iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.54466

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Children Speech Disorder Detection Using Speech Analytics: A Literature Survey

Poornima A[1], Praveena Vallivel[2]
Student, REVA Academy for Corporate Excellence (RACE), REVA University, Bengaluru, Karnataka, India

Abstract: ***Children speech impairments can have a big impact on how well a kid communicates and develops as a person. Early diagnosis of speech impairments is essential for prompt intervention and efficient therapy. Speech analytics approaches have become promising tools in recent years for the identification and evaluation of various speech problems in children. This literature review seeks to present an overview of the state-of-the-art in speech analytics-based child speech disorder detection. It explores the approaches and technology used in speech analytics for disorder detection and covers a wide spectrum of speech disorders, such as articulation abnormalities, phonological disorders, and language disorders. The survey emphasises the potential impact of speech analytics in enhancing the diagnosis while highlighting the obstacles, trends, and future directions in this quickly expanding sector.***

Keywords: Speech disorder detection, MFCC, HMM, GMM, CNN, DNN, LPC, Feature Extraction, Machine Learning, open Smile, Deep Learning

I. INTRODUCTION

Children's social interactions, academic progress, and overall well-being are all influenced by their ability to communicate successfully. Language acquisition proceeds in a predictable order, beginning with simple sounds and syllables and progressing to the formation of words, sentences, and, finally, complex language structures. It is critical to monitor this growth to detect potential speech abnormalities, developmental delays, or language impairments at an early stage. Early intervention can improve children's speech development greatly, allowing them to attain their full potential.

In the past, speech-language pathologists, linguists, or educators used manual evaluations to a large extent to determine how children's speech was organised. For these evaluations, audio recordings were analysed, speech samples were transcribed, and the utterances were manually compared to predetermined developmental stages. However, this manual method requires a lot of work, takes a long time, and is prone to biases and mistakes made by humans. Furthermore, the lack of experienced assessors may cause assessments to be delayed and proper intervention to be postponed for longer periods of time.

Fortunately, recent developments in artificial intelligence, machine learning, and speech processing have completely changed the field of speech evaluation. Large-scale speech data analysis is now possible using automated systems, which can also spot patterns and spot variations from expected developmental trajectories. These systems can accurately and impartially evaluate young children's speech order in real time by utilising the power of AI algorithms, giving insightful data on how their linguistic development is progressing.

II. SPEECH FEATURE EXTRACTION ALGORITHMS:

Feature extraction is an essential part of many audios analysis tasks, e.g., Automatic Speech Recognition (ASR), analysis of paralinguistics in speech, and Music Information Retrieval (MIR). Feature extraction methods typically provide a multidimensional feature vector for each spoken input. To parametrically characterise the speech signal for the recognition process, a variety of techniques are available, including perceptual linear prediction (PLP), linear prediction coding (LPC), and mel-frequency cepstrum coefficients (MFCC). The most well-known and well-liked is MFCC. Feature extraction is the most important aspect of speaker recognition. Speech characteristics play an important role in distinguishing a speaker from others. Feature extraction minimises the magnitude of the voice signal while preserving the signal's power. Feature extraction is achieved by converting the speech waveform to a parametric representation at a lower data rate for later processing and analysis [1]. This is commonly referred to as front end signal processing. It converts the processed speech signal into a short yet logical representation that is more discriminative and reliable than the original signal. Because the front end is the first element in the sequence, the quality of the subsequent features (pattern matching and speaker modelling) is heavily influenced by it [2].

©IJRASET: All Rights are Reserved

MFCC is a feature that is frequently used in automatic speech recognition and speaker identification because it is appropriate for understanding people and the frequency at which they speak [3]. Mel frequency analysis and cepstral analysis are the two main steps in the MFCC feature extraction process. A group of crucial coefficients known as MFCC are used to determine Mel cepstral. A set of cepstrums can be created so that are sufficient to represent the audio signal from the segments in the audio signal. Since the frequency band on the Mel cepstrum is evenly distributed on the Mel scale, it will be more linear than what we typically observe. This contrasts with the general cepstrum. The cepstral representation approach is more analogous to the nonlinear auditory system in humans. The MFCC features are anchored in the recognised discrepancy of the human ear's crucial bandwidths, and frequency filters separated linearly at low frequencies and logarithmically at high frequencies have been utilised to keep the phonetically vital qualities of the speech signal. Speech transmissions frequently include tones of various frequencies. The subjective pitch of each tone is calculated using the Mel scale, and each tone has an actual frequency, f (Hz). Below 1000 Hz, the mel-frequency scale has linear frequency spacing, while above 1000 Hz, it has logarithmic frequency spacing. The reference point is the pitch of a tone at 1 kHz and 40 dB above the perceptual audible threshold, which is specified as 1000 mels. MFCC is based on signal disintegration with the help of a filter bank. The MFCC gives a discrete cosine transform (DCT) of a real logarithm of the short-term energy displayed on the Mel frequency scale [4].

