Dysarthria Severity detection Using Recurrent and Convolutional Neural Networks

Amina Hamza*
Speech and Signal Processing Lab University of Sciences and Technology, Houari Boumediene Algiers, Algeria
Email: ahamza1@usthb.dz

Djamel Addou
Speech and Signal Processing Laboratory University of Sciences and Technology, Houari Boumediene Algiers, Algeria
Email: daddou@usthb.dz

Received: 01/12/2024 **Accepted:** 13/12/2024 **Published:** 30/12/2024

Abstract:

The diagnosis and monitoring of dysarthria, a speech disorder caused by neuro-motor problems that affect articulation, depend on a precise evaluation of its severity. When creating automated systems to identify and categorize dysarthric speech, accurate severity classification is essential. Using neural network models, specifically recurrent neural networks (RNN) and convolutional neural networks (CNN), this paper offers a thorough investigation of how to distinguish dysarthric voices among a collection of normal voice samples and categorize the severity of dysarthria. Among the features used in the study are voice quality, prosodic parameters, formants, Mel frequency cepstral coefficients (MFCC), and spectrograms. Comparing the ability of convolutional networks and reccurents to identify abnormalities in normal data, as well as the hybrid model that combines convolutional and reccurent neural networks (CRNN), is our goal. The Nemours corpus database is used to assess these neural network models' performances. Notably, 99.8% is the highest classification accuracy attained with this corpus.

Keywords: Dysarthria classification - CNN - RNN - CRNN - Acoustic parameters - Automatic speech assessment.

*Corresponding author: Amina Hamza

اكتشاف شدة عسر التلفظ باستخدام الشبكات العصبية المتكررة والتلافيفية

الملخص:

يعتمد تشخيص ومراقبة اضطراب التلفظ، وهو اضطراب في الكلام ناجم عن مشاكل عصبية حركية تؤثر على النطق، على تقييم دقيق لدرجة اضطرابه القصوى. عند إنشاء أنظمة آلية لتحديد وتصنيف الكلام عند اضطراب التلفظ، فإن التصنيف الدقيق لدرجة الاضطراب أمر ضروري. باستخدام نماذج الشبكة العصبية، وتحديدًا الشبكات العصبية المتكررة (RNN) والشبكات العصبية التلافيفية (CNN)، تقدم هذه الورقة تحقيقًا شاملاً حول كيفية التمييز بين الأصوات ذات الاضطراب التلفظي بين مجموعة من عينات الاصوات الطبيعية وتصنيف شدة اضطراب التلفظ. من بين السمات المستخدمة في الدراسة، استعملنا جودة الصوت، والمؤشرات النغمية، والبواني الصوتية، ومؤشرات التردد الميلاني (MFCC)، والمنحنيات الطيفية. إن مقارنة قدرة الشبكات التلافيفية والشبكات المتكررة على تحديد الشذوذ في البيانات الطبيعية، بالإضافة إلى النموذج الهجين الذي يجمع بين الشبكات العصبية التلافيفية والشبكات المتكررة (CRNN)، يعتبر هدفنا من هذه الدراسة. قمنا باستغلال قاعدة البيانات Nemours لتقييم أداء نماذجنا للشبكات العصبية. في النهاية لاحظنا نسبة 99.8% التي تحصلنا عليها تعتبر أعلى دقة تصنيف تم تحقيقها باستخدام هذه القاعدة.

كلمات مفتاحية: تصنيف اضطراب التلفظ - CNN - RNN - CRNN - المؤشرات الصوتية - العلاج الآلي للكلام.

