Multimedia Tools and Applications (2022) 81:24145–24166
https://doi.org/10.1007/s11042-022-12817-z

Intelligent stuttering speech recognition: A succinct review

Nilanjan Banerjee [1] · Samarjeet Borah [2] · Nilambar Sethi [1]

Received: 11 February 2021 / Revised: 21 February 2022 / Accepted: 9 March 2022 /
Published online: 19 March 2022

Abstract
Stuttering speech recognition is a well-studied concept in speech signal processing. Classification of speech disorder is the main focus of this study. Classification of stuttered speech is becoming more important with the enhancement of machine learning and deep learning. In this study, some of the recent and most influencing stuttering speech recognition methods are reviewed with a discussion on different categories of stuttering. The stuttering speech recognition process is divided mainly into four segments-input speech pre-emphasis, segmentation, feature extraction, and stutter classification. All these segments are briefly elaborated and related researches are discussed. It is observed that different traditional machine learning and deep learning classification approaches are employed to recognize stuttered speech in last few decades. A comprehensive analysis is presented on different feature extraction and classification method with their efficiency.

Keywords Stuttering · Speech recognition · Feature extraction · Machine learning · Deep learning · Classification

1 Introduction

Human speech is employed for communication to precise their feelings, ideas, and thoughts. A sort of speech problem where the flow of speech is interrupted is understood as stuttering or

✉ Samarjeet Borah
samarjeet.b@smit.smu.edu.in

Nilanjan Banerjee
nilanjan.banerjee@giet.edu

Nilambar Sethi
nilambar@giet.edu

[1] Department of Computer Science and Engineering, GIET University, Odisha, India

[2] Department of Computer Applications, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim, India

generally heard as stammering. It is a speech disorder where the sufferers want to say but have difficulty saying it. Stutterers feel same of difficulty while communicating with other people, which often affect a person's quality of life and interpersonal relationships. It creates negative vibes influencing job performance and opportunities. A huge number of people i.e., more than 70 million people worldwide are affected by this problem. This number is about 1% of the total population [41]. It is observed whenever they communicate, receiver person feels irritated by hearing the prolonged words and most of the time don't understand. E. Charles Healey, in his article, sought a discussion of children reaction to stuttering, impacts of stuttering with listener recall and comprehension of story information listeners', interferes stuttering on listeners' reactions and listeners' reaction on strategies and therapy programs on stuttering [21]. An enormous source of evidence-based information about the cited things has been provided in the extant literature. Stuttering, aging processes and several neurological diseases in relation to speech can be identified by muscular stiffness and analyzing the latency times in verbal reactions, their coordination and their patterns of the muscles (respiratory, glottal, oromandibular) involved in speaking [50]. Being an interdisciplinary field of research among different domains like speech pathology, psychology, speech physiology, acoustics and signal analysis, the field of stuttering speech recognition is one area of interest for the researches over previous few decades. Traditionally, the assessment of stuttering is done by manually counting and classifying the occurrence of disturbances in stuttering speech. Time of disfluency in total speech is also considered as a measurement to assess stuttered speech. But this type of manually stuttering assessment varies depending on different speech language pathologist (SLP). So, it is time consuming and liable to error. Therefore, an Automatic Speech Recognition system (ASR) system for stuttered speech is used to automate the dysfluency count and type of dysfluency classification for assessment of stuttered speech [41]. Such approach can support Speech Language Pathology (SLP). Therefore, in the past two decades, the main focus in this field of researches is on developing objective methods using DSP and AI concepts to assist the SLP during stuttering assessment. Classification of speech disorder is the main focus of this study. By extracting features from the speech and using different classifier we can easily classify non-stuttered and different types of stuttered speech. Therefore, artificial intelligent has become a significant role to classify form of dysfluencies in automatic stuttering recognition system which can support Speech Language Pathology (SLP) [7]. It is observed that different traditional machine learning approaches are employed to recognize stuttered speech in last few decades. Mostly, appearance of three major classifier HMMs, SVM and ANNs was notable in recognition of stuttering speech. Whereas LPC, LPCC, PLP and MFCC feature extraction methods were employed in the previous studies. Now, Deep learning algorithms have become very popular over traditional machine learning algorithms for stuttering speech recognition.