Linear prediction coefficients (LPC) mimic the human vocal tract and provide robust speech features [5]. It assesses the speech signal by approximating the formants, removing their effects from the speech signal, and estimating the concentration and frequency of the residue left behind. Each sample of the signal is stated as a direct incorporation of previous samples in the result. The formants are defined by the coefficients of the difference equation, so LPC must approximate these coefficients. LPC is an effective speech analysis method that has gained popularity as a formant estimation method [6].

Linear prediction cepstral coefficients (LPCC) are cepstral coefficients generated from the spectral envelope determined by LPC [7]. The coefficients of the Fourier transform visualisation of the logarithmic magnitude spectrum of LPC are denoted as LPCC. Because of its ability to properly symbolise speech waveforms and properties with a minimum number of parameters, cepstral analysis is widely used in speech processing [8].

Line Spectrum Pairs (LSPs) are an alternative LP spectral representation of speech frames that have been found to be perceptually meaningful in coding systems. LSPs can be quantized using perceptual criteria and have good interpolation properties. Line spectral frequencies (LSF) are the frequencies of individual lines in a Line Spectral Pair (LSP). LSF describes the two resonance scenarios that occur in the human vocal tract's interconnected tube model. The model considers the nasal cavity and mouth shape, which provides the foundation for the linear prediction illustration's essential physiological value. The two resonance conditions specify whether the vocal tract is entirely open or totally closed at the glottis [9]. The two scenarios generate two groups of resonant frequencies, with the number of resonances in each group determined by the number of linked tubes. The resonances of each circumstance are the odd and even line spectra, which are interlaced into a uniquely rising collection of resonances.

The Wavelet Transform (WT) theory is based on signal analysis at different scales in the time and frequency domains [10]. A wavelet is a waveform with an effective duration of zero and an average value of zero. WT is a signal processing approach that can be used to efficiently represent real-life non-stationary data. It can extract information from transient signals in both the time and frequency domains at the same time. The continuous wavelet transform (CWT) is a technique for dividing a continuous-time function into wavelets. However, there is information duplication and massive computational effort is necessary to calculate all possible scales and translations of CWT, limiting its applicability. The discrete wavelet transform (DWT) is a WT variant that adds flexibility to the decomposition process. It was introduced as a very flexible and efficient approach for signal sub-band breakup. The DWT parameters give information on various frequency scales. This improves the quality of the speech information acquired in the relevant frequency band. The DWT's ability to split the variance of the input items on a scale-by-scale basis is an added benefit. This segmentation results in the scale-dependent wavelet variance, which is analogous to the more familiar frequency-dependent Fourier power spectrum in many aspects [11].

To extract important information from speech, the perceptual linear prediction (PLP) technique combines the critical bands, intensity-to-loudness compression, and equal loudness pre-emphasis. By removing the speaker-dependent elements, it was originally designed for use in voice recognition tasks. Its origins are in the nonlinear bark scale. PLP provides a representation that resembles the MFCC by conforming to a smoothed short-term spectrum that has been equalised and compressed to simulate human hearing. The PLP technique replicates several key aspects of hearing, and an autoregressive all-pole model is used to simulate the resulting auditory spectrum of speech [12]. PLP produces orthogonal results that are comparable to cepstral analysis but with minimal resolution at high frequencies, indicating a technique based on auditory filter banks. Perceptual linear prediction refers to the method's use of linear predictions for spectral smoothing. PLP combines both linear prediction analysis and spectral analysis [13].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

openSMILE is an open-source software for speech and music signal classification and automatic feature extraction from audio signals. "Speech & Music Interpretation by Large-space Extraction" is what "SMILE" stands for. The software is primarily used in the field of automatic emotion recognition and is popular among researchers studying affective computing. openSMILE is widely applied in automatic emotion recognition for affective computing. openSMILE is used for academic research as well as for commercial applications to automatically analyze speech and music signals in real-time. In contrast to automatic speech recognition which extracts the spoken content out of a speech signal, openSMILE is capable of recognizing the characteristics of a given speech or music segment [14].