Détection de la gravité de la dysarthrie à l'aide de réseaux neuronaux récurrents et convolutionnels

Résumé:

Le diagnostic et le suivi de la dysarthrie, un trouble de la parole causé par des problèmes neuromoteurs qui affectent l'articulation, dépendent d'une évaluation précise de sa gravité. Lors de la création de systèmes automatisés pour identifier et catégoriser la parole dysarthrique, une classification précise de la gravité est essentielle. En utilisant les modèles de réseaux neuronaux, en particulier les réseaux neuronaux récurrents (RNN) et les réseaux neuronaux convolutionnels (CNN), cet article propose une étude approfondie de la manière de distinguer les voix dysarthriques parmi une base d'échantillons de voix normales et de catégoriser la gravité de la dysarthrie. Parmi les caractéristiques utilisées dans l'étude figurent la qualité de la voix, les paramètres prosodiques, les formants, les coefficients cepstraux Mel (MFCC) et les spectrogrammes. Nous Comparons la capacité des réseaux convolutionnels et récurrents à identifier les anomalies dans les données normales, ainsi que le modèle hybride qui combine les réseaux neuronaux convolutionnels et récurrents (CRNN), qui est notre objectif. La base de données du corpus Nemours est utilisée pour évaluer les performances de ces modèles de réseaux neuronaux. Nous avons enregistré un taux de 99,8 % de précision de classification qui est le taux le plus élevée atteint avec ce corpus.

Mots clés: Classification de la dysarthrie - CNN - RNN - CRNN - Paramètres acoustiques - Evaluation automatique de la parole.

INTRODUCTION

The integrity of the central or peripheral nervous system is necessary for the control of the vocal apparatus. Because the speech muscles are paralyzed, weak, or poorly coordinated, this condition makes oral communication difficult (Freed, 2018). In particular, many neurological conditions, such as cerebral palsy and neurodegenerative illnesses like Parkinson's disease, are linked to dysarthria (Rudzicz, 2011). Due to the impairment of motor speech functions, this condition causes irregular speech patterns, fluctuating speech rates, reduced audibility, and imprecise articulation. When taken as a whole, these elements degrade speech quality (Palmer & Enderby, 2007). In order to track patient progress and plan speech therapy, it is essential to evaluate speech intelligibility in order to determine the severity of dysarthria (Schu et al., 2023). However, trained speech-language pathologists' subjective assessments can be expensive and unreliable, which highlights the need for an automated system to classify the severity of dysarthria. Additionally, people with dysarthria often have physical disabilities and difficulties with muscle coordination, which can limit their ability to use interactive applications that use touch screens or keyboards. This highlights the importance of automatic speech recognition (ASR) systems, since correctly classifying the degree of dysarthria can improve the performance of these systems, as demonstrated by (Martinez et al., 2015).

Various techniques have been put proposed in the literature to objectively evaluate the intelligibility of dysarthric speech. Automatic speech recognition (ASR) systems trained on both dysarthric and normal speech data are one example of such methods. The goal of these systems is to improve accuracy and categorize the severity of dysarthria. While these systems have benefits including automated intelligibility evaluation, the flexibility to adapt to varying severity levels, and the potential for accuracy gains through a variety of data, they also have drawbacks. These difficulties include restricted data availability, the inability to properly capture the subtle components of dysarthric speech that are necessary for an effective diagnosis and therapy, and difficulties in addressing the unique characteristics of dysarthric speech (Seong et al., 2016). Analyzing dysarthric speech signals acoustically is another strategy. This entails examining a variety of acoustic characteristics, such as Mel Frequency Cepstral Coefficients (MFCC), speech pace, prosodic qualities, fundamental frequency, and formant frequencies (Al-Qatab & Mustapha, 2021), and i-vectors by employing techniques such as v-support vector regression (vSVR) (Bai et al., 2013) or probabilistic linear discriminant analysis (PLDA) (Hernandez et al., 2020) to identify patterns that may indicate dysarthria. Acoustic analysis records spectral information, models intricate correlations in data, and provides insightful information about acoustic characteristics. The potential for incompletely capturing the entire spectrum of dysarthria, variations in the efficacy of acoustic characteristics across various speech patterns, and a heavy dependence on data representativeness and quality for machine learning models are some of its limitations. On the other hand, deep learning algorithms, such as long short-term memory (LSTM) networks, convolutional neural networks (CNN), and deep neural networks (DNN) (Deng & Platt, 2014), have been studied for the purpose of dysarthric automatic speech recognition (ASR) using dysarthric datasets like TORGO (Deng & Platt, 2014), Nemours, and UASpeech. These machine learning algorithms show mastery of speech tasks by learning hierarchical representations that include speech patterns and acoustic features, and they are robustly adaptive to complex dysarthric speech data. They do, however, have several drawbacks, such as the requirement for large datasets, sensitivity to building and hyperparameter choices, and computing requirements that could make them difficult to apply in contexts with restricted resources (Joy et al., 2017). Furthermore, research has looked at perceptual assessment techniques, which use speech-language pathologists or trained listeners to gauge how intelligible dysarthric speech is. In order to measure the severity of dysarthria, these evaluations usually use rating scales or scoring systems. Overall, the majority of existing research has advanced the development of objective methods for determining the intelligibility of dysarthric speech and categorizing it. More accurate and consistent assessments are now possible because to these advancements, which also make it easier to track patients' progress and create customized speech therapy programs (Mehrish et al., 2023).

This framework's objective is to investigate, apply, and select the best acoustic properties, such as voice quality, formants, prosodic parameters, mel-spectrograms, and Mel Frequency Cepstral Coefficients (MFCCs). Dysarthric speech is often assessed and categorized using these characteristics since they can capture significant aspects of the condition. This drive results from these parameters' advantages: Formants are used to provide information about articulatory coordination, prosodic parameters are used to reflect irregularities in speech rhythm, voice quality parameters are used to evaluate vocal qualities, mel-frequency cepstral coefficients (MFCCs) are used to capture spectral information, and spectrograms provide a comprehensive view of the temporal and spectral aspects of the speech signal. Making use of these characteristics enables a thorough depiction of dysarthric speech, which promotes more accurate classification and evaluation of intelligibility. To create a classification system for the severity of dysarthria, these characteristics are combined with convolutional neural networks (CNN), recurrent neural networks (RNN), and their hybrid model. The study was carried out on the Nemours dysarthric speech database. The selection of these techniques was justified by the goal of using the chosen acoustic features to create a thorough representation of dysarthric speech, which should improve the precision of intelligibility evaluation and classification. Deep learning algorithms, such as recurrent neural networks (RNNs), are utilized to capture complex temporal dependencies and patterns within data, while convolutional neural networks (CNNs) are employed for tasks like image classification, including the analysis of mel spectrograms, which play a vital role in the effective classification of dysarthria.

1. PROPOSED SYSTEM DESCRIPTION

This section is devoted to giving a thorough rundown of the methodology used to determine the severity of dysarthria. It discusses the process of choosing and extracting pertinent linguistic and acoustic characteristics, as well as the specifics of the model's development and application for determining the severity of dysarthria.

A. Prepocessing

The following parameters are employed on the basis of their efficacy in facilitating the distinction between voices:

- *MFCCs:* Mel Frequency Cepstral Coefficients measure articulation and vocal tract resonance variations in dysarthria, offering vital information about the intelligibility and clarity of speech.

- *Prosodic Parameters:* Dysarthria frequently causes disruptions in prosodic parameters, such as stress, intonation, and rhythm patterns in speech, which impact phrasing, pitch variability, and speech rate.
- *Formants:* The frequencies at which a vocal tract resonates are known as formants. Vowel perception greatly depends on formants, which can be impacted in dysarthric speech due to changes in vocal tract control. Formant frequency analysis makes it possible to evaluate vowel quality and articulatory accuracy.
- *Voice Quality parameters:* Hoarseness, breathiness, or strained vocal quality are some of the changes that can occur in dysarthric speech (Salhi & Cherif, 2013). These variations in voice quality are measured using parameters like jitter, shimmer, and Harmonics-to-Noise Ratio (HNR), which reveal how severe dysarthria is.
- *Mel Frequency Spectrograms:* When frequency and perceived loudness are not linearly proportional, mel scale corresponds to human auditory perception. Derived from the input speech waveform, the mel-spectrogram offers a frequency representation that is more sensitive to the way that people perceive sound.