1.1 Characteristic of stuttering

Stuttering is a speech disorder. Usually stuttering has been detected from the age of 18 months to 24 months. It's mainly the problem of fluency and delivery of speeches in case of stuttering varies considerably across different speaking situation like diplomatic, official or presentation mode of conversation and home atmosphere conversation. There are several reasons of stuttering. Some of those along with types are discussed in this section.

Concentrating on the previous studies and researches some of the following causes of stuttering are being explained. Based on the literature, some of the common causes of

stuttering are Genetic, Physiological, Congenital, Auditory and Environmental [31]. The brief discussion on the same is presented as follows:

- Genetics – On the basis of recent international researches there are few certain genes identified for the stuttering and the genetic family linkage plays a role.
- Physiological – From the field of brain imaging study a little bit of dis-functioning of brain during speaking is the reason for lack of speech production and unable to keep the fluency of saying words or sentences weakness of human neuro system may act in the process.
- Congenital – Congenital factors like physical trauma at the time of birth, cerebral palsy, retardation may cause the stuttering. The conditions are found in case of sibling and sudden growth in linguistic ability.
- Auditory – Deafness and hard of hearing have an impact on stuttering. Slow response to audio increases the stuttering habit.
- Environmental – An uncomfortable and stressful situation is a significant reason for development of stuttering behaviors.

1.2 Types of stuttering

Stuttering is also interwoven with the language, phonetics, social, emotional, cognitive and physiological domains, among others. Stuttering can be classified based on different ways. The extant literature has provided subtypes of stuttering based on the following: Subtypes classification based on Etiology, subtypes depending on stuttering phenomena, subtypes on the basis of biological characteristics of the stuttered person, subtypes related to concomitant disorders, subtypes corresponding to developmental course and subtypes based on statistically generated models [74]. The way of stuttering varies from person to person. It is observed that the foremost common types of stuttering are:

- Interjection (irrelevant and insignificant extra sounds or words like – uh, um, well): like "He um must stop talking now."
- Revisions (the Changes in phrase of a sentence or going back to the beginning of a sentence for correcting initial phrase.): like "She—her mother...",," I had – I lost my keys."
- Incomplete phrases: "What are the …. let start writing"
- Repetition (Phrase–repetitions "I was-I was walking past the garden.", "I tell—I tell—I tell you.", Word-repetitions: "Go-go-go away.",," Here is my puppy-puppy-puppy.", Part-word repetitions: "I w-w-w-want a drink.", un-un-under, o-o-open)
- Prolonged sounds: "Weather is so breeeeeeeezy.", "mmmmmmmmmom"
- Broken words (silent pause within a word): "He was eat[pause]ing over there"

With the advancement of artificial intelligent, recognition of stuttering speech has become more convenient. This paper reviews present progression associated with analysis of stuttered speech recognition on different database with different features extraction techniques. Stuttering recognition system is mainly divided into two sections i.e., feature extraction and classification. This paper is organized as follows – different feature extraction methods employed in stutter recognition process and logic on pre-emphasis & segmentation processes are discussed in section 2; Section 3 highlights some of the pioneering works using different classification

methods to recognize stutter speech. Analysis and Discussion is presented in section 4 considering various parameters. Finally, the paper is concluded with future work in section 5.

2 Methods of feature extraction

Feature extraction can be called as the fundamental process in speech recognition. The speech features can be categorized as time domain (temporal)and frequency (spectral) based features. The improvement of accuracy in recognition always depends on the proper selection of the features. The best known spectral based feature extraction methods employed in stuttering speech recognition are Mel Frequency Cepstral Coefficient (MFCC), Linear predