IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 21, 2020, accepted January 13, 2021, date of publication January 21, 2021, date of current version February 1, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3053335

Classification of Dysarthric Speech According to the Severity of Impairment: an Analysis of Acoustic Features

BASSAM ALI AL-QATAB, (Member, IEEE), AND MUMTAZ BEGUM MUSTAFA
Department of Software Engineering, Faculty of Computer Science and Information Technology University of Malaya, Kuala Lumpur 50603, Malaysia
Corresponding author: Bassam Ali Al-Qatab (bassam_qatab@hotmail.com)

ABSTRACT The automatic speech recognition (ASR) system is increasingly being applied as assistive technology in the speech impaired community, for individuals with physical disabilities such as dysarthric speakers. However, the effectiveness of the ASR system in recognizing dysarthric speech can be disadvantaged by data sparsity, either in the coverage of the language, or the size of the existing speech database, not counting the severity of the speech impairment. This study examines the acoustic features and feature selection methods that can be used to improve the classification of dysarthric speech, based on the severity of the impairment. For the purpose of this study, we incorporated four acoustic features including prosody, spectral, cepstral, and voice quality and seven feature selection methods which encompassed Interaction Capping (ICAP), Conditional Information Feature Extraction (CIFE), Conditional Mutual Information Maximization (CMIM), Double Input Symmetrical Relevance (DISR), Joint Mutual Information (JMI), Conditional redundancy (Condred) and Relief. Further to that, we engaged six classification algorithms like Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Artificial Neural Network (ANN), Classification and Regression Tree (CART), Naive Bayes (NB), and Random Forest (RF) in our experiment. The classification accuracy of our experiments ranges from 40.41% to 95.80%.

INDEX TERMS Acoustic features, automatic dysarthric speech recognition system, dysarthria, classification algorithms, feature selection methods.

I. INTRODUCTION

Speech impairment is a condition in which the ability to produce speech sounds that are necessary for communicating with others is impaired. The condition may be acquired or developed. Speech impairment could be mild, such as occasionally mispronouncing a couple of words, or it can be severe, such as not being able to produce speech sounds at all.

Many terms are used in reference to speech impairment, such as childhood apraxia of speech (CAS), dysarthria, stuttering voice, and others. Of these few terms, the term dysarthria refers to impairment of the neuron-motor speech, where the muscles controlling the speech organs are weak. These muscles either move slowly, or they may not move at all. The causes of dysarthria can be attributed to muscle dystrophy, cerebral palsy, head injury, and also stroke [1]–[4].

Dysarthria can be categorized based on the presentation of symptoms, which can be hypokinetic, hyperkinetic, ataxic, flaccid-spastic mix, spastic, and flaccid [5]–[8]. The severity level concerns the degree of dysarthric impairment which needs the experts' perception [9], [10]. A common assessment tool is the Frenchay Dysarthria Assessment [11], [12] and the Computerized Assessment of Intelligibility of Dysarthric Speech (CAIDS) [13].

The classification of dysarthria has gained importance among researchers due to a number of reasons. Firstly, it has helped us to fully understand the types of impairment which can result in empirical features that can be used to develop programs that can easily identify the disorder and its characteristics [3], [10], [14]–[16]. Secondly, classifications are needed to compare the different types of dysarthria with each other or with controlled speech, thereby resulting in more accurate identification of the impairment [14], [17]. However, thus far, there has been no comprehensive works done to examine the influence of acoustic features and feature selection methods on the classification of dysarthric speech based on the speech impairment's level of severity [17], [18].

The associate editor coordinating the review of this manuscript and approving it for publication was Gang Li.

To fill the gap, the current study examines the influence of acoustic features and feature selection methods on the classification of dysarthria speech based on the severity of the speech impairment. The outcome of this study will enhance the classification accuracy of spastic dysarthria because it is one of the most common types of dysarthria generally [4]. Spastic dysarthria is associated with a variety of disabilities such as, but not limited to, cerebral palsy and traumatic brain injury [4]. The remainder of the article is organized as follows: Section 2 focuses on related works that describe the acoustic features of dysarthric speech, Section 3 explains the method used towards achieving the objective of this research, Section 4 presents and discusses the major findings, and Section 5 concludes the article.

II. RELATED WORKS

One of the main challenges in differentiating the severity of the types of dysarthria is the lack of relevant analysis derived from a sufficient number of speakers with different types of dysarthria and various levels of severity [14]. Due to this inadequacy, it is thus important to characterize a particular speech impairment's effect on speech intelligibility [19]. The severity of the different types of dysarthria cannot be determined in terms of standard, and yet speech intelligibility has been frequently used to determine the level of speech mechanisms affected by the neurological disease [20]. Even though the number of speakers with different types of dysarthria with various severity levels is sufficiently large, the low number of associated analysis has made it difficult for professionals to differentiate the severity effects and the dysarthria types [14].

It appears that each severity level has its characteristics which can be used to classify speech impairment [3]. This has been noted in past studies [21]. For instance, the Kurtosis of Linear Prediction (LP) residual (κLP) signal has been used to distinguish the excitation of the atypical vocal source (referring to vocal breathing and harshness). Likewise, the rate-of-change of the signal in log-energy has also been used to characterize speech with short-term temporal dynamics. This is because the temporal impairments of speech are concentrated on an unclear distinction between the adjacent phonemes caused by the articulation's inaccurate placement [19]–[22].

The Low-to-High Modulation Energy Ratio (LHMR) has also been used to characterize the speech temporal impairments associated with the long-term temporal dynamics. Representation of the modulation spectral signal, which is auditory-inspired, is used to represent the modulation spectral energy's ratio of frequencies which are lower than 4 Hz to frequencies greater than 4 Hz [19]–[22]. Prosody features, for example, the standard deviation of the fundamental frequency (σ f0), range of the fundamental frequency f0 (Δf0), and percentage of segments of voice in words uttered ($\%v$) were used as parameters to identify speech impairments [19].

Harmonics-to-Noise Ratio (HNR), the Glottal-to-Noise Excitation ratio (GNE), and Mel Frequency Cepstral Coefficients (MFCCs) are speech features that have been used for classifying dysarthric speech based on the severity of the impairments [19], [23]. It appears that the MFCCs have the capability to capture the movements of the irregular vocal folds or the lack of closure of vocal-folds caused by a change in the mass/tissue [23]. Here, the GNE quantifies the excitation ratio due to vocal fold oscillations, as opposed to turbulent noise [24], and the HNR uses the difference in the ratio between the components of the periodic signal's energy and the component of the aperiodic signal's energy [25]. The combination of all these features into one dimension was proposed by [19].

Among some of the measures used to identify the severity of dysarthria is the Low-to-High Modulation energy Ratio (LHMR) [21]. The higher LHMR values are affected by the intelligibility level, depending on how the modulation spectral frequency contents are set (greater or lower than 4 Hz).

Some of the features like perturbations in temporal dynamics (long and short term), atypical excitation of the vocal source, separation of information of vocal tract and source, nasality, prosody, and composite measures can also be used to classify dysarthric speech based on the severity of the impairment [3], [21]. Nonetheless, it was stressed in [21], [26] that a linear combination of the characteristics of dysarthria speech tends to perform better than when using any single measure.

The Variability Index (VI) is defined as the average syllable variability for a given utterance, after the duration of neighboring syllables is compared with the normalized duration of each syllable [27]. When compared to a control group of speakers, the VI values were found to be lower for the group with ataxic dysarthria. This implies that controlled speech and ataxic dysarthria have different intersubjective variability in VI values [28].

A. ACOUSTIC FEATURES OF DYSARTHRIC SPEECH

There are many speech parameters, like voicing contrasts, nasalization, and vowel height which all play an important role in decreasing speech intelligibility [29]. Voice Onset Time (VOT) [28], second formant frequency (F2) slope [31], [32], and acoustic vowel space [33]–[35] are some of the acoustic features used to determine the speech intelligibility of speakers with dysarthria. The severity of the speech impairment is characterized according to acoustic measurements, such as slow rate of speaking, VOT with high variability, almost similar duration of utterance with regards to vowel/syllable, and fundamental frequency (F0) range across utterances that are abnormally large. These have also been associated with ataxic dysarthria [36].