III. SOME OF THE COMMONLY USED MACHINE LEARNING MODELS FOR SPEECH DISORDER DETECTION.

The Gaussian Mixture Model (GMM) is a statistical model used for speech disorder detection. It captures the statistical distribution of speech features and classifies them into different categories. Overall, the GMM provides a flexible and probabilistic framework for modeling and analyzing data with a mixture of Gaussian distributions, making it useful for various applications, including speech disorder detection.

Hidden Markov Models (HMM) have been traditionally used for speech disorder detection, particularly in cases where temporal modeling is crucial. They model speech sequences probabilistically, allowing for the detection of abnormalities or deviations from normal speech patterns.

Convolutional Neural Networks (CNN) are widely used for speech disorder detection tasks. They can analyze spectrogram or other acoustic representations of speech signals to extract relevant features and classify them into different categories.

Recurrent Neural Networks (RNN) particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) variants, are commonly employed for speech disorder detection. They can capture temporal dependencies in speech data, making them effective for sequence-based classification tasks.

Support Vector Machines (SVM) are widely used in speech disorder detection, especially for classification tasks. They can separate different classes of speech disorders by finding an optimal hyperplane in the feature space.

Naive Bayes (NB) is a classification method that is often used for detecting speech disorders. Given the class designation, it presupposes that characteristics are independent. A fresh speech sample is assigned to the class with the highest probability after the algorithm collects pertinent characteristics from the speech signal and assesses class probabilities. Naive Bayes is a straightforward and effective method that works well in high-dimensional feature spaces, although it may have trouble with intricate feature interactions. While it offers a simple answer, more sophisticated methods, such as deep learning, are being investigated for enhanced performance and accuracy in speech problem diagnosis.

XGBoost is a gradient boosting framework generally used for tabular data, however it may be modified for speech disease diagnosis by transforming speech characteristics to a tabular format. The effectiveness of employing XGBoost for speech disorder identification is dependent on parameters such as feature quality, dataset size, and hyperparameter tweaking. Collaboration with domain experts is critical for reaching peak performance.

IV. DATASET

The SLI dataset is designed for studying Specific Language Impairment, a developmental language disorder. It includes language tasks, assessments, and standardized tools to evaluate various linguistic domains affected by SLI. The dataset covers a specific age range, often includes a control group for comparison, and may have longitudinal data. Ethical guidelines are followed to protect participant privacy. The SLI dataset enables researchers and clinicians to investigate language profiles, identify linguistic deficits, study developmental trajectories, and develop targeted interventions for individuals with SLI.

ASR datasets are collections of audio recordings and transcriptions used for training and evaluating speech recognition models. They encompass diverse speakers, languages, and speech characteristics. The datasets vary in size, often include annotations and metadata, and serve as benchmarks for evaluating ASR techniques. Open availability promotes collaboration and advancements in the field of ASR.

The Saarbruecken Voice Database (SVD) is a large-scale database created by Saarland University, Germany, for speech and voice analysis research. It contains diverse recordings of individuals with speech disorders and normal speech. The database offers a broad representation of different speech disorders, multimodal data including audio and video recordings, comprehensive annotations, and metadata. It serves various applications, such as speech disorder diagnosis, therapy evaluation, and the development of speech analysis algorithms. The SVD is a valuable resource for researchers and professionals in speech pathology and related fields, enabling the study of speech production, disorders, and the development of assessment tools and therapies.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The ABIDE1 dataset is a publicly available collection of resting-state functional magnetic resonance imaging (fMRI) data from individuals with autism spectrum disorder (ASD) and typically developing controls. It includes diverse participants spanning different age groups and utilizes various MRI scanners, ensuring a broad representation of the population. Alongside the imaging data, the dataset provides extensive clinical and phenotypic information. The ABIDE1 dataset undergoes rigorous quality control and promotes open access and collaboration among researchers. It has been utilized in various research studies to investigate brain connectivity patterns, develop classification algorithms, and explore biomarkers for ASD diagnosis and prognosis. Overall, the ABIDE1 dataset plays a vital role in advancing our understanding of the neural underpinnings of ASD and supports the development of new approaches for early detection and intervention.