B. Classifier Design

1) Convolutional Reccurent Neural Network : Convolutional and recurrent neural networks are combined to create the CRNN (Zuo et al., 2017; Xiao et al., 2016). As seen in the graphical representation of CRNN below, it is made up of convolutional (and pooling) layers followed by a few recurrent layers. The benefits of both recurrent and convolutional networks are combined in CRNN. The input sequence's middle-level, abstract, and locally invariant features can be effectively extracted by the convolutional layers. The pooling layers aid in overfitting control and computation reduction. From the feature sequence produced by the earlier convolutional layers, the recurrent layers extract contextual information. The hyperspectral sequence's dependencies between various bands are captured by contextual information, which makes it more stable and helpful for classification. The architecture used is as follow :

TABLE 1 . CRNN ARCHITECTURE

Layer (type)	Output Shape
conv2d (Conv2D) (Relu)	(None, 256, 256, 32)
max_pooling2d (MaxPooling2D)	(None, 128, 128, 32)
conv2d (Conv2D) (Relu)	(None, 128, 128, 64)
max_pooling2d_4 (MaxPooling2D)	(None, 64, 64, 64)
conv2d (Conv2D) (Relu)	(None, 64, 64, 128)
max_pooling2d (MaxPooling2D)	(None, 32, 32, 128)
flatten (Flatten)	(None, 131072)
dense (Dense)	(None, 128)
dropout (Dropout)	(None, 128)
Reshape (reshape)	(None, 128,1)
Simple_rnn	(None, 64)
dense_2 (Dense) (sigmoid or softmax)	(None, 1)

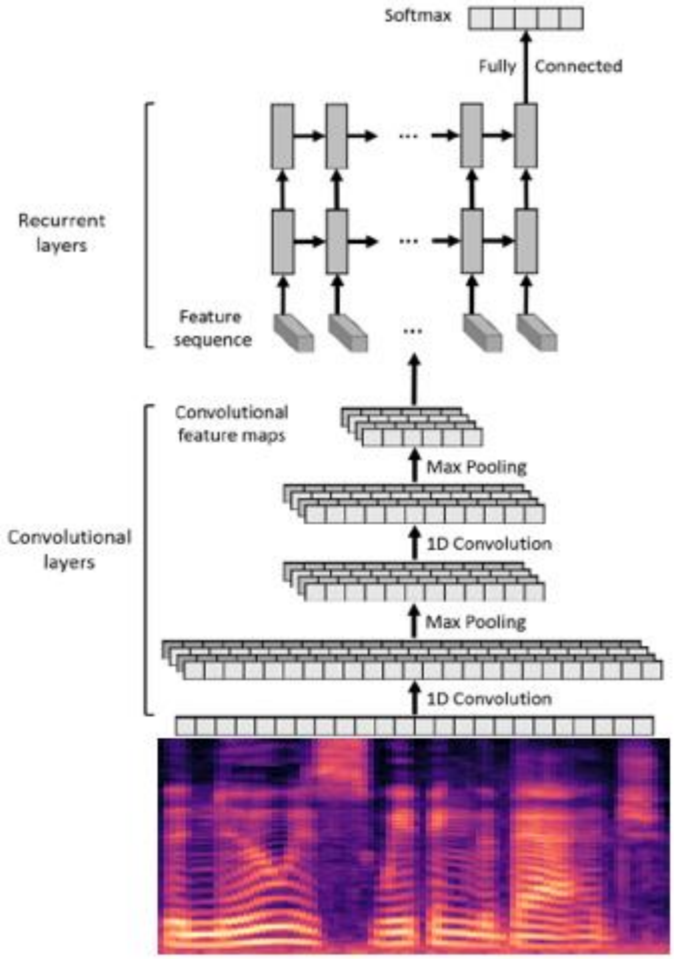

Figure 1. Architecture of the convolutional recurrent neural network (CRNN).

2) Reccurent Neural network classifier : it is used in the study's classification phase. By preserving an internal memory state, RNNs are made to process sequential data. This enables the model to take temporal context into account and identify dependencies within the sequence. Because the "tanh" (hyperbolic tangent) activation function is nonlinear and can handle both positive and negative values, it is utilized in the RNN model's hidden layer, which has 64 neurons. One neuron with a "sigmoid" activation function, chosen for its applicability in binary classification tasks (Kheddar et al., 2019), and a softmax for three-class classification make up the RNN model's output layer. In order to capture and take advantage of temporal dependencies in the data, the RNN model (fig. 2) has a single recurrent hidden layer with 64 neurons that allows connections between neurons in the temporal sequence. With a batch size of 32 and 15 iterations, the model is backpropagated during training in order to minimize the cost function using gradient descent. By efficiently capturing temporal relationships in speech samples, this recurrent structure enables the RNN to increase the classification accuracy of dysarthria severity.

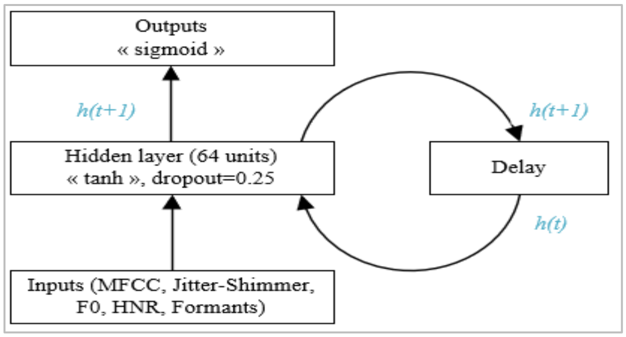

Figure 2. Structure of RNN

1) Convolutional Neural Network classifier: spectrograms were fed into the CNN model. Both pathological and healthy speech samples are represented by these spectrograms. A sigmoid activation function was employed for binary classification, which sought to differentiate pathological samples from normal ones. The softmax activation function was used for the three-class classification (severe, moderate, and mild). The model distinguishes dysarthric speech from normal speech by taking advantage of color changes in the spectrograms that are associated with vocal energy. The table below displays the CNN model's intricate structure :

TABLE 2. CNN ARCHITECTURE

Layer (type)	Activation	Output Shape
(Conv2D)	Relu	(None, 256, 256, 32)
(MaxPooling2D)	/	(None, 128, 128, 32)
(Conv2D)	Relu	(None, 128, 128, 64)
(MaxPooling2D)	/	(None, 64, 64, 64)
(Conv2D)	Relu	(None, 64, 64, 128)
(MaxPooling2D)	/	(None, 32, 32, 128)
Flatten)	/	(None, 131072)
(Dense)	/	(None, 128)
(Dropout)	/	(None, 128)
(Dense)	Sigmoid or softmax	(None, 2 or 3)

2. EXPERIMENTAL EVALUATION

A. Dataset

The Nemours database (Seong et al., 2016) is used in the proposed study for evaluation. 814 recordings of 74 sentences uttered by 11 people with dysarthria—a condition caused by conditions like cerebral palsy and head trauma—make up the dataset. Nonsensical phrases like "The sin is sitting the who," which are part of the Nemours corpus, are used as stimuli to evaluate the intelligibility of dysarthric speech and to examine patterns of production errors. While phoneme-level annotations are available for sentences from 10 out of 11 speakers, word-level annotations are included for the entire dataset. Additionally, information from the Frenchay Dysarthria Assessment version 1 (FDA-1) (Mazari et al., 2023), a standardized tool for assessing dysarthric speech in the English language, is incorporated into the Nemours database.

The identifications and recognition scores from a study (Kadi & Selouani, 2019) with subjects divided into three groups—severe, moderate, and mild—are shown in Table 3. The proposed study can learn a great deal about the traits and intelligibility of dysarthric speech by using this extensive dataset.

TABLE 3. SEVERITY OF DYSARTHRIC SPEAKERS FROM THE NEMOURS DATABASE REFLECTING THE FDA-1 ASSESSMENT TOOL (Mazari et al., 2023).

LEVEL OF SEVERITY	SEVERE			MODERATE			MILD			
PATIENT	SC	BV	BK	RK	RL	JF	LL	BB	MH	FB

B. Experimental Framework

As shown in Figure 3, the experimental framework is separated into two stages: the learning phase and the testing phase. Three supervised machine learning techniques RNN, CNN, and CRNN are suggested in order to create a highly accurate discriminator. The implementation of the RNN model makes use of 37 acoustics parameters. In addition to mastering the nonlinear mapping between inputs and outputs, it also demonstrates exceptional comprehension of the underlying data structure, enabling it to handle speech signal variabilities with ease. A Hanning window size of 2048 samples, a frame spacing of 512 samples, and MFCCs sampled at a frequency of 16 Hz are used to characterize each speech frame for the front-end representation. F0, HNR, Jitter, and Shimmer are among the other features that are extracted. Mel-spectrograms, which are fed into convolutional networks, were created with the Librosa library, sampled at 16 kHz, and then resized to 256x256 pixels. The objective is to create a reliable and accurate system for classifying the severity of dysarthria by utilizing this experimental framework and the strengths of the three models.

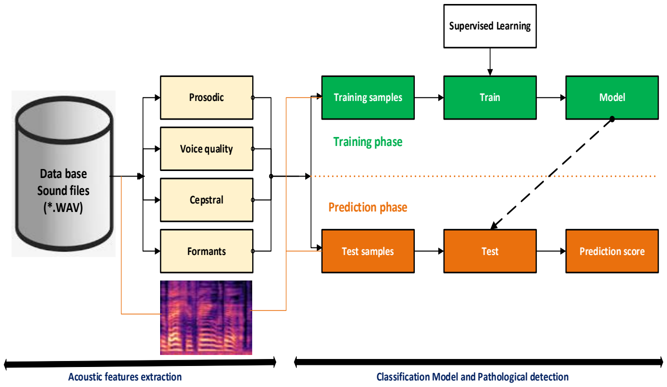

Figure 3. Proposed classification model (Hamza et al., 2023).

C. Results and discussion

Choosing the right amount of training and test data is essential for supervised learning techniques in order to avoid overfitting and underfitting. About 70% of the data in this study is used for training, and the remaining 30% is used for testing when the three methods are used to build classification rules.

The models are put into practice in phases. The results of pathological voice detection, particularly in the context of cross-validation, are the main focus of this evaluation. By splitting the dataset into k-folds, cross-validation is used to perform multiple tests.

The study also looks at using additional voice parameters as feature sets for RNN classifier models, such as jitter, shimmer, F0, and NHR measures. This investigation sheds light on how various feature sets affect the classifiers' performance.

Finally, the three models used to classify the severity of dysarthric voice are assessed. The goal of this in-depth analysis is to determine how well each model classifies the severity levels of dysarthric voices.

1) *Cross validation results* : K-fold cross-validation, which divides the sample set into K subsets, was used to conduct a number of tests. The remaining subsets are used for training, and each subset is used as a validation set. K values of 5, 8, and 10 were chosen for this investigation. Both the RNN and CRNN models were evaluated, with 15 iterations per fold. Table 4 presents the findings.

TABLE 4. DYSARTHRIC DETECTION RATE BY K-FOLD CROSS-VALIDATION

Model	**Classification rate (%)**		
	K=5	**K=8**	**K=10**
RNN	99.51	99.54	99.69
CNN	99.73	**99.80**	99.53
CRNN	98.45	91.82	84.39

The table presents the dysarthric detection rates of RNN, CNN, and CRNN models using k-fold cross-validation with k=5, k=8, and k=10. Both RNN and CNN exhibit excellent and consistent performance, with classification rates above 99%, though CNN slightly outperforms RNN at k=5 and k=8, achieving its highest rate (99.80%) at k=8. However, CRNN demonstrates significantly lower performance, with a sharp decline as k increases, dropping from 98.45% at k=5 to 84.39% at k=10. Overall, CNN emerges as the most reliable model, while CRNN shows limitations in generalization with larger k-values.

2) *Influence of acoustic parameters on the detection of abnormal speech:* The effect of various parameters on the detection and classification system's performance was evaluated through an analysis. In particular, the 10-fold cross-validation classifier model was assessed. Based on the selected acoustic parameters, the results are summarized in Fig. 4.