Kim *et al.* [14] had examined the Root-Mean-Square (RMS) intensity contour, F0 contour, F2 transitions extent and duration, M1 for fricatives (/s/ and /$\int$/) during the three 50-ms-long windows approaching the vocalic nucleus (25-ms overlap between adjacent windows), first and second formant frequencies from four corner vowels, voiceless interval durations, and vowel and sentence durations. These measurements need to produce the necessary variables, such as RMS intensity range of utterance, F0 range (maximum-minimum) of utterance, F2 slope, M1 difference between /s/ and /ʃ/, acoustic vowel space, Pairwise Variability analysis (PVI), and rate of articulation for analysis.

Kim *et al.* [14] also noted that F2 slope, vowel space, the difference of M1 for /s/ and /ʃ/, rate of articulation, Voiceless Interval Duration, and the range of F0 interquartile, were significantly correlated with speech intelligibility. All clinical groups, except for Parkinson's disease (PD), had shown a significant rate of articulation, and the score of the speech intelligibility for all four disease groups showed a significant regression of the F2 slope.

As there were many different features recommended and used in previous studies such as the work in [37], it is thus vital to determine the features that would contribute to the highest classification accuracy for dysarthric speech.

III. METHOD

Since this study aims to examine the influence of acoustic features and feature selection methods on the classification of dysarthric speech according to the severity of speech impairment, the methodology adopted will include the speech corpus selection, acoustic features extraction, classification, and evaluation of the classification accuracy and ranking.

A. SPEECH CORPUS

The database used in this research contains the recorded speech of one dysarthric speaker with different levels of severity. The NEMOURS database [38] meets the above criteria and is used for feature extraction and classification. The NEMOURS speech database is a collection of 814 short nonsense sentences spoken by 11 male speakers. Each speaker was prompted to utter 74 sentences. The sentences are the form of "The X is Ying the Z" where $X \neq Z$ [38]. The target words X, Y, and Z had the constraints to provide closed-set phonetic contrasts (e.g. place, manner, and voicing contrasts) similar to [20].

The speakers of the NEMOURS database have been categorized according to three types of severity, which are mild, moderate, and severe dysarthria. The speakers are assessed and classified according to their severity levels by the speech-language pathologist based on the Frenchay Dysarthria Assessment [11], [12]. Four speakers were classified as severe, one speaker as moderate-to-severely dysarthria, one as moderate, and two speakers with very mild dysarthria [39]. More information about the speakers' severity levels and their intelligibility score can be found in [38]. The intelligibility score is computed as the average of scores for three sessions by 12 non-hearing impaired listeners.

We have used the recorded speech of nine speakers out of the original 11 dysarthric speakers to extract the speech acoustic features and testing of the classifiers. One of the speakers has some missing data and the other was left out to balance the number of speakers for each severity.

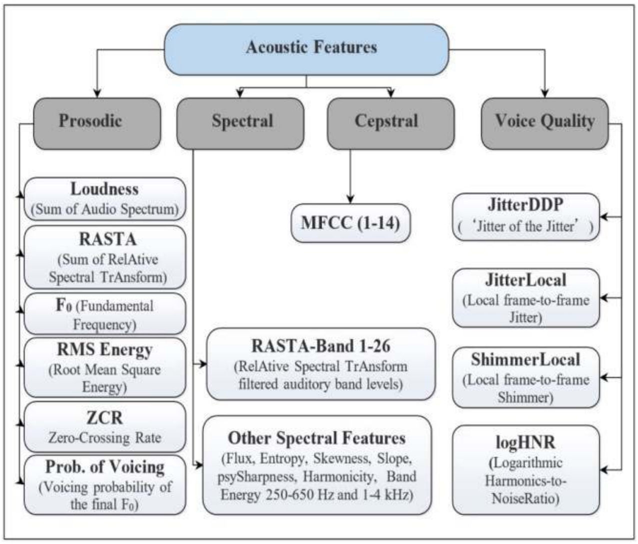

FIGURE 1. The acoustic features applied in this study.

B. ACOUSTIC FEATURES EXTRACTION

The large number of features extracted in this study made it difficult to perform the classification according to the severity of the speech impairment. As a result, feature selection is a possible solution for creating different training sets that can identify the most significant features related to the specific type of severity level. One alternative to doing this is to reduce the number of features used in the classification algorithms.

However, the existing literature does not suggest suitable methods for selecting the optimal number of feature parameters. In view of this, we have adopted a method proposed in [40], [41], represented by the formula below for better the computation cost.

$$NOF = log_2 n \tag{1}$$

where NOF is the number of feature parameters to be picked up for classification algorithms, and the total number of extracted features is n. As we have extracted 5673 features (n=5673), the number of feature parameters used in classification algorithms is13.

$$\begin{aligned} \mathrm{NOF} &= \log_2 5673 \\ NOF &= 12.47 \approx 13 Features \end{aligned}$$

The 13 features are; prosodic (Loudness, RASTA, Fundamental Frequency, RMS Energy, ZCR, Prob. of Voicing), spectral (RASTA-Band 1-26, Other Spectral Features), cepstral (MFCC (1-14)), and voice quality (JitterDDP, JitterLocal, ShimmerLocal, logHNR)

These 13 feature parameters are categorized into four acoustic features as shown in Fig. 1. They include: prosody, spectral, cepstral, and voice quality. For each feature, there are parameters computed for a short time frame of an audio signal at a given time, called the acoustic Low-Level Descriptors (LLD) [42], [43].

C. FEATURES SELECTION METHODS

The feature selection methods were applied prior to the running of the classification algorithms. The objective of using the different feature selection methods was to create different training sets and to increase the diversity among the classifiers, which is a key feature in improving the performance of the multi-classifiers system. In addition to this, the selection methods of two different features may give rise to two different sets of features.

Presenting only one feature set can be misleading; it may also produce suboptimal results [44], hence the seven feature selection method was used in this study. They include: Interaction Capping (ICAP) [45], Conditional Information Feature Extraction (CIFE) [46], Conditional Mutual Information Maximization (CMIM) [47], Double Input Symmetrical Relevance (DISR) [48], Joint Mutual Information (JMI) [49], [50], Conditional Redundancy (Condred) [51], and Relief [52].

D. CLASSIFICATION ALGORITHMS

This study used six classification algorithms which include Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Artificial Neural Network (ANN) as well as other well-known algorithms used in previous research, like Classification and Regression Tree (CART), Naive Bayes (NB), and Random Forest (RF) to make comparisons. The classification algorithms would classify the severity level of a given dysarthric speaker based on the acoustic features extracted in terms of mild, moderate, and severe.

More recently, deep learning models like neural network (DNN), convolutional neural network (CNN), and long short-term memory network (LSTM) have been explored for dysarthric speech classification [53]. This research did not adopt deep Learning as the data used for classification is structured data. Moreover, the Nemours database is very small and may not sustain deep learning that requires relatively large data for high classification accuracy [53].

E. PROCEDURES AND TOOLS

In speech analysis, the typical frame lengths range from 20 to 60 milliseconds (ms), with the most commonly chosen frame period being 10ms [54], [55]. For the proposed solution, 60ms were used as the frame length, with 10ms as the frame period. To compute LLD, the frame must contain enough data, and the quasi-stationary of the signal has to be within the length of the frame of the LLD of interest [42].

The procedure for features extraction encompasses three steps. First, the samples pronounced by each speaker are listed into one individual file for each speaker. This file is used as an input for the openSMILE tool which then produces the features for each separate file (the total number of sample files per speaker is 74). Second, each file generated in the first step is then combined into three separate files according to their severity level. Third, the three separate files produced in the second step are next combined into one feature file, including the class types which are severe, moderate, and mild. This file is then used as an input for the features selection step for the classification algorithm.

The various toolbox used for the classification algorithms includes: statistical toolbox which is used to build LDA, and the CART classification methods. The neural network toolbox was used to build the ANN models. Libvm version 3.22 which was developed by Chang and Lin [56] was used to build the SVM classification model (can be downloaded from http://www.csie.ntu.edu.tw/∼cjlin/libsvm). The Naive Bayes code uses the default algorithms that were developed in the MATLAB program, while the code for the Random Forest can be downloaded from https://code.google.com/archive/p/randomforest-matlab/downloads.

F. PROCEDURES AND TOOLS

The performance of each acoustic feature and the feature selection methods meant for the effective classification of dysarthria speech based on the severity of the speech impairment was evaluated in terms of classification accuracy and classification ranking.

1) CLASSIFICATION ACCURACY

To calculate the classification accuracy for each classifier algorithm, the k- fold cross-validation, where k is assigned to 10 [57], [58] was used. It is commonly used to calculate the rate of accuracy of the classifier algorithm for assessing the severity level of dysarthria speakers. In this method, the features extracted from dysarthric speech (including all severity levels) were randomly divided into 10 equal sizes of set samples, where nine partitions were assigned for model training, and the remaining one was used as the test set for model evaluation. For each run, one partition would be used as test data, and the remaining partitions would be used as training data. To ensure that all 10 partitions were used as test data, this procedure was repeated 10 times. For the need to produce one single estimation, the mean score of all the 10 runs was calculated. Compared to a repeated random sub-sampling, the advantage of this method is that, for both training and validation, all observations were used, with each observation being used for validating once only. the average classification accuracy rate was then calculated using the equation below:

$$\text{Average Classification Accuracy Rate} \\ = 100 \times (TNCF/(TNF)) \tag{2}$$

where TNCF is the Total Number of Correctly-testing features, and TNF is the Total Number of Features used

A confusion matrix for the classification of dysarthric speech is created to evaluate the overall classification of dysarthric speech according to the level of severity as well as for speech features categorized as prosodic, voice quality, and spectral. The number of test data for each severity level is 222 data (74 sentences × 3 speakers).

TABLE 1. The Classification Accuracy Based on Classification Algorithms, Features Selection Methods, and Acoustic Features

Classifier algorithm	Feature Selection Algorithm	Acoustic Features: Prosodic: audspec (Loudness)	audspecRasta-Sum	F0final	PCM-RMS Energy	PCM-ZCR	voicingFinalUnclipped	All	Voice Quality: jitterDDP	JitterLocal	shimmerLocal	logHNR	All	Spectral: audspecRasta-Band 1-26	PCM- OSF	All	Cpl: MFCC	All
LDA	jmi	77.16	71.94	65.77	72.98	57.70	66.37	65.14	44.62	49.50	52.24	71.00	66.35	77.98	74.47	78.10	70.72	75.22
	disr	77.16	71.94	65.77	72.98	57.70	66.37	65.14	44.62	49.50	52.24	71.00	66.35	77.98	74.47	78.10	70.72	75.22
	cmim	85.00	72.38	66.22	75.38	56.62	67.13	65.49	43.54	56.11	56.91	70.71	68.62	76.42	74.65	80.32	72.39	76.59
	cife	77.64	70.56	66.53	76.29	57.06	66.33	81.82	46.70	51.94	55.06	71.80	69.04	78.51	77.79	80.17	69.37	75.06
	icap	69.64	69.53	71.02	63.40	58.25	63.94	63.65	44.30	46.86	50.57	72.22	66.38	70.87	71.48	65.77	60.22	62.62
	condred	75.53	74.33	66.67	76.57	56.78	66.80	81.25	40.83	49.55	52.41	71.65	68.80	80.32	78.06	81.51	61.12	74.15
	relief	74.02	69.21	60.80	62.90	58.39	65.02	62.92	40.41	48.36	50.47	68.92	66.82	73.12	71.79	67.70	68.46	76.23
CART	jmi	86.61	65.92	62.32	85.30	77.02	79.90	87.55	53.76	54.07	69.98	68.00	69.04	81.55	95.64	93.86	81.09	75.22
	disr	70.71	59.90	83.18	73.55	55.88	62.49	82.13	49.70	54.64	54.95	74.77	71.78	71.02	68.92	72.51	64.72	74.48
	cmim	75.08	62.60	81.67	75.38	53.60	63.83	80.65	52.69	59.60	62.33	75.21	74.61	72.96	69.84	75.66	63.86	76.57
	cife	74.15	61.24	82.90	73.89	56.90	63.52	84.82	52.10	54.19	62.91	73.89	75.82	75.09	77.04	73.10	65.14	76.57
	icap	71.87	59.78	79.75	63.09	55.41	55.11	79.85	49.43	55.10	52.81	72.98	66.37	68.78	62.92	60.65	61.71	68.20
	condred	71.79	60.50	81.85	70.74	53.28	65.20	83.34	49.53	54.07	52.70	77.95	76.01	71.77	75.53	75.08	62.88	79.41
	relief	69.06	59.28	80.47	60.65	51.18	57.35	80.17	50.29	54.65	54.36	73.11	72.06	66.81	65.77	59.16	64.40	77.02
NB	jmi	85.26	68.02	60.52	78.40	66.22	75.52	86.20	59.18	54.95	64.86	69.65	64.84	74.48	90.85	90.09	75.69	77.77
	disr	72.22	59.92	72.52	63.81	62.49	64.43	71.31	52.42	54.02	57.36	74.02	70.69	73.58	70.88	75.38	65.76	79.89
	cmim	76.29	62.61	72.35	70.26	61.25	68.04	69.68	52.39	57.05	62.61	73.40	73.72	72.68	72.07	77.16	66.99	79.89
	cife	74.15	62.47	72.53	61.81	62.61	66.33	79.58	52.86	55.70	63.50	75.99	72.67	75.97	76.58	81.07	67.27	80.77
	icap	67.86	60.52	79.57	56.05	59.90	59.88	73.72	53.74	50.17	55.97	69.81	68.77	66.20	69.23	65.92	60.95	66.55
	condred	71.31	62.15	71.81	59.31	61.89	65.19	78.96	54.03	56.34	57.09	74.03	72.52	75.84	77.32	80.91	65.92	83.18
	relief	83.15	66.70	56.89	70.59	65.05	80.20	82.14	54.03	52.25	67.71	65.45	64.38	75.38	85.28	80.00	78.99	77.34
ANN	jmi	80.01	67.72	74.62	77.78	53.76	63.05	76.27	43.36	45.17	46.07	77.49	71.87	74.40	77.17	77.81	70.86	77.47
	disr	79.12	72.21	73.07	75.39	55.10	67.13	74.74	44.31	54.80	55.20	78.80	79.73	71.49	76.88	81.20	70.27	75.06
	cmim	77.91	65.77	73.99	75.55	58.11	65.72	83.48	45.94	49.68	54.27	78.97	76.87	73.13	79.58	80.63	72.08	78.67
	cife	71.60	67.11	72.10	64.57	53.16	64.40	72.50	41.01	48.79	46.67	73.86	67.31	72.21	72.38	63.35	60.37	64.31
	icap	73.42	69.08	71.69	75.97	54.38	65.62	85.27	42.30	47.45	50.01	77.61	79.74	75.82	81.22	81.52	64.14	81.96
	condred	78.80	66.22	75.80	64.53	52.27	62.00	74.89	42.02	45.17	44.80	74.30	71.47	73.73	73.29	67.85	70.26	78.21
	relief	89.01	73.58	58.71	80.94	75.40	80.37	87.10	51.94	51.51	68.63	70.26	64.40	82.29	95.64	94.45	79.17	74.91
SVM	jmi	79.42	71.48	69.38	77.79	58.45	66.20	64.42	43.26	48.01	49.70	73.25	69.35	80.21	76.27	79.45	70.41	76.58
	disr	78.68	75.83	65.32	76.72	58.28	65.93	68.93	43.55	48.02	48.20	73.72	72.07	76.57	77.04	81.81	67.43	78.97
	cmim	78.66	70.86	67.27	75.84	60.22	64.55	82.12	44.44	48.35	48.92	73.00	69.63	78.81	81.39	82.42	69.37	74.45
	cife	71.89	70.88	74.31	62.18	55.55	64.23	65.30	43.72	46.56	49.52	72.22	69.38	72.21	69.99	67.44	60.38	59.31
	icap	75.83	74.17	69.85	77.93	56.45	65.76	81.53	43.25	48.36	50.01	74.21	69.54	78.68	81.82	81.06	63.51	77.02
	condred	76.88	71.32	66.10	65.60	56.91	64.56	64.41	43.38	44.45	47.92	70.12	69.24	76.13	72.69	69.35	68.61	74.31
	relief	76.09	62.35	58.86	62.34	50.00	56.61	80.03	50.90	49.86	50.58	61.38	49.83	75.24	92.95	89.92	81.11	63.07
RF	jmi	81.52	69.98	87.83	82.13	68.37	71.04	87.97	55.25	63.50	61.43	81.38	80.63	79.01	78.83	81.99	74.31	83.19
	disr	82.30	71.91	86.63	82.58	66.84	73.73	88.74	58.54	66.05	70.41	80.14	83.03	80.17	77.79	79.72	73.59	82.58
	cmim	82.25	70.87	87.53	81.97	66.97	72.48	91.14	58.56	61.40	69.94	82.29	81.67	79.57	83.94	84.23	74.63	87.82
	cife	80.00	68.63	86.91	71.32	63.81	66.02	86.31	55.86	59.46	61.24	78.53	75.97	77.33	69.97	71.04	72.83	78.38
	icap	81.68	72.67	88.15	80.33	64.43	73.30	91.15	58.40	63.06	60.65	82.00	81.56	80.18	81.08	81.98	70.88	89.64
	condred	77.02	69.52	86.93	72.35	62.46	67.39	87.52	57.06	62.03	64.12	80.47	79.74	74.03	71.78	69.68	72.36	85.27
	relief	92.01	75.09	69.99	84.68	75.55	86.07	89.66	64.69	60.08	71.03	75.51	72.51	83.93	95.80	95.79	83.50	83.48

Cpl=Cepstral, OSF= Other Spectral Features, ZCR= Zero-Crossing Rate.

2) CLASSIFICATION RANKING

To select the best classifier or best feature selection method, the ranking method of friedman's m statistics was used [59], [60]. In this method, each classifier would receive a rank based on the measured accuracy rate of each feature group, where the classifier with the highest accuracy rate of the feature group is assigned rank 1. The classifier with the second-highest accuracy rate is assigned rank 2, and so on. In the case of two classifiers achieving equal accuracy rates, then the rank is divided between them. For example, considering that the accuracy rate of 50%, 60%, 62%, 62%, and 67% was achieved by five different classifiers focusing on different group features, then their ranking score would be 5, 4, 2.5, 2.5, and 1, respectively. The performance of the classifier is then evaluated by using the ranking method, as represented by the following equation, (3), as shown at the bottom of the page, where x_1^n is the set of accuracy rate for the classification algorithms used, and where the number of classification algorithms used is n, and the current value in the x set is i.

For calculating the final ranking of a classifier for the different feature groups, the mean score of each classifier is then calculated. Therefore, the lowest average ranking score would be considered to be the best classifier. The following equation is used to calculate the best classifier based on the average ranking score:

$$Best\ Classifier\ (X_1^n) = MIN\ (Average(Ranking(x_1^n))) \quad (4)$$

where, X_1^n is the set of classification algorithms used, n is the total number of the classifier, and ranking (x_1^n) is the ranking score of the accuracy rate of different feature groups.

IV. RESULTS

The results were analyzed in two parts. The first part focused on the classification results of dysarthric speech. The second part looked at the performance of the acoustic features and the best classification algorithms which can be used to classify the dysarthric speech.

A. CLASSIFICATION ACCURACY

The first evaluation was the classification accuracy of both the acoustic features and the classification algorithms, as shown in Table 1. The classification accuracy ranged from 40.41 (LDA; condred; jitterDDP) to 95.80 (RF; relief; PCM- Other Spectral Features). For the combination of sub-features, the 13 features would be selected all the sub-features were combined. The selection of these features would then be based on the feature selection algorithms.

The results were analyzed based on six classification algorithms, which are LDA, CART, NB, ANN, SVM, and RF. The results would be able to highlight the effectiveness of each classification algorithm based on the feature selection methods and the acoustic features.

Table 2 shows the confusion matrix for classification of dysarthric speech according to the level of severity for the various speech features.

TABLE 2. Confusion Matrix of the Classification Based on Overall Speech Features, Prosodic, Voice Quality, Spectral and Cepstral Features

		Mild	Moderate	Severe
OVERALL SPEECH FEATURES	Mild	176	28	18
	Moderate	33	159	30
	Severe	26	60	136
PROSODIC FEATURES	Mild	186	23	13
	Moderate	36	165	21
	Severe	17	66	139
VOICE QUALITY FEATURES	Mild	172	32	18
	Moderate	48	154	20
	Severe	32	79	111
SPECTRAL FEATURES	Mild	183	24	15
	Moderate	27	164	31
	Severe	26	47	149
CEPSTRAL FEATURES	Mild	162	33	27
	Moderate	23	151	48
	Severe	29	48	145

From Table 2, the confusion matrix of the mild speech was better than severe and moderate due the fact that the mild speech has more common speech features among the speakers.

This is not surprising as many of the existing works on dysarthria have discussed the difficulties in classifying severe dysarthric speech.

In terms of specific speech features, it was found that cepstral features were the least effective in the classification of dysarthric speech according to the level of severity. Prosodic features were found to have a marginal advantage over spectral features in classification of mild and moderate speech, while the spectral features have better classification result for severe dysarthric speech.

B. CLASSIFICATION RANKING

Table 3 reports the ranking score obtained from the classification accuracy. The number of ranking varied from 1 to 42 score, according to the number of classification algorithms. There were six classification algorithms, and each

$$Ranking\ (x_1^n) = \begin{cases} Ranking\ based\ on\ highest, & x_i\ is\ identical\ value \\ \frac{n}{2}, & x_i\ for\ each\ equal\ value \end{cases} \quad (3)$$

TABLE 3. Average Ranking Score for all Classification Algorithms

Classifier algorithm	Feature Selection Algorithm	Acoustic Features																	Average Ranking
		Prosodic							Voice Quality					Spectral			Cpl		
		audspec (Loudness)	audspecRasta-Sum	F0final	PCM-RMS Energy	PCM-ZCR	voicingFinalUnclipped	All	jitterDDP	JitterLocal	shimmerLocal	logHNR	All	audspecRasta-Band 1-26	PCM- OSF	All	MFCC	All	
LDA	jmi	20	9	35	23	23	15	38	27	30	29	33	38	13	26	23	16	29	25.12
	disr	5	7	33	20	29	13	36	33	11	18	34	32	16	25	19	11	23	21.47
	cmim	19	16	32	14	25	17	18	25	24	21	30	29	12	15	20	21	31	21.71
	cife	39	18	26	33	21	32	41	30	38	31	28	36	38	33	38	42	41	33.24
	icap	26	3	31	13	28	14	20	41	29	28	31	30	4	14	12	38	36	23.41
	condred	30	20	38	35	19	26	42	42	33	32	39	35	30	31	35	23	27	31.59
	relief	3	28	37	1	1	4	6	11	19	3	40	28	3	3	3	3	28	13.00
CART	jmi	38	39	7	22	31	36	16	22	17	23	15	20	37	39	30	31	33	26.82
	disr	27	31	10	19	37	33	21	15	7	12	14	12	31	37	26	34	25	23.00
	cmim	28	35	8	21	27	34	12	18	18	10	20	11	23	20	29	30	26	21.76
	cife	34	40	12	34	33	42	24	24	13	26	27	37	39	42	41	37	37	31.88
	icap	35	37	9	26	38	24	14	23	20	27	9	9	35	24	28	36	12	23.88
	condred	40	41	11	40	41	40	22	21	16	24	25	18	40	41	42	32	22	30.35
	relief	4	23	39	8	7	5	10	2	14	7	38	39	24	5	4	6	17	14.82
NB	jmi	32	38	20	32	12	29	32	16	21	16	19	22	28	34	27	29	11	24.59
	disr	23	30	21	28	15	10	33	17	9	11	23	13	32	30	25	26	10	20.94
	cmim	29	32	19	39	11	16	25	13	12	9	12	14	18	22	14	25	9	18.76
	cife	42	36	13	42	17	38	30	12	26	19	37	31	41	38	37	39	38	31.53
	icap	37	34	23	41	14	25	26	9	10	17	18	15	19	17	16	28	6	20.88
	condred	41	42	25	37	24	39	35	14	22	22	32	34	42	40	40	27	19	31.47
	relief	6	26	42	27	8	3	15	10	23	6	41	41	21	6	18	5	20	18.71
ANN	jmi	11	24	15	11	36	35	27	35	41	41	11	19	25	18	24	15	18	23.88
	disr	14	8	18	18	34	12	29	29	15	20	7	7	36	21	13	18	30	19.35
	cmim	18	29	17	17	22	22	13	26	28	25	6	8	29	12	17	13	14	18.59
	cife	36	25	22	30	39	30	31	40	31	40	21	33	34	29	39	41	39	32.94
	icap	31	21	24	15	35	23	11	38	37	33	10	6	20	10	11	33	8	21.53
	condred	15	27	14	31	40	37	28	39	40	42	16	21	27	27	34	19	16	27.82
	relief	2	5	41	6	3	2	8	19	25	5	35	40	2	2	2	4	32	13.71
SVM	jmi	13	11	29	10	18	18	39	36	36	35	24	26	5	23	22	17	24	22.71
	disr	16	1	36	12	20	20	34	32	35	38	22	17	15	19	10	24	13	21.41
	cmim	17	15	30	16	16	28	17	28	34	37	26	23	10	9	7	20	34	21.59
	cife	33	13	16	38	32	31	37	31	39	36	29	25	33	35	36	40	42	32.12
	icap	25	4	28	9	30	21	19	37	32	34	17	24	11	8	15	35	21	21.76
	condred	22	12	34	29	26	27	40	34	42	39	36	27	17	28	33	22	35	29.59
	relief	24	33	40	36	42	41	23	20	27	30	42	42	22	4	5	2	40	27.82
RF	jmi	10	17	2	4	4	9	5	8	2	13	3	4	9	13	8	8	5	7.29
	disr	7	10	6	3	6	6	4	4	1	2	5	1	7	16	21	9	7	6.76
	cmim	8	14	3	5	5	8	2	3	5	4	1	2	8	7	6	7	2	5.29
	cife	12	22	5	25	10	19	9	7	8	14	8	10	14	36	31	10	15	15.00
	icap	9	6	1	7	9	7	1	5	3	15	2	3	6	11	9	14	1	6.41
	condred	21	19	4	24	13	11	7	6	4	8	4	5	26	32	32	12	3	13.59
	relief	1	2	27	2	2	1	3	1	6	1	13	16	1	1	1	1	4	4.88

Cpl=Cepstral, OSF= Other Spectral Features, ZCR= Zero-Crossing Rate.

classification algorithm contained seven feature selection methods, amounting to 42 ranking scores (six classification algorithms × seven features selection methods = 42 ranking score).

The average ranking score depicted in Table 3 showed that the Random Forest (RF) algorithms with the "Relief" feature selection method had obtained the highest performance for classifying the severity level of dysarthric speech, with an average ranking score of 4.88. The second and third highest performing algorithm for classifying the severity level of dysarthric speech was the RF algorithms, with the "cmim" and "icap" features selection method, with an average ranking score of 5.29 and 6.41 respectively. Table 3 also illustrated that the RF algorithms had obtained the highest performance for classifying the severity level of dysarthric speech. The RF algorithms were used to identify the most relevant features for the pathophysiology of parkinsonian dysarthria. It had also obtained the highest classification accuracy for classifying Parkinson's disease among healthy speakers [61].

When the present results were compared to the results in [62], it was found that pronunciation and voice quality for the binary classification of dysarthric speech was varied, based on acoustic features.

The binary classification of speech intelligibility was 73.5% for unweighted average recall, and 72.8% for weighted average recall for the SVM classification. This highlighted the classifier's best performance. The results from this study, as shown in Table 1 above indicates that the SVM classification algorithms had obtained an average classification accuracy of 71.96%. The results of this study were computed as average classification accuracy rather than the best recognition accuracy because seven feature selection algorithms were used for each classifier, with the highest classification accuracy being 78.97%. These results suggest that the RF algorithms had obtained a high performance, as previously described.

Narendra and Alku [63] used almost the same acoustic features as the current study, including glottal features for classifying dysarthric speech and the speech of non-impaired speakers. The classification accuracy detected by Narendra and Alku [63] was 94.29% classification accuracy when using the SVM classification algorithms, and 89.64% classification accuracy when using the RF classification algorithm. The difference between the results derived from the current study and those of Narendra and Alku [63] can be attributed to the fact that the current study had classified the dysarthric speech and the speech of non-impaired speakers into words, non-words, and sentences.

C. CLASSIFICATION RANKING OF THE ACOUSTIC FEATURES

The main goal of this analysis is to show the effectiveness of the sub acoustic features in classifying based on the severity of the speech impairment.

1) PROSODIC ACOUSTIC FEATURES

Table 4 shows the ranking scores which varied from 1 to 7, according to the highest classification accuracy of the number of sub acoustic features used in this study. As noted in Table 4, the best prosodic acoustic feature to be used for classifying the dysarthric speech is the audspec (Loudness), with the lowest average ranking score of 2.24. The results also showed that the combination of the prosodic acoustic features had the second-highest score, with a 2.26 average ranking score. This was followed by the F0-Final which ranked third, with an average ranking score of 3.38.

TABLE 4. Average Classification Ranking for Prosodic Acoustic Features

	audspec (Loudness)	AUDSPECRASTA-SUM	F0final	PCM-RMS Energy	PCM-Zero-Crossing Rate (ZCR)	voicingFinal Unclipped	All
Average Ranking	2.24	4.71	3.38	3.71	6.64	5.05	2.26

TABLE 5. Average Ranking Score for Voice Quality Acoustic Features

	jitterDDP	JITTERLOCAL	shimmerLocal	logHNR	All
Average Ranking	4.81	3.95	3.02	1.19	2.02

TABLE 6. Average Ranking Score for Spectral Acoustic Features

	audspecRasta -Band 1-26	PCM- OHER SPECTRAL FEATURES	All
Average Ranking	2.19	1.98	1.83

2) VOICE QUALITY ACOUSTIC FEATURES

Table 5 shows the average ranking score of the voice quality in sub acoustic features. Here, the best voice quality acoustic features for the classification of dysarthric speech was the loghnr voice quality acoustic features, with the lowest average ranking score of 1.19.

This showed that the combination of voice quality acoustic features can be a competitor to sub voice quality acoustic features. This is because it was ranked second, followed by lorHNR with an average ranking score of 2.00. The shimmerlocal ranked third, with an average ranking score of 3.02. The average ranking score was computed as the average of the ranking score obtained, based on the classification accuracy for each classification algorithm and feature selection method used in this experiment.

3) SPECTRAL ACOUSTIC FEATURES

The performance of the spectral acoustic features is presented in table 6. It depicts the best spectral acoustic features for the classification of dysarthric speech. This was attributed to the combination of all the spectral acoustic features which

TABLE 7. **Average Ranking Score of Overall Acoustic Features**

Verification Method	Acoustic Features																
	Prosodic							Voice Quality					Spectral			Cpl	All
	audspec (Loudness)	audspecRasta-Sum	F0final	PCM-RMS Energy	PCM-ZCR	voicingFinalUnclipped	All	jitterDDP	JitterLocal	shimmerLocal	logHNR	All	audspecRasta-Band 1-26	PCM- OSF	All	MFCC	
Average Ranking Score	4.83	10.57	7.24	8.00	13.76	11.19	4.48	16.79	15.88	14.48	6.17	8.40	6.17	5.17	4.95	9.98	4.95

Cpl=Cepstral, OSF= Other Spectral Features, ZCR= Zero-Crossing Rate.

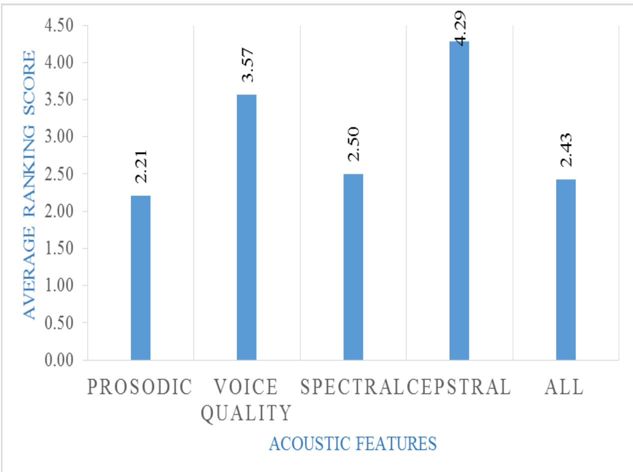

FIGURE 2. **Average ranking score for all acoustic features groups.**

had achieved the lowest average ranking score of 1.83. The results also showed that the pcm-other spectral features of the spectral acoustic features were ranked second, followed by the combination of the spectral acoustic features which had achieved an average ranking score of 1.98. The audspecrasta-band 1-26 was ranked third, with an average ranking score of 2.19.

D. DISCUSSION

This part of the analysis focused on the performance of four acoustic features which were prosodic, voice quality, spectral, and cepstral features. In each feature, the combination of the sub-features was selected for making comparisons.

The classifier which had achieved the best performance for classifying the dysarthric speech according to the severity of the impairment was the prosodic acoustic features, with an average ranking score of 2.21, as shown in Fig. 2. It appears that the combination of the acoustic features had enabled it to be second in performance, with an average ranking score of 2.40. The third-ranking feature was the spectral acoustic features with an average ranking score of 2.50.

The binary classification of speech intelligibility based on prosodic acoustic features was 71.3% and 75.5%, for unweighted and weighted average recalls, using the SVM classification algorithms. The LDA classification algorithm had obtained 65.3% for unweighted average recalls, and 69.1% for weighted average recalls. The results from this study, as listed in Table 1 above, showed that the prosodic acoustic features had obtained average classification accuracy of 72.39 % and 72.55% when using the SVM and LDA classification algorithms.

For voice quality features, the binary classification of speech intelligibility was 66.3% and 66.0%, respectively, for the unweighted and weighted average recall, using the SVM classification algorithms. The LDA classification algorithm had obtained 68.9% for the unweighted average recall and 71.7% for the weighted average recall. The results from this study, as listed in Table 1, showed that the voice quality acoustic features had obtained 67.39% and 67.86%, for the average classification accuracy when using the SVM and LDA classification algorithms, respectively.

The overall acoustic features were listed together so as to show the comparison of all the acoustic features used for classifying the severity level of the dysarthric speech. The comparison includes all of the acoustic features used in this study. The sub-features, as well as the combination of the sub-features, were also included. The main objective of the analysis was to report on the best performance of the acoustic features, for classifying the severity level of dysarthric speech. The total number of acoustic features used was 13 which included all the features that were discussed in the previous section above.

The best performance shown by the overall acoustic features which were used to classify the dysarthric speech, based on the severity of impairment, was the combination of prosodic acoustic features. These had obtained the lowest average ranking score among the overall features as shown in Table 7. The combination of prosodic acoustic features had an average ranking score of 4.48.

The second-best performance was obtained by the sub-features of prosodic acoustic features, namely the audspec (Loudness), with an average ranking score of 4.83. The loudness acoustic feature was also considered to be one of the acoustic features used in the perceptual (subjective) studies to identify voice quality in dysarthric speech [64].

Both the combination of all the features of spectral and the overall combination of all the features, had obtained an average score of 4.95. This puts them in the third-highest performance for classifying the severity level of dysarthric speech or dysarthria speakers. As such, it can be said that prosodic features, voice quality, spectral and cepstral acoustic features, all have a significant impact on the classification of dysarthric speech and its severity level. The combination of all the acoustic features had achieved a high average ranking score for classifying the severity level of acoustic features. For example, the combination of acoustic features had been noted to achieve the third-highest average ranking score in all the previous results as well as in the overall acoustic feature analysis, where it had achieved the first and third highest performance among all the acoustic features used. This is shown in Table 7.

V. CONCLUSION

This study has presented the findings of the classification accuracy of dysarthric speech based on the severity of the impairment by examining the acoustic features and feature selection methods. It was found that the different combinations of acoustic features, feature selection methods, and classification algorithms had produced different classification accuracy. This outcome thus strengthens the notion that there is no one best method for improving the classification accuracy of an ASR system. In this study, the best classification accuracy was generated when we combined all the prosodic acoustic features of dysarthric speech. This means that all the prosodic acoustic features were relevant in classifying the dysarthric speech, based on the severity of the impairment.

In our study, the combination of Random Forest as the classifier, Relief as the feature selection method, and PCM-Other Spectral features had resulted in the highest classification accuracy. On the other hand, the combination of LDA as the classifier, Condred as the feature selection method, and jitterDDP as the acoustic feature, had resulted in the lowest classification accuracy. It appears that the combination which provided the highest classification accuracy was only applicable for the classification of dysarthric speech, based on the severity of the impairment. It may not produce the same result when sued in combination with other forms of speech or other speech databases.

This research has several merits that add knowledge to the classification of dysarthric speech according to the level of severity. First of all, this research has identifies the features that can work in most of the classifiers. Secondly, it looks at the importance of feature selection in the classification of dysarthric speech. Finally, it looks at the best combination that gives the best classification accuracy in the classification of dysarthric speech according to the level of severity. However, there are a number of disadvantages of this works particularly on the use of Nemours as the database for this study. Nemours is a very small database as compared to other databases though it focused only on spastic dysarthria. The other disadvantage of this work is not adopting the state of the art classifiers such as deep learning.

VI. FUTURE WORKS

The limitation of this work can be the opportunity for future research direction including the use of several databases of dysarthric speech to confirm the importance of the features and feature selection methods in the classification of dysarthric speech according to the level of severity. There is also the opportunity for using the state of the art classifiers such as the deep learning classifiers for the classification of dysarthric speech according to the level of severity.

REFERENCES

[1] P. Green, J. Carmichael, A. Hatzis, P. Enderby, M. S. Hawley, and M. Parker, "Automatic speech recognition with sparse training data for dysarthric speakers," in *Proc. Eur. Conf. Speech Commun. Technol.*, 2003, pp. 1–4.

[2] E. K. Hanson and S. K. Fager, "Communication supports for people with motor speech disorders," *Topics Lang. Disorders*, vol. 37, no. 4, pp. 375–388, 2017.

[3] G. E. Castillo and D. F. Lovey, "A modern approach to dysarthria classification," in *Proc. 25th Annu. Int. Conf. Eng. Med. Biol. Soc.*, Cancun, Mexico, Sep. 2013, pp. 2257–2260.

[4] M. Hasegawa-Johnson, J. Gunderson, A. Perlman, and T. Huang, "HMM-based and SVM-based recognition of the speech of talkers with spastic dysarthria," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Toulouse, France, May 2006, p. 3.

[5] F. L. Darley, A. E. Aronson, and J. R. Brown, "Clusters of deviant speech dimensions in the dysarthrias," *J. Speech Hearing Res.*, vol. 12, no. 3, pp. 462–496, Sep. 1969, doi: 10.1044/jshr.1203.462.

[6] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," *J. Speech Hearing Res.*, vol. 12, no. 2, pp. 246–269, Jun. 1969, doi: 10.1044/jshr.1202.246.

[7] R. D. Kent, H. K. Vorperian, J. F. Kent, and J. R. Duffy, "Voice dysfunction in dysarthria: Application of the multi-dimensional voice program," *J. Commun. Disorders*, vol. 36, no. 4, pp. 281–306, 2003, doi: 10.1016/S0021-9924(03)00016-9.

[8] K. P. Connaghan and R. Patel, "The impact of contrastive stress on vowel acoustics and intelligibility in dysarthria," *J. Speech, Lang., Hearing Res.*, vol. 60, no. 1, pp. 38–50, Jan. 2017.

[9] P. Kayasith, T. Theeramunkong, and N. Thubthong, "Recognition rate prediction for dysarthric speech disorder via speech consistency score," in *Proc. 9th Pacific Rim Int. Conf. Artif. Intell.*, Q. Yang and G. Webb, Eds., Guilin, China. Berlin, Germany: Springer, Aug. 2006, pp. 885–889.

[10] S. D. Barreto and K. Z. Ortiz, "Speech intelligibility in dysarthrias: Influence of utterance length," *Folia Phoniatrica Logopaedica*, vol. 72, no. 3, pp. 202–210, 2020.

[11] P. Enderby, "Frenchay dysarthria assessment," *Int. J. Lang. Commun. Disorders*, vol. 15, no. 3, pp. 165–173, 1980.

[12] P. Enderby, "Frenchay dysarthria assessment," *Brit. J. Disorders Commun.*, vol. 15, no. 3, pp. 165–173, 1980.

[13] K. M. Yorkston, D. R. Beukelman, and C. Traynor, *Computerized Assessment of Intelligibility of Dysarthric Speech*. Tigard, Oregon: C.C. Publications, 1984.

[14] Y. Kim, R. D. Kent, and G. Weismer, "An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria," *J. Speech, Lang. Hearing Res.*, vol. 54, no. 2, pp. 417–429, 2011, doi: 10.1044/1092-4388(2010/10-0020).

[15] I. Calvo, P. Tropea, M. Viganò, M. Scialla, A. B. Cavalcante, M. Grajzer, M. Gilardone, M. Corbo, "Evaluation of an automatic speech recognition platform for dysarthric speech," *Folia Phoniatr Logop*, 2020, doi: 10.1159/000511042.

[16] A. Cavalcante and M. Grajzer, "Proof-of-concept evaluation of the mobile and personal speech assistant for the recognition of disordered speech," *Int. J. Adv. Intell. Syst.*, vol. 9, p. 589, May 2016.

[17] J. M. Liss, L. White, S. L. Mattys, K. Lansford, A. J. Lotto, S. M. Spitzer, and J. N. Caviness, "Quantifying Speech Rhythm Abnormalities in the Dysarthrias," *J. Speech, Lang., Hearing Res.*, vol. 52, no. 5, pp. 1334–1352, 2019, doi: 10.1044/1092-4388(2009/08-0208).

[18] M. N. S. Niimi, ''Speaking rate and its components in dysarthric speakers,'' *Clin. Linguistics Phonetics*, vol. 15, no. 4, pp. 309–317, Jan. 2001, doi: 10.1080/02699200010024456.
[19] M. O. S. Paja and T. H. Falk, ''Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech,'' in*Proc. Annu. Conf. Int. Speech Commun. Assoc.*, 2012, pp. 1–4.
[20] R. D. Kent, G. Weismer, J. F. Kent, and J. C. Rosenbek, ''Toward phonetic intelligibility testing in dysarthria,'' *J. Speech Hearing Disorders*, vol. 54, no. 4, pp. 482–499, Nov. 1989.
[21] T. H. Falk, W.-Y. Chan, and F. Shein, ''Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility,'' *Speech Commun.*, vol. 54, no. 5, pp. 622–631, Jun. 2012.
[22] T. H. Falk and W.-Y. Chan, ''Temporal dynamics for blind measurement of room acoustical parameters,'' *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 978–989, Apr. 2010, doi: 10.1109/TIM.2009.2024697.
[23] J. I. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, ''Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters,'' *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1943–1953, Oct. 2006, doi: 10.1109/TBME.2006.871883.
[24] J. I. Godino-Llorente, V. Osma-Ruiz, N. Sáenz-Lechón, P. Gómez-Vilda, M. Blanco-Velasco, and F. Cruz-Roldán, ''The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders,'' *J. Voice*, vol. 24, no. 1, pp. 47–56, Jan. 2010, doi: 10.1016/j.jvoice.2008.04.006.
[25] J. P. Teixeira and P. O. Fernandes, ''Jitter, shimmer and HNR classification within gender, tones and vowels in healthy voices,'' *Procedia Technol.*, vol. 16, pp. 1228–1237, Dec. 2014, doi: 10.1016/j.protcy.2014.10.138.
[26] M. S. De Bodt, M. E. Hernández-Daz Huici, and P. H. Van De Heyning, ''Intelligibility as a linear combination of dimensions in dysarthric speech,'' *J. Commun. Disorders*, vol. 35, no. 3, pp. 283–292, 2002, doi: 10.1016/S0021-9924(02)00065-5.
[27] D. Deterding, ''The measurement of rhythm: A comparison of singapore and british english,'' *J. Phonetics*, vol. 29, no. 2, pp. 217–230, Apr. 2001, doi: 10.1006/jpho.2001.0138.
[28] S. Stuntebeck, ''Acoustic analysis of the prosodic properties of ataxic speech,'' M.S. thesis, Dept. Commun. Sci. Disorders, Fac. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, 2002, unpublished.
[29] G. Weismer, R. Martin, and R. Kent, ''Acoustic and perceptual approaches to the study of intelligibility,'' *Intell. Speech Disorders*, vol. 16, pp. 67–118, Apr. 1992.
[30] F. Huei-Mei LiuChin-Hsing Tseng, ''Perceptual and acoustic analysis of speech intelligibility in mandarin-speaking young adults with cerebral palsy,'' *Clin. Linguistics Phonetics*, vol. 14, no. 6, pp. 447–464, Jan. 2000.
[31] J. F. Kent, R. D. Kent, J. C. Rosenbek, G. Weismer, R. Martin, R. Sufit, and B. R. Brooks, ''Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis,'' *J. Speech, Lang., Hearing Res.*, vol. 35, no. 4, pp. 723–733, Aug. 1992.
[32] Y. Kim, G. Weismer, R. D. Kent, and J. R. Duffy, ''Statistical models of F2 slope in relation to severity of dysarthria,'' *Folia Phoniatrica Logopaedica*, vol. 61, no. 6, pp. 329–335, 2009.
[33] P. A. McRae, K. Tjaden, and B. Schoonings, ''Acoustic and perceptual consequences of articulatory rate change in parkinson disease,'' *J. Speech, Lang., Hearing Res.*, vol. 45, no. 1, pp. 35–50, Feb. 2002.
[34] K. Tjaden and G. E. Wilding, ''Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings,'' *J. Speech, Lang., Hearing Res.*, vol. 47, no. 4, pp. 766–783, Aug. 2004.
[35] G. Weismer, J.-Y. Jeng, J. S. Laures, R. D. Kent, and J. F. Kent, ''Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders,'' *Folia Phoniatrica Logopaedica*, vol. 53, no. 1, pp. 1–18, 2001.
[36] R. D. Kent, J. F. Kent, J. R. Duffy, J. E. Thomas, G. Weismer, and S. Stuntebeck, ''Ataxic dysarthria,'' *J. Speech, Lang., Hearing Res.*, vol. 43, no. 5, pp. 1275–1289, 2002.
[37] V. F. M. Mujumdar and R. Kubichek, ''Design of a dysarthria classifier using global statistics of speech features,'' in *Int. Conf. Acoust., Speech Signal Process.*, Dallas, TX, USA, Mar. 2010, pp. 582–585.
[38] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, and H. T. Bunnell, ''The Nemours database of dysarthric speech,'' in *Proc. 4th Int. Conf. Spoken Lang.*, Oct. 1996. pp. 1962–1965.
[39] K. T. Mengistu and F. Rudzicz, ''Adapting acoustic and lexical models to dysarthric speech,'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Dec. 2011, pp. 4924–4927.
[40] T. M. Khoshgoftaar, M. Golawala, and J. V. Hulse, ''An empirical study of learning from imbalanced data using random forest,'' in *Proc. 19th IEEE Int. Conf. Tools with Artif. Intell. (ICTAI)*, Oct. 2007, pp. 310–317.
[41] S. H. Samsudin, H. Z. M. Shafri, A. Hamedianfar, and S. Mansor, ''Spectral feature selection and classification of roofing materials using field spectroscopy data,'' *J. Appl. Remote Sens.*, vol. 9, no. 1, May 2015, Art. no. 095079.
[42] F. Eyben, *Real-Time Speech and Music Classification by Large Audio Feature Space Extraction*. Cham, Switzerland: Springer, 2014.
[43] B. Schuller, *Intelligent Audio Analysis*. Berlin, Germany: Springer, 2013.
[44] L. I. Kuncheva, ''A stability index for feature selection,'' in *Proc. Artif. Intell. Appl.*, 2007, pp. 1–5.
[45] A. Jakulin, ''Machine Learning based on attribute interactions,'' Ph.D. dissertation, Dept. Comput. Inf. Sci., Artif. Intell. Lab., Univ. Ljubljana, Ljubljana, Slovenia, 2005.
[46] D. Lin and X. Tang, ''Conditional infomax learning: An integrated framework for feature extraction and fusion,'' in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 68–82.
[47] F. Fleuret, ''Fast binary feature selection with conditional mutual information,'' *J. Mach. Learn. Res.*, vol. 5, pp. 1531–1555, Nov. 2004.
[48] P. E. Meyer and G. Bontempi, ''On the use of variable complementarity for feature selection in cancer classification,'' in *Proc. Workshops Appl. Evol. Comput.*, 2006, pp. 91–102.
[49] C. Parmar, P. Grossmann, D. Rietveld, M. M. Rietbergen, P. Lambin, and H. J. W. L. Aerts, ''Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer,'' *Frontiers Oncol.*, vol. 5, p. 272, Dec. 2015.
[50] W. Gao, L. Hu, and P. Zhang, ''Class-specific mutual information variation for feature selection,'' *Pattern Recognit.*, vol. 79, pp. 328–339, Jul. 2018.
[51] G. Brown, A. Pocock, M.-J. Zhao, and M. Lujan, ''Conditional likelihood maximisation: A unifying framework for information theoretic feature selection,'' *J. Mach. Learn. Res.*, vol. 13, pp. 27–66, Jun. 2012.
[52] K. Kira and L. A. Rendell, ''A practical approach to feature selection,'' in *Machine Learning Proceedings*. Amsterdam, The Netherlands: Elsevier, 1992, pp. 249–256.
[53] A. A. Joshy, ''Automated dysarthria severity classification using deep learning frameworks,'' in *Proc. Eur. Signal Process. Conf.*, Amsterdam, The Netherlands, 2020, pp. 116–120.
[54] L. R. Rabiner, ''A tutorial on hidden Markov models and selected applications in speech recognition,'' *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Dec. 1989.
[55] V. Young and A. Mihailidis, ''Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review,'' *Assistive Technol.*, vol. 22, no. 2, pp. 99–112, Jun. 2010, doi: 10.1080/10400435.2010.483646.
[56] C.-C. Chang and C.-J. Lin, ''LIBSVM: A library for support vector machines,'' *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.
[57] H. Ishibuchi and Y. Nojima, ''Repeated double cross-validation for choosing a single solution in evolutionary multi-objective fuzzy classifier design,'' *Knowl.-Based Syst.*, vol. 54, pp. 22–31, Dec. 2013.
[58] G. McLachlan, K.-A. Do, and C. Ambroise, *Analyzing Microarray Gene Expression Data*, vol. 422. Hoboken, NJ, USA: Wiley, 2005.
[59] H. R. Neave and P. L. Worthington, *Distribution-Free Tests*. London, U.K.: Unwin Hyman, 1988.
[60] P. B. Brazdil and C. Soares, ''A comparison of ranking methods for classification algorithm selection,'' in *Proc. Eur. Conf. Mach. Learn.*, May 2000, pp. 63–75.
[61] A. Rueda, J. C. Vásquez-Correa, C. D. Rios-Urrego, J. R. Orozco-Arroyave, S. Krishnan, and E. Nöth, ''Feature representation of pathophysiology of parkinsonian dysarthria,'' in *Proc. Interspeech*, 2019, pp. 3048–3052.
[62] J. Kim, N. Kumar, A. Tsiartas, M. Li, and S. S. Narayanan, ''Automatic intelligibility classification of sentence-level pathological speech,'' *Comput. Speech Lang.*, vol. 29, no. 1, pp. 132–144, Jan. 2015, doi: 10.1016/j.csl.2014.02.001.
[63] N. P. Narendra and P. Alku, ''Dysarthric speech classification from coded telephone speech using glottal features,'' *Speech Commun.*, vol. 110, pp. 47–55, Jul. 2019.
[64] L. Hartelius, B. Runmarker, and O. Andersen, ''Prevalence and characteristics of dysarthria in a multiple-sclerosis incidence cohort: Relation to neurological data,'' *Folia Phoniatrica Logopaedica*, vol. 52, no. 4, pp. 160–177, 2000.

BASSAM ALI AL-QATAB (Member, IEEE) was born in Taiz, Yemen, in 1980. He received the B.S. degree in computer science from Babylon University, Iraq, in 2002, and the master's degree in software engineering and the Ph.D. degree from the University of Malaya, Kuala Lumpur, Malaysia. in 2010 and 2020, respectively.

From 2009 to 2010 and in 2012, he was a Research Assistant with the Faculty of Computer Science and Information Technology. Since 2009, he has been a Researcher with the Department of Software Engineering, Faculty of Computer Science, University of Malaya. He has published his work in many of prestigious international journals. His research interests include software development, automatic speech recognition, deep learning methods, and signal processing.

Dr. Al-Qatab was an IEEE Young Professionals.

MUMTAZ BEGUM MUSTAFA received the B.Sc. degree in software engineering from University Putra Malaysia (UPM), in 2002, and the M.Sc. degree in software engineering and the Ph.D. degree in computer science from University of Malaya (UM), in 2006 and 2012, respectively. She is currently an Associate Professor with UM. She has undertaken several Speech Synthesis research and holds grants from Ministry of Higher Education. Her research and development of the HMM-based Malay speech synthesis system received The Most Prestigious Award (MPA) for Excellent Research 2012 from MIMOS Bhd. the National R&D Centre in ICT and received gold medals for several national level competitions. She has published several articles in prestigious speech conferences and journals. She has established network with a number of the International Speech Research Laboratory, Japan and Singapore. She supervises a group of master's and Ph.D. students working on speech synthesis, speech recognition, and speech signal processing. Her research interests include emotional speech synthesis and speech assistive tools for disabled individuals.