V. EVALUATION METRICS

1) *Word Error Rate (WER):* It measures the accuracy of automatic speech recognition (ASR) systems by comparing the transcribed output with the ground truth. It calculates the percentage of incorrectly recognized words, including substitutions, insertions, and deletions. Lower WER indicates better performance.
2) *Phoneme Error Rate (PER):* Like WER, PER evaluates the accuracy of phoneme-level recognition in ASR systems. It measures the percentage of incorrectly recognized phonemes compared to the ground truth. Lower PER values indicate better performance.
3) *Accuracy:* Accuracy measures the overall correctness of the classification system by calculating the ratio of correctly predicted instances (e.g., correctly identified speech disorders) to the total number of instances. It provides a general assessment of system performance but may not be suitable for imbalanced datasets.
4) *Precision:* It focuses on the true positive rate, i.e., the ratio of correctly identified instances (e.g., correctly identified speech disorders) to the total number of instances predicted as positive. It indicates the system's ability to correctly identify positive cases.
5) *Recall (Sensitivity):* Recall, also known as sensitivity or true positive rate, calculates the ratio of correctly identified instances (e.g., correctly identified speech disorders) to the total number of actual positive instances. It represents the system's ability to correctly detect positive cases.
6) *Specificity:* It measures the true negative rate, i.e., the ratio of correctly identified negative instances (e.g., correctly identified normal speech) to the total number of actual negative instances. It indicates the system's ability to correctly identify negative cases.
7) *F1 Score:* It is the harmonic mean of precision and recall. It provides a balance between the two metrics and is useful when there is an imbalance between positive and negative instances.

These evaluation metrics can help assess the effectiveness and accuracy of child speech disorder detection systems using speech analytics.

VI. LITERATURE SURVEY

In this section, we present an overview of several influential papers that have made significant contributions to the field of children speech disorder detection using acoustic features. The summarized information from these papers is presented in the Table 1.

Paper Title	Dataset	Models	Accuracy	Key Findings
An Automated Assessment Tool for Child Speech Disorders	CU-Child127 dataset	Deep Neural Network (DNN). Multi-task learning (MTL) trained acoustic model was implemented by the Kaldi speech recognition toolkit.	89%	An automated evaluation tool that mimics the clinical evaluation of child SSD is presented in this paper. It works with a mobile application, and DNN-based ASR technology supports automatic assessment. The use of domain knowledge in the creation of automatic assessment is described in this work and the effectiveness of the demonstration system was evaluated in a real-world setting [15].

Automatic Screening to Detect 'At Risk' Child Speech Samples using a Clinical Group Verification framework	Voice recordings of 164 children via the iOS application.	Gaussian Mixture Model (GMM) using MFCC	79.88%	This paper uses Gaussian Mixture Models to propose a unique paradigm for clinical-group verification. On a dataset with short-duration utterances, the clinical-group verification architecture and unique scoring algorithms were able to produce encouraging subject-level classification results. Future work would be to examine the feature space that produced such precise discrimination both before and after using any subject-level scoring strategies. [16].
Prediction of Specific Language Impairment in Children Using Speech Linear Predictive Coding Coefficients.	SLI database (LANNA).	Naive-Bayes (NB) and Support Vector Machine (SVM)	NB with 5 folds-97.9%	This research highlights the potential of linear predictive coding (LPC) coefficients as a useful tool for diagnosing SLI in children. This study demonstrates that a model trained on LPC features can effectively differentiate between children with SLI and typically developing children, providing a step toward early identification and intervention for language impairments in children [17].
Detection of Specific Language Impairment in Children Using Glottal Source Features		Support Vector Machine (SVM) and feed-forward neural network (FFNN), are trained separately for the MFCC, openSMILE and glottal features	98.82%	A novel technique for SLI detection from speech signals is presented in this paper distinguishes between healthy child speakers and kids with SLI using the glottal and acoustic features extracted from speech. For each speech utterance, three sets of acoustic features (one MFCC feature set, two openSMILE-based acoustic feature sets, and one set of glottal features) are extracted. According to experimental findings, the classification accuracies produced by glottal parameters were good and somewhat inferior to those produced by acoustic features [18].
Automated detection of childhood speech disorders	Saarbruecken voice database (SVD) including their sources, sizes, and character-	Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)	93.05%	This paper provides a comprehensive analysis of the reviewed literature and present the main findings related to automated detection of childhood speech disorders. This research work discusses the effectiveness of different approaches, the strengths, and limitations of the existing methods, and potentially identify areas for further research [19].