Interestingly, performance improved across all metrics when all 37 parameters were used. While using the 13 MFCC parameters separately led to poorer performance in comparison to the full parameter set, the comprehensive parameter set improved system accuracy.

The classification score increased when jitter and shimmer were added to MFCC, yielding a classification rate of 99.69% that was comparable to the full parameter set. The classifier operated satisfactorily but was marginally less efficient than the other two configurations when taking into account 24 parameters pertaining to jitter, shimmer, F0, and HNR without MFCCs. This suggests that combining all of the parameters yielded the best accuracy by utilizing the advantages of each one separately.

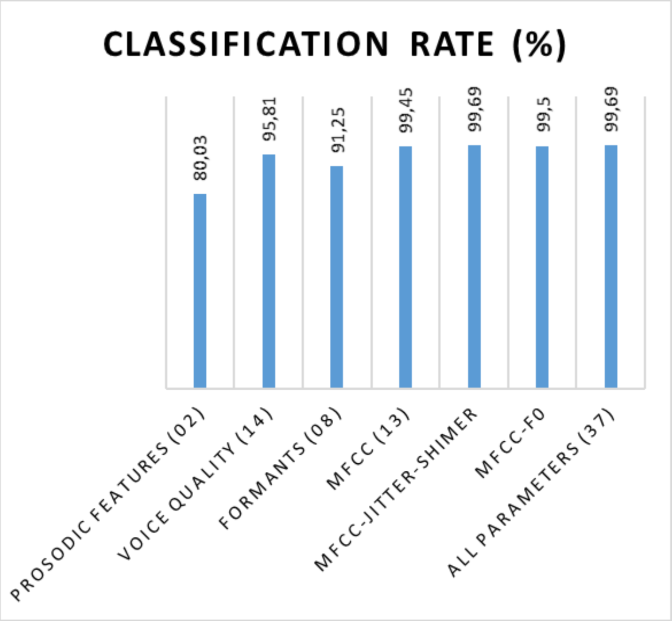

Figure 4. Influence of acoustic parameters on the detection of abnormal speech.

Overall, the findings demonstrate how important it is to employ a variety of acoustic characteristics in order to enhance the functionality of dysarthria classifi

Dysarthria severity classification performance comparison: Data on dysarthric voice is categorized using severity levels, with each level being labeled as "Mild," "Moderate," or "Severe." Thirty percent of the dysarthric database was used for the classification tests, which produced the outcomes detailed on the table 5 :

TABLE 5. DYSARTHRIA SEVERITY-WISE ACCURACY FOR PROPOSED CLASSIFICATION MODELS (100 epochs)

Severity Class	RNN Accuracy (%)	CNN Accuracy (%)	CRNN Accuracy (%)
Severe	0	100	96.97
Moderate	100	97	91.04
Mild	95.50	100	100
overall	65.16	99	96

The table evaluates the severity-wise and overall accuracy of RNN, CNN, and CRNN models for dysarthria classification over 100 epochs. CNN achieves near-perfect accuracy across all severity classes, with 100% for severe, mild, and overall, and 97% for moderate cases, making it the most consistent and reliable model. CRNN also performs well, achieving 96% overall accuracy and perfect results for mild cases (100%), though it slightly underperforms for moderate (91.04%) and severe cases (96.97%). In contrast, RNN struggles with severe dysarthria, achieving 0% accuracy, and its overall performance is the lowest at 65.16%, despite performing well for moderate (100%) and mild (95.50%) cases. These results highlight CNN's superior generalization and robustness, while CRNN shows good potential with some variability, and RNN faces significant challenges, especially with severe cases.

3. CONCLUSION

This study employs supervised learning, leveraging labeled data to train a model capable of learning from examples, to address vocal pathology detection and dysarthria classification. The results highlight the effectiveness of this approach in developing accurate detection and classification systems. The experiments demonstrated that incorporating multiple acoustic parameters significantly improved detection performance, with the RNN achieving an impressive accuracy of over 99%. However, for classifying dysarthria severity, the RNN was surpassed by CRNN and CNN, which achieved accuracy rates of 96% and 99%, respectively. Ultimately, the study successfully developed a reliable model for vocal pathology detection, with future work aiming to enhance pathological voice recognition further.

