margin classifiers, constitute a supervised learning approach primarily employed for classification tasks. They are widely regarded as one of the most robust and effective methods in pattern recognition and classification, and are particularly well suited to automatic speech processing. Their theoretical foundation is grounded in the principles of statistical learning theory and the theory of generalization bounds developed by Vapnik and Chervonenkis [16].

Originally introduced by Boser et al. in 1992 [17], SVMs have achieved substantial success due to the strength and rigor of their theoretical underpinnings. They are applicable to a wide range of problems, especially classification tasks, and have demonstrated remarkable effectiveness in handling very high-dimensional data.

4.1 Operating Principle

The core objective of SVM-based classification is to determine an optimal hyperplane that separates data points belonging to different classes within a feature space, while maximizing the distance between these classes (Figure 01).
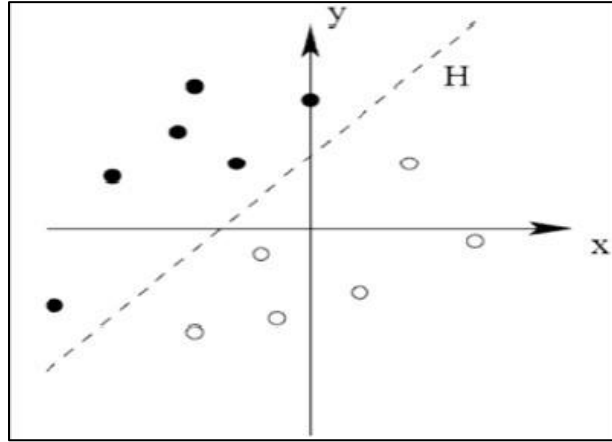

Figure 01: Separation of two point sets by a hyperplane H [16].

The closest points, which are the only ones used to determine the hyperplane, are called support vectors (Figure 02).
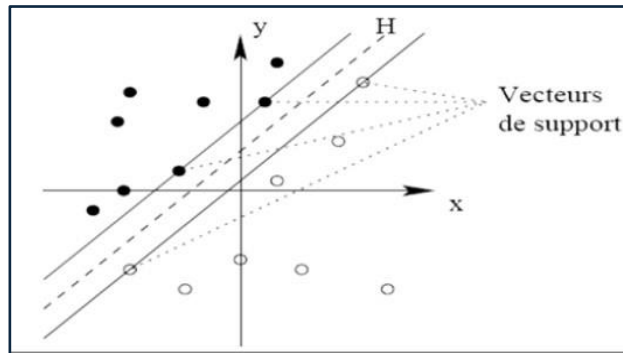

Figure 02: Support vectors [16].

Although infinitely many separating hyperplanes may exist, the fundamental criterion of SVM is to identify the unique optimal hyperplane. Geometrically, this corresponds to the hyperplane that maximally separates the two classes by passing through the "middle" of the margin, as illustrated in Figure 03.

This is equivalent to seeking a hyperplane whose minimum distance to the training examples is maximized. This distance is called the margin.

The optimal separating hyperplane is thus the one that maximizes this margin, which is why SVMs are referred to as large margin *classifiers*.
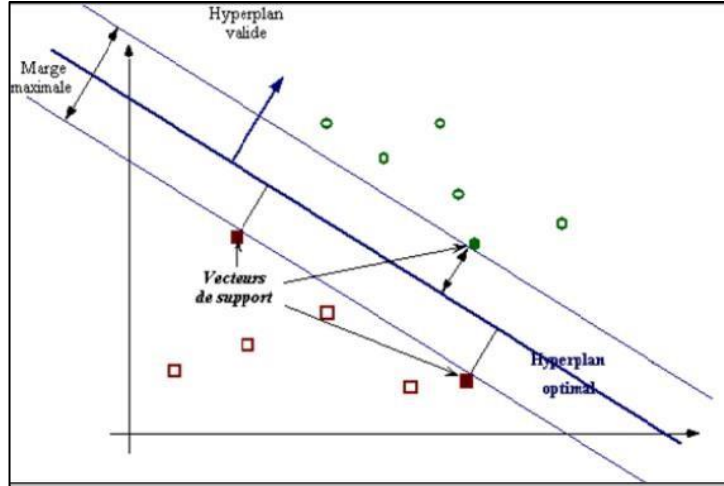
Figure 03: Optimal hyperplane, margin, and support vectors [16].