	istics, as well as how they were used in the studies related to automated detection of childhood speech disorders			
Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network	ABIDE dataset containing 505 ASD patients and 530 typical controls.	Support Vector Machine (SVM), K-nearest neighbors (KNN) and Random Forest (RF) classifiers and Convolutional Neural Networks (CNN)	85%	The proposed model in this paper can detect ASD correctly with an accuracy of 70.22% using the ABIDE-I dataset and the CC400 functional parcellation atlas of the brain. Also, the CNN model developed used fewer parameters than the state-of-art techniques and is hence computationally less intensive. CNN with state of art gave high accuracy [20].
Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis	Schizophrenia dataset	Multilevel Bayesian Modelling	79%	Utilizing Bayesian meta-analysis, the researchers analyzed the collected data to determine the overall effects and associations of voice patterns in schizophrenia. This study findings contribute to a better understanding of voice patterns in schizophrenia, potentially offering insights into the underlying mechanisms of the disorder and providing a foundation for further research in this area [21].
An Analytical Study of Speech Pathology Detection Based on MFCC and Deep Neural Networks	Saarbruecken Voice Database (SVD) to detect abnormal voices	Deep Neural Network (DNN) used for model training. Acoustic Features used are chroma, mel spectrogram and mel frequency cepstral coefficient (MFCC).	96.77%	This research work presents a customized deep neural network (DNN) algorithm for classifying pathology voices from healthy based on samples from the publicly available Saarbruecken Voice Database (SVD). The major drawback is that they fail in generalization to a real-world scenario involving variable voice patterns. Future work will be to incorporate more accurate data from other publicly available datasets and update the model to learn meaningful features to produce more accurate results [22].

A Deep Learning Based Evaluation of Articulation Disorder and Learning Assistive System for Autistic Children.	ASR dataset	Deep Neural Network (DNN) with auto-encoder for the evaluation of articulation disorder	95%	MFCC feature extraction has acquired the highest accuracy rate of 85%, whereas ZCPA has acquired 38% and LPC has achieved 82%. An ASR experiment in Malay language with MFCC features was conducted and that gave an average accuracy of 95% using Multilayer Perceptron [23].
Voice disorder classification using speech enhancement and deep learning models	Saarbruecken voice database (SVD)	CNN-LSTM, Gaussian Mixture Model (GMM) and SVM, XGBoost	CNN-85.2%, SVM-69.9%	This experiment results highlights the effectiveness of signal enhancement techniques and the selection of appropriate classification algorithms for improving the accuracy of automatic voice disorder classification. The combination of various input features and the use of advanced machine learning algorithms contribute to enhanced classification performance in the detection and assessment of voice pathologies [24].

Table 1 : Paper Summary

VII. CONCLUSION

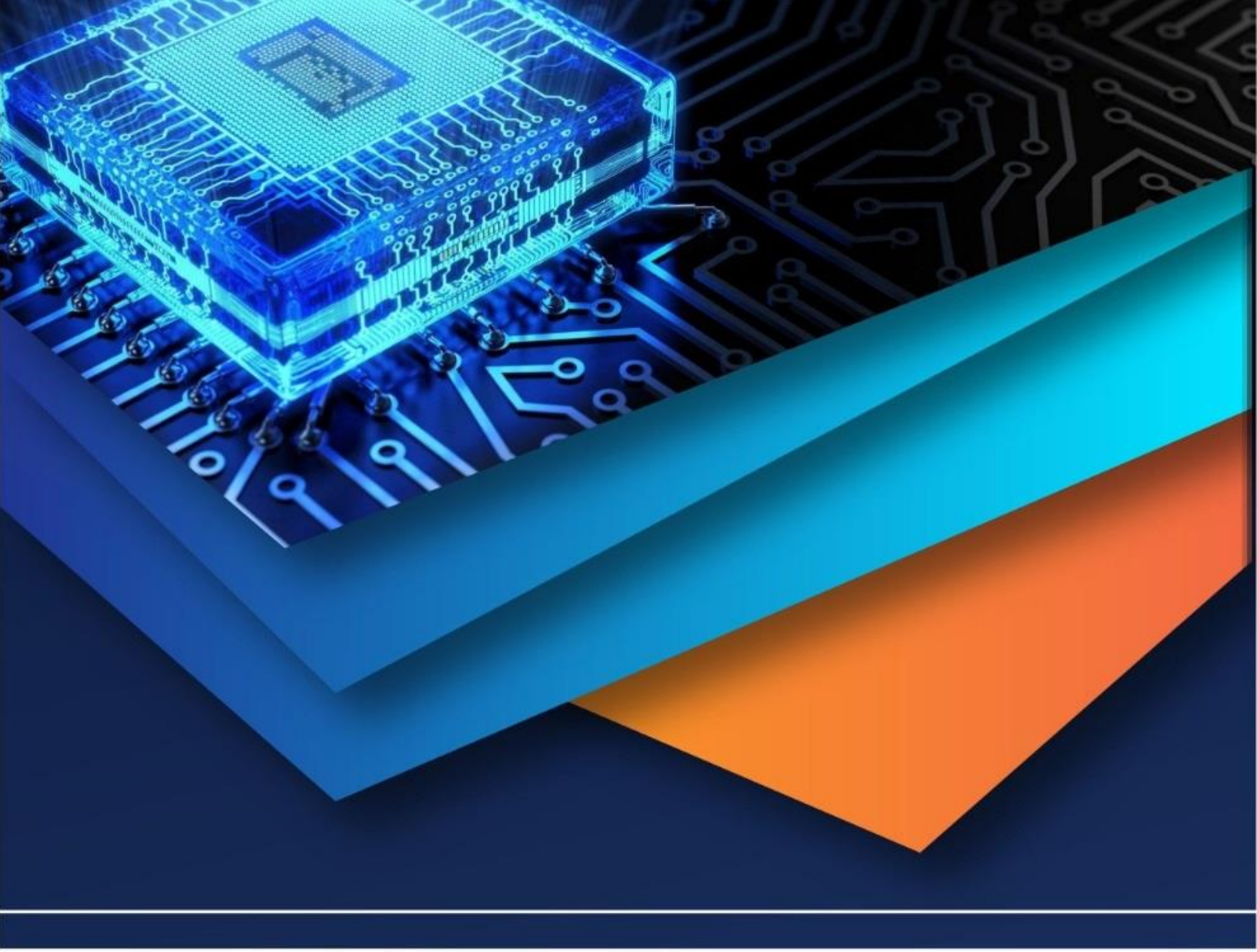
This survey paper provides a comprehensive overview of the current state-of-the-art techniques for child speech disorder detection. It emphasizes the need for further research and collaboration among experts in the fields of speech pathology, machine learning, and signal processing to develop more effective and accessible detection systems that can make a positive impact on the lives of children with speech disorders. The survey revealed that different types of speech disorders require different detection methodologies. For articulation disorders, techniques such as acoustic analysis, phonetic transcription, and articulatory modeling have shown promising results. On the other hand, language disorders necessitate the application of natural language processing and language modeling techniques. This paper identified several important factors to consider when designing and implementing speech disorder detection systems. These include the availability and size of annotated datasets, feature selection, algorithm robustness, and system usability.