References

Al-Qatab, B. A., & Mustafa, M. B. (2021). Classification of dysarthric speech according to the severity of impairment: an analysis of acoustic features. *IEEE Access*, *9*, 18183-18194.

Bai, J., Wang, J., & Zhang, X. (2013). A Parameters Optimization Method of v-Support Vector Machine and Its Application in Speech Recognition. *J. Comput.*, *8*(1), 113-120..

Deng, L., & Platt, J. (2014, September). Ensemble deep learning for speech recognition. In Proc. interspeech.

Freed, D. B. (2023). Motor speech disorders: diagnosis and. treatment. plural publishing

Hamza, A., Addou, D., & Kheddar, H. (2023, November). *Machine learning approaches for automated detection and classification of dysarthria severity*. In 2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM) (Vol. 1, pp. 1-6). IEEE.

Hernandez, A., Kim, S., & Chung, M. (2020). Prosody-based measures for automatic severity assessment of dysarthric speech. *Applied Sciences*, *10*(19), 6999.

Joy, N. M., & Umesh, S. (2018). Improving acoustic models in TORGO dysarthric speech database. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(3), 637-645.

Kadi, K. L., & Selouani, S. A. (2019). Distinctive auditory-based cues and rhythm metrics to assess the severity level of dysarthria. *Signal and Acoustic Modeling for Speech and Communication Disorders, edited by Patil and Amy Neustein, De Gruyter*, 205-226.

Kheddar, H., Bouzid, M., & Megías, D. (2019). Pitch and fourier magnitude based steganography for hiding 2.4 kbps melp bitstream. *IET Signal Processing*, *13*(3), 396-407.

Martínez, D., Lleida, E., Green, P., Christensen, H., Ortega, A., & Miguel, A. (2015). Intelligibility assessment and speech recognizer word accuracy rate prediction for dysarthric speakers in a factor analysis subspace. *ACM Transactions on Accessible Computing (TACCESS)*, *6*(3), 1-21.

Mazari, A. C., & Kheddar, H. (2023). Deep learning-based analysis of Algerian dialect dataset targeted hate speech, offensive language and cyberbullying. *International Journal of Computing and Digital Systems*.

Mehrish, A., Majumder, N., Bharadwaj, R., Mihalcea, R., & Poria, S. (2023). A review of deep learning techniques for speech processing. Information Fusion, 99, 101869.

palmer, R., & Enderby, P. (2007). Methods of speech therapy treatment for stable dysarthria: A review. *Advances in Speech Language Pathology*, *9*(2), 140-153

Rudzicz, F. (2010). Articulatory knowledge in the recognition of dysarthric speech. IEEE Transactions on Audio, Speech, and Language Processing, 19(4), 947-960

Salhi, L., & Cherif, A. (2013, April). Selection of pertinent acoustic features for detection of pathological voices. In 2013 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO) (pp. 1-6). IEEE.

Schu, G., Janbakhshi, P., & Kodrasi, I. (2023, June). On using the UA-Speech and TORGO databases to validate automatic dysarthric speech classification approaches. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.

Seong, W. K., Kim, N. K., Ha, H. K., & Kim, H. K. (2016, December). A discriminative training method incorporating pronunciation variations for dysarthric automatic speech recognition. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)* (pp. 1-5). IEEE.

Tu, M., Wisler, A., Berisha, V., & Liss, J. M. (2016). The relationship between perceptual disturbances in dysarthric speech and automatic speech recognition performance. The Journal of the Acoustical Society of America, 140(5), EL416-EL422.

Xiao, Y., & Cho, K. (2016). Efficient character-level document classification by combining convolution and recurrent layers. *arXiv preprint arXiv:1602.00367*.

Zuo, Z., Shuai, B., Wang, G., Liu, X., Wang, X., Wang, B., & Chen, Y. (2015). Convolutional recurrent neural networks: Learning spatial dependencies for image representation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 18-26).