Within the SVM framework, two principal cases of separability are typically distinguished (Figure 04):

- **Linearly separable case:** the two data classes can be separated by a simple linear boundary, such as a straight line or a hyperplane (A).

- **Non-linearly separable case:** which characterizes the majority of real-world problems, where class boundaries are intrinsically complex and cannot be captured by linear separation alone (B).
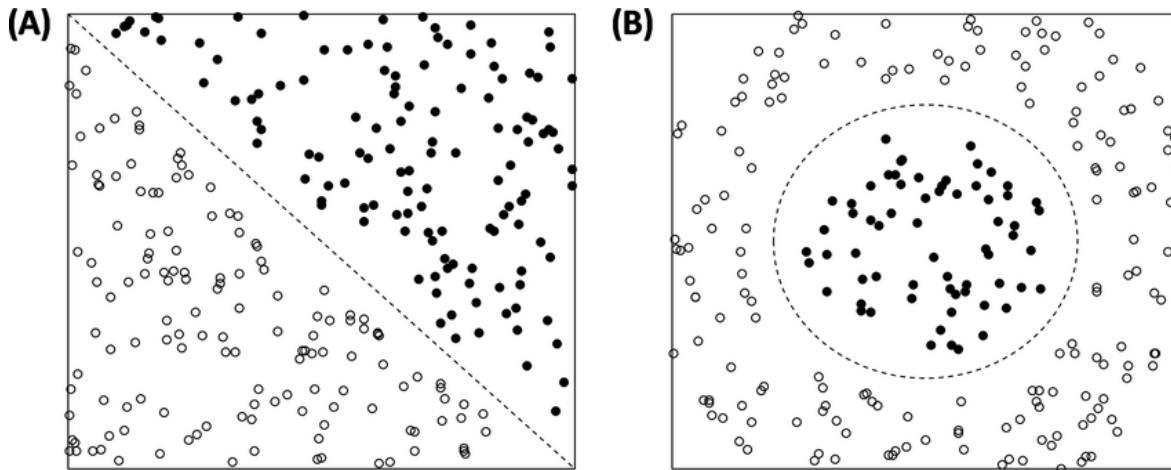

Figure 04: Linear and non-linear separability [16].

In practice, classification problems often involve more than two classes. Several strategies have therefore been proposed to extend SVMs to multi-class classification. Among the most commonly adopted approaches are:

- **One-vs-One (OvO):** This decomposition strategy trains a dedicated binary classifier for every pair of classes, resulting in N(N−1)/2 models. During prediction, each classifier votes for one of its two classes. The final decision is determined by majority voting. A key advantage of OvO is that each classifier is trained on a relatively balanced subset of data (only two classes), which can be beneficial for the learning process. However, this comes at the cost of increased computational overhead during training due to the quadratic number of models.

- **One-vs-All (OvA) / One-vs-Rest (OvR):** In this more computationally efficient approach, N binary classifiers are trained. The *k*-th classifier is tasked with distinguishing instances of class *k* from all other classes combined. Formally, each classifier learns a decision function fkfk, typically outputting a confidence score. The predicted class for a new sample is the one whose corresponding classifier yields the highest score (or largest margin). While OvA requires training fewer models, each one faces a class-imbalance problem, as the "rest" class typically contains many more samples than the positive class *k*.

A crucial component of SVM performance is the choice of the kernel function, which implicitly maps the input data into a higher-dimensional feature space where linear separation becomes feasible. Several families of parameterized kernel functions are commonly used. The selection of an appropriate kernel, as well as the optimization of its parameters, is application-dependent and typically achieved through statistical procedures such as cross-validation . Table 01 summarizes the most widely used kernels.

Table 01: Commonly used kernel functions [16]

Kernel type	Kernel function
Laplace	$K(x_i, x_j) = \exp(-\| x_i - x_j \|/\sigma)$
Linear kernel	$K(x_i, x_j) = x_i \cdot x_j$
Polynomial kernel	$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d$
RBF (Gaussian) kernel	$K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2)$
Sigmoid kernel	$K(x_i, x_j) = \tanh(\gamma x_i \cdot x_j + r)$

5. Acoustic Analysis of Pathological Speech (PS)

The present study aims to quantitatively characterize acoustic alterations in speech produced by post-stroke speakers, through a systematic comparison with a matched group of healthy speakers. This comparative approach enables the analysis of a range of articulatory–acoustic indices observable in a continuous acoustic signal and facilitates the identification of

discriminative acoustic markers likely to contribute to both clinical assessment and longitudinal monitoring of patients.

5.1 Participants

Our research was conducted in collaboration with two hospital institutions in Algiers: the Neurology Department of EHS Ali Aït Idir and the EPH Djillali BelKhenchir.

Participants were divided into two main groups: 10 post-stroke speakers (07 females and 03 males) and 18 speakers without speech disorders (14 females and 04 males), who served as the control group. The control speakers were matched to the pathological group in terms of age and gender. The mean age of the post-stroke participants was 51.1 years (± 15.50).

Audio recordings of the patients were conducted under identical conditions in the speech therapists' offices of both hospitals, following rehabilitation sessions, in order to ensure recording consistency.

5.2 Speech Corpus

The corpus consists of carrier words (logatomes) containing [CV] syllabic pairs, where the consonant [C] corresponds to one of the three stop consonants [k, q, ṭ], combined with one of the three vowels [a, u, i].

The primary function of the carrier word is to ensure the contextual independence of the unit under study from its phonetic environment, resulting in relative invariance in the time–frequency spectrum, particularly in vowel formant contours.

Accordingly, the carrier word structure [# A – CV – C #] was adopted. The resulting corpus comprises the following nine logatomes:

[aṭaṭ], [aṭuṭ], [aṭiṭ], [aqaq], [aquq], [aqiq], [akak], [akuk], and [akik].

It should be recalled that stop consonants are produced by a momentary closure of the vocal tract, followed by a sudden release. The closure is achieved through the coordinated action of various speech articulators at different places within the oral cavity. Stop consonants thus result from two distinct physical processes: (i) airflow blockage at the point of closure, and (ii) a rapid release of the occlusion, which may generate an explosive or fricative component.

From an acoustic perspective, a stop consonant consists of a sequence of acoustic events:

- A silence interval, corresponding to the articulatory hold phase of complete vocal tract closure. In the case of voiced stops, this silence is not absolute, as vocal fold vibrations during the closure phase produce low-frequency energy (approximately 100–300 Hz), commonly referred to as the voicing bar.

- A burst, occurring at the release of the occlusion, which corresponds to a short-duration pressure wave generated by the sudden release of compressed air behind the articulatory blockage.

- A frication noise component, resulting from the relatively slow separation of the articulators, during which the constrictive channel remains sufficiently narrow to generate turbulent airflow.

5.3 Sound Recording and Database Construction

Acoustic data were collected using a TASCAM DR-05 digital recorder. Recordings were stored as 16-bit ".wav" files with a sampling frequency of 44.1 kHz. During the recording sessions, participants were seated comfortably and instructed to articulate naturally while minimizing body movement, thereby maintaining a relatively stable posture with respect to the portable recorder. Frequent breaks were provided to allow participants to drink water and/or rest, with particular attention paid to fatigue effects, given the clinical status of the post-stroke speakers.

Each word in the corpus was repeated at least three times by each speaker during the recording session, ensuring sufficient data redundancy for reliable acoustic analysis.

The acoustic assessment focused on parameters with well-established discriminative relevance for the characterization of stop consonants in the Arabic language. These parameters were selected to capture both the source and filter components of the speech signal, as well as indices reflecting suprasegmental stability. Specifically, the analysis included:

- Source-related measures: fundamental frequency (Pitch Mean, Pitch Min, and Pitch Max), harmonics-to-noise ratio (HNR), jitter (cycle-to-cycle variability of fundamental frequency), and shimmer (cycle-to-cycle amplitude variability).

- Filter-related measures: the first three formant frequencies (F_1–F_3).

- Intensity and temporal indices: overall energy (E_0), durations of the logatome ([ACVC]), the consonant–vowel sequence ([CV]), and the vowel ([V]), as well as voice onset time (VOT) and the silence interval corresponding to the articulatory holding phase of complete vocal tract occlusion during stop production.

5.4 Extraction of Acoustic Parameters

The corpus was segmented according to speech type. Table 02 presents the distribution of audio files between pathological speakers and control speakers.

Table 02. Number of sound files recorded for pathological speech (PP) and normal speech (PN)

Speech type	Pathological speech		Normal speech		Total
	Speakers	Files	Speakers	Files	Files
Sound files	10	360	18	567	927

The detailed acoustic extraction is presented in Table 03, which illustrates representative samples highlighting variability in voice onset time (VOT), segmental durations, spectral measures, and perturbation indices. Parameters such as fundamental frequency (F0) and formant distribution provide insight into vocal tract resonance configuration, whereas jitter, shimmer, and the harmonics-to-noise ratio (HNR) reveal phonatory instability, a phenomenon frequently observed in speech disorders associated with stroke.

Table 03. Example of acoustic analysis of pathological speech recordings

Word	CV	V	Silence	VOT	E0	F0	F0 min	F0 max	F1	F2	F3	Jitter	Shimmer	HNR	Word duration
[aṭat]	0.17	0.10	0.042	0.0264	55.171	198.138	172.945	207.856	629	1575	2600	1.005	5.631	12.816	0.471
[aṭat]	0.182	0.099	0.063	0.0268	53.949	193.141	165.518	210.508	652	1304	2624	1.509	8.149	10.604	0.569
[aṭat]	0.219	0.096	0.101	0.0128	57.288	198.477	168.777	220.020	641	1337	2645	1.726	8.337	10.824	0.611
[aṭut]	0.185	0.099	0.053	0.0207	57.214	198.637	168.556	216.928	563	1189	2684	1.381	6.337	11.833	0.536
[aṭut]	0.199	0.098	0.070	0.0221	55.919	193.668	167.306	208.087	555	1166	2654	1.129	7.682	12.670	0.526
[aṭut]	0.166	0.094	0.053	0.0245	47.161	159.283	139.692	173.259	438	1115	2467	1.721	8.70	11.670	0.420
[aṭit]	0.177	0.098	0.057	0.0183	56.146	195.604	163.058	212.806	503	1630	2614	1.548	7.859	11.214	0.538
[aṭit]	0.142	0.084	0.039	0.0179	47.235	165.780	138.563	186.840	392	1596	2440	2.080	7.214	9.902	0.425
[aṭit]	0.193	0.097	0.070	0.0178	56.644	194.586	163.282	212.627	508	1525	2596	1.642	5.655	10.728	0.502

The values of the investigated acoustic parameters obtained from the acoustic analysis are reported as mean ± standard deviation (SD). We subsequently compare the mean values of the acoustic parameters extracted from pathological speech with those obtained from healthy speech, as summarized in Table 04.

Table 04. Mean ± standard deviation of the investigated acoustic parameters

Acoustic Parameters	Pathological Speech	Normal Speech
CV	0.203 ± 0.049	0.163 ± 0.037
V	0.084 ± 0.024	0.095 ± 0.023
Silence	0.081 ± 0.024	0.043 ± 0.018
VOT	0.023 ± 0.008	0.021 ± 0.016

E0	45.4 ± 7.8	81.4 ± 9.7
F0	188.6 ± 37.0	207.3 ± 46.2
Pitch min	170.4 ± 34.0	193.0 ± 48.3
Pitch max	206.0 ± 40.2	228.7 ± 56.3
F1	601.9 ± 113.1	581.1 ± 97.2
F2	1455.0 ± 222.4	1430.0 ± 186.7
F3	2754.2 ± 353.3	2797.7 ± 285.3
Jitter	1.62 ± 1.39	1.64 ± 0.98
Shimmer	8.39 ± 6.11	8.30 ± 5.79
HNR	12.77 ± 4.4	11.63 ± 3.91
Word duration	0.578 ± 0.102	0.590 ± 0.104

Based on Table 04, several observations can be made.

- CV duration appears to be significantly longer in pathological speakers, suggesting a disruption in consonant–vowel coordination rather than a uniform slowing of speech.

- Vowel duration tends to be slightly shorter in the pathological group. The difference is modest, yet it may indicate compensatory temporal adjustments within the syllable structure.

- Silence durations (consonantal hold phases) are markedly longer in pathological speech. This finding is consistent with impaired articulatory transitions and delayed segmental release.

- Voice Onset Time (VOT) is prolonged in pathological speakers, which may reflect deficits in laryngeal–supralaryngeal timing and reduced precision in voicing control.

- Fundamental frequency (F0) is globally lower and less variable in pathological speech. This reduction in pitch level and variability suggests diminished laryngeal control and reduced prosodic flexibility.

- Minimum pitch values are also lower, reinforcing the tendency toward an overall lower and less dynamic voice.

- Maximum pitch values are reduced, resulting in a substantially narrower pitch range. This restricted range limits the speakers' ability to produce intonational modulation, a key component of expressive speech.

- First formant (F1), which is primarily related to vowel aperture, does not appear to be systematically affected in this logatome corpus, indicating relative preservation of vertical tongue displacement.

- Second formant (F2) shows a slight decrease in pathological speech. This reduction may reflect diminished articulatory precision, particularly in anterior–posterior tongue movements, which are often slower or reduced in amplitude in motor speech disorders.

- Third formant (F3) exhibits a slight decrease, which may be associated with alterations in vocal tract configuration, including tongue posture and velopharyngeal control.

- Jitter values are higher and more variable in pathological speakers. This parameter directly indexes irregularities in vocal fold vibration, pointing to reduced phonatory stability and increased voice roughness.

- Shimmer is also elevated, supporting the hypothesis of impaired neuromuscular control of the larynx, leading to abnormal variations in glottal amplitude.

- The mean difference in HNR is minimal. This result may appear counterintuitive given the observed reduction in vocal energy. It suggests that, despite hypophonia, the signal-to-noise ratio of the voiced segments may be relatively preserved, or that noise components scale proportionally with the reduced signal amplitude.

- Total word duration is remarkably similar between the two groups. This is a crucial finding: it indicates that temporal alterations are localized at the segmental level (particularly CV units) rather than reflecting a global reduction in speech rate. Such a pattern points toward a deficit in internal syllabic timing and coordination, rather than generalized motor slowing.

Taken together, these acoustic alterations are consistent with a diagnosis of dysarthria, in which fine motor control, strength, and coordination across the respiratory, phonatory, and articulatory subsystems are compromised. From an engineering and computational perspective, these parameters constitute highly informative acoustic biomarkers, making them strong candidates for inclusion in an automatic pathological speech classification framework.

6. Support Vector Machine (SVM) Application

After data acquisition and preprocessing—consisting of normalization, filtering, and removal of null values for the selected acoustic parameters—a post-processing stage was carried out through dimensionality reduction, using both Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

Classification was then performed using Support Vector Machines (SVMs), combined with temporal smoothing based on a sliding-window voting strategy and the computation of a confidence score for each decision.

Since the extracted data are not linearly separable—as illustrated, for example, by the distribution of the acoustic parameter V (see Figure 05)—a polynomial kernel was adopted instead of the commonly used Radial Basis Function (RBF).
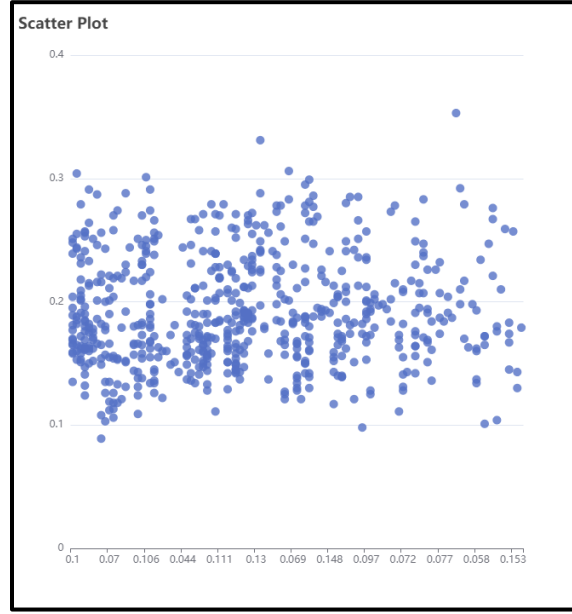