REFERENCES

[1] S. Narang and M. Gupta, “Speech feature extraction techniques: A review;,” International Journal of Computer Science and Mobile Computing., pp. 4(3):107-114, 2015.
[2] S. Shah, u. A. A and S. Shaukat, “Neural network solution for secure interactive voice response,” World Applied Sciences Journal., pp. 6(9):1264-1269, 2009.
[3] C. Sandipan, R. Anindya and S. Goutam, “Fusion of a complementary feature set with MFCC for improved closed set text-independent speaker identification.,” IEEE International Conference on Industrial Technology, . ICIT, pp. pp. 387-390, 2006.
[4] K. Ravikumar, B. Reddy, R. Rajagopal and H. Nagaraj, “Automatic detection of syllable repetition in read speech for objective assessment of stuttered Disfluencies.,” Proceedings of World Academy Science, Engineering and Technology., 2008.
[5] K. Al-Sarayreh, R. Al-Qutaish and B. Al-Kasasbeh, “Using the sound recognition techniques to reduce the electricity consumption in highways.,” Journal of American Science., pp. 5(2):1-12, 2009.
[6] S. Agrawal, A. Shruti and C. Krishna, “Prosodic feature based text dependent speaker recognition using machine learning algorithms.,” International Journal of Engineering Science and Technology., pp. 2(10):5150-5157, 2010.
[7] K. Ravikumar, R. Rajagopal and H. Nagaraj, “An approach for objective assessment of stuttered speech using MFCC features.,” ICGST International Journal on Digital Signal Processing, DSP., pp. 9(1):19-24, 2009.
[8] Q.-Z. Wu, I.-C. Jou and S.-Y. Lee, “On-line signature verification using LPC cepstrum and neural networks,” IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics., pp. 27(1):148-153, 1997;.
[9] I. McLoughlin, “Line spectral pairs;,” Signal Processing, pp. 88(3):448-467, 2008.
[10] M. Oliveira and A. Bretas, “Application of discrete wavelet transform for differential protection of power transformers.,” IEEE PowerTech. Bucharest: IEEE, pp. pp.1-8, 2009.
[11] D. Gupta and S. Choubey, “Discrete wavelet transform for image processing;,” International Journal of Emerging Technology and Advanced Engineering., pp. 4(3):598-602, 2015.
[12] H. Hermansky, “Perceptual linear predictive (PLP) analysis of speech.,” The Journal of the Acoustical Society of America;, pp. 87(4):1738-1752, 1990.

[13] K. Ravikumar, R. Rajagopal and H. Nagaraj, "An approach for objective assessment of stuttered speech using MFCC features;," ICGST International Journal on Digital Signal Processing, DSP., pp. 9(1):19-24, 2009.
[14] S. Abhishek, C. Suraj and S. Aditya, "Audio Feature Extraction Tools.," International Research Journal of Modernization in Engineering Technology and Science, vol. 03, no. 04, 2021.
[15] N. S. Ioi, T. Dehua, W. Jiarui, J. Yi, N. W. Yee and L. Tan, "An Automated Assessment Tool for Child Speech Disorders," 11th International Symposium on Chinese Spoken Language Processing, pp. pp. 493-494, 2018.
[16] P. V. Kothalkar, R. J, D. C, M. J and H. L. H. T. F. Campbell, "Automatic Screening to Detect 'At Risk' Child Speech Samples using a Clinical Group Verification framework," 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. pp. 4909-4913, 2018.
[17] Y. Sharma and K. S. B., "Prediction of Specific Language Impairment in Children Using Speech Linear Predictive Coding Coefficients," First International Conference on Power, Control and Computing Technologies, pp. pp. 305-310, 2020.
[18] M. K. Reddy, P. Alku and K. S. Rao, "Detection of Specific Language Impairment in Children Using Glottal Source Features," IEEE Access, vol. 8, pp. pp. 15273-15279, 2020.
[19] S. Mostafa, Z. Usman and A. Beena, "The Automatic Detection of Speech Disorders in Children: Challenges, Opportunities, and Preliminary Results," IEEE Journal of Selected Topics in Signal Processing, vol. 14, pp. pp. 400-412, 2020.
[20] S. Zeinab, A. M. Sadegh, S. Soorena, Zomorodi, A. Mariam, A. Moloud, Khosrowabadi, U. Rajendra, S. Reza and Vahid, "Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network.," Frontiers in Neuroscience., vol. 13, p. 1325, 2020.
[21] Alberto Parola , Arndis Simonsen and Vibeke Bliksted, "Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis.," Schizophrenia research., vol. 216, 2019.
[22] Z. Mohammed, B. Reshma, A. Y. Ajmi, G. Yanhui, T.-T. Kiet and E. M. Mamun, "An Analytical Study of Speech Pathology Detection Based on MFCC and Deep Neural Networks.," Computational and Mathematical Methods in Medicine, vol. 2022, p. 15, 2022.
[23] L. Pillai and E. Sherly, "A Deep Learning Based Evaluation of Articulation Disorder and Learning Assistive System for Autistic Children," International Journal on Natural Language Computing, vol. 6, 2017.
[24] C. Mounira, S. S. Ahmed, B. Malika and Y. M. Sidi, "Voice disorder classification using speech enhancement and deep learning models," Biocybernetics and Biomedical Engineering, vol. 42, no. 2, pp. 463-480, 2022.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)