Figure 05. Scatter plot representation of the acoustic parameter V values.

SVM Decision Function

The decision function of the SVM classifier with a polynomial kernel is defined as:

$$f(x) = \text{sign}\left[\sum_{i=1}^{N_{sv}} \alpha_i\, y_i K(x_i, x) + b\right]$$

where the polynomial kernel is expressed as:

$$K(x_i, x_j) = (\gamma\langle x_i, x_j\rangle + r)^d$$

with:

- α_i: Lagrange multipliers
- $y_i \in \{-1, +1\}$: class labels
- b: bias term
- N_{sv}: number of support vectors

The SVM hyperparameters were fixed a priori, without further optimization, as summarized in Table 05.

Table 05. SVM hyperparameter configuration

Parameter	Symbol	Value	Description
Polynomial degree	d(power)	1.0	Polynomial order
Kernel coefficient	γ(gamma)	1.0	Scaling factor of inner products
Constant term	r(bias)	0.1	Added constant shift
Margin penalty	C	1.0	Regularization parameter

This configuration represents a deterministic and reproducible choice, corresponding to a homogeneous polynomial kernel of degree 1, with a positive constant bias and unit-margin regularization.

While simple, this setup provides a controlled baseline and avoids overfitting in a relatively limited dataset.

Operational Procedure

The implementation followed the steps below:

- **Pre-processing:** Acoustic descriptors were standardized (centered to zero mean and scaled to unit variance).
- **Training:** The dual SVM optimization problem was solved using the Sequential Minimal Optimization (SMO) algorithm
- **Stopping criteria:** The solver iterated until the Karush–Kuhn–Tucker (KKT) conditions were satisfied within a tolerance of 10^{-3}
- **Outputs:** The training procedure yielded the set of support vectors, their corresponding Lagrange multipliers (α_i), and the decision bias (b)

Decision Rule

For a new test sample x_{test}:

- Compute similarities with each support vector:

$$s_j = K(x_{\text{test}}, x_j)$$

Compute the decision score:

$$\text{score} = \sum_{j=1}^{N_{sv}} \alpha_j \, y_j s_j + b$$

14

Assign class label:

$$\hat{y} = \begin{cases} +1 & \text{if score} \geq 0(\text{pathological speech}) \\ -1 & \text{otherwise(normal speech)} \end{cases}$$

Performance Evaluation

Classification performance was quantified using the accuracy metric:

$$\text{Accuracy} = \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}} \times 100\%$$

Evaluation was conducted using a leave-one-subject-out (LOSO) cross-validation protocol, ensuring strict subject independence between training and testing sets—a critical requirement in clinical speech analysis.

6.1 Learning Phase

The learning phase enables the system to ingest reference parameters derived from the recorded audio files used in the acoustic analysis for this stage. These reference vectors are obtained from acoustic models designed to characterize the various speech sounds produced by the speakers. The underlying principle consists in providing the classifier with a set of input examples x and their corresponding labels y, and subsequently learning a model that approximates the expected outcomes, with the dual objective of achieving a high recognition rate and, more critically, strong generalization capability. On the basis of the learned examples, the system becomes capable of processing previously unseen data that are nonetheless acoustically similar to the training instances.

For the present classifier, the training phase was carried out using a matrix of acoustic feature vectors extracted from 638 audio files, comprising 252 pathological speech (PS) files and 386 normal speech (NS) files (Table 06).

Table 06. Number of PS and NS audio files used for learning

Speech type	PS	NS	Total
Sound files	252	386	638

6.2 Classification Phase

Feature extraction is treated as input data to the classifier, whose objective is to determine an optimal separating hyperplane that discriminates between examples during the learning phase, and subsequently to make a classification decision during the identification phase [18].

An automatic classification of pathological speech (PS) versus normal speech (NS) was conducted. This procedure relied on a quantitative classification of the values obtained directly from the acoustic analysis.

This phase makes it possible to assess the generalization capability of the classifier, that is, its ability to produce reliable results on any dataset drawn from the same statistical distribution.

To validate the classification experiments, we employed an independent test set composed of sound files produced by the speakers who participated in the recordings, comprising 289 acoustic files, including 108 PS files and 181 NS files (Table 07).

Table 07. Number of PS and NS audio files used for classification

Speech type	PS	NS	Total
Sound files	108	181	289

The resulting confusion matrix is reported in Table 08.

Table 08. Number of PS and NS audio files used for classification

Speech type	PS	NS
PS	104	04
NS	15	166

7. Discussion of Classification Results

For the evaluation of our classifier, we followed the complete workflow from reading acoustic signals to classification testing. The parameters extracted during the acoustic analysis were fed into the automatic classifier, which then determined the nature of the spoken word.

From the test set of 108 pathological speech (PS) files, 104 were correctly recognized, while 4 files were misclassified as normal speech (NS) (Table 08).

The performance of the automatic classifier is assessed in terms of the percentage of correct classifications for the test files presented as input (Table 09).

Table 09. Classification accuracy for PS and NS audio files

Speech type	PS	NS
PS	96.3	3.7
NS	8.7	91.3

According to the results, the proposed SVM-based system achieves a high recognition rate for PS relative to NS.

On the test data (representing 30% of the samples), the SVM model attained an overall accuracy of 93.4%, with a sensitivity of 96.3% for pathological cases (Table 09).

8. Conclusions

This study emphasizes the specific requirements of assessment in the Arabic language, proposing that any comprehensive framework for speech pathology must incorporate language-specific articulatory features.

Post-stroke acoustic evaluation should prioritize measurements of signal energy, cyclical stability, and temporal fluency, with clearly defined clinical thresholds to guide therapeutic intervention. The measurement of selected acoustic parameters enables comparisons across subjects and allows longitudinal tracking of speech production in these patients. Such acoustic analysis can support speech rehabilitation, providing objective metrics to evaluate patients' progress over time and furnishing clinicians with concrete data to improve therapeutic outcomes.

The investigated acoustic markers clearly distinguish post-stroke pathological speech, which is primarily characterized by:

1. A marked reduction in signal energy (extreme hypophonia).

2. Internal temporal disorganization within syllables, manifested through selective segmental lengthening.

3. Increased phonatory instability, reflected in elevated jitter and shimmer values.

4. Reduced prosodic variability, limiting expressive and intonational modulation.

These manifestations are further associated with :

- Reduced articulatory coordination,

- Increased laryngeal instability,

- Decreased phonatory efficiency,

- Greater acoustic unpredictability.

Our SVM-based classification system, employing a feature vector including VOT, consonantal silence, CV and V durations, E_0, F_0, F_1–F_3, jitter, shimmer, HNR, and word duration, demonstrated notable performance, achieving a recognition accuracy of 96.3% for pathological speech (PS) and 91.3% for normal speech (NS). These results indicate that SVMs represent a robust and well-established method for automatic speech processing, particularly suited for classification tasks with moderately sized corpora and well-defined features. Their application to

vocal pathology detection, as in our study on Arabic stop consonants in post-stroke speakers, yields excellent performance grounded in solid theoretical principles.

In conclusion, SVMs provide an efficient, robust, and theoretically sound framework for automatic dysarthria detection, especially in data-constrained clinical scenarios. While deep learning excels with big data, the interpretability, computational efficiency, and strong generalization of SVMs from limited samples make them an indispensable tool for bridging acoustic phonetics and clinical speech processing. Future work could explore hybrid models or kernel optimizations to further enhance performance.

References

[1] BEHNAS, Lynda. *Prise en charge des accidents vasculaires cérébraux au service de réanimation médicale de l'hôpital militaire régional universitaire de Constantine*. Thèse de doctorat. Faculté de Médecine, Université de Constantine 3, 2024, Algérie.

[2] DUFFY, J. R. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. 4th ed. Elsevier, 2019.

[3] DE COCK, Elien; OOSTRA, Kristine; BLIKI, Lisa; VOLKAERTS, Anne-Sophie; HEMELSOET, Dimitri; DE HERDT, Veerle; BATENS, Katja. Dysarthria following acute ischemic stroke: Prospective evaluation of characteristics, type and severity. *International Journal of Language & Communication Disorders*, 2021. DOI: 10.1111/1460-6984.12607.

[4] KIM, Y.; WEISMER, G. Speech intelligibility and kinematic variability in dysarthria. *Journal of Speech, Language, and Hearing Research*, 2011, 54(4), p. 1101–1111.

[5] BAKEN, R. J.; ORLIKOFF, R. F. *Clinical Measurement of Speech and Voice*. 2nd ed. Cengage Learning, 2000.

[6] WHITFIELD, J. A.; GOBERMAN, A. M. Vowel centralization in dysarthria: A systematic review. *American Journal of Speech-Language Pathology*, 2017, 26(2), p. 643–659.

[7] ABDALLA, I. M., et al. VOICE: A prehospital automated speech-based stroke screening system. *IEEE Journal of Biomedical and Health Informatics*, 2023, 27(4), p. 1641–1652.

[8] SACCO, R. L.; KASNER, S. E.; BRODERICK, J. P.; CAPLAN, L. R.; CONNORS, J. J. B.; CULEBRAS, A., et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 2013, 44(7), p. 2064–2089.

[9] MURPHY, Stephen J. X.; WERRING, David J. Stroke: causes and clinical features. *Medicine*, 2020, 48(9). DOI: 10.1016/j.mpmed.2020.06.002.

[10] EUSTACHE, Francis; FAURE, Sylvane; DESGRANGES, Béatrice. *Manuel de neuropsychologie*. Dunod, Univers Psy, 2023.

[11] ROSS, E. D. The right hemisphere and affective prosody. In: CUMMINGS, J. (Ed.). *Handbook of Clinical Neurology*, vol. 70. Elsevier, 2010.

[12] PELL, M. D. Prosody–emotion dissociation after unilateral brain damage. *Brain and Language*, 2007, 101(1), p. 64–79.

[13]   MOST, Tova, LEVIN, Iris and SARSOUR, Mutie. 2007. The Effect of Modern Standard Arabic Orthography on Speech Production by Arab Children With Hearing Loss. *Journal of Deaf Studies and Deaf Education*. vol. 13, no. 3, pp. 417–431. DOI 10.1093/deafed/enm060. Available from: https://doi.org/10.1093/deafed/enm060

[14]   AL-BATAINEH, Hussein. 2019. Emphasis Harmony in Arabic: A Critical Assessment of Feature-Geometric and Optimality-Theoretic Approaches. *Languages*. vol. 4, no. 4, p. 79. DOI 10.3390/languages4040079. Available from: https://doi.org/10.3390/languages4040079

[15]   SHRIBERG, Lawrence D., PAUL, Rhea, BLACK, Lindsey M. and VAN SANTEN, Jan P. H. 2010. The Hypothesis of Apraxia of Speech in Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*. vol. 41, no. 4, pp. 405–426. DOI 10.1007/s10803-010-1117-5. Available from: https://doi.org/10.1007/s10803-010-1117-5

[16]   ZIDELMAL, Zahia. *Reconnaissance d'arythmies cardiaques par Support Vector Machines (SVMs)*. Thèse de doctorat, 2012.

[17]   BOSSER, M.; GUYMON, et VAPNIK, V. A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual ACM Conference on Computational Learning Theory*, 1992.

[18]   ZAIZ, F. *Les Supports Vecteurs Machines (SVM) pour la reconnaissance des caractères manuscrits arabes*. Mémoire de magistère en Informatique, Intelligence Artificielle et Systèmes Distribués, Université Mohamed Khider, Biskra, 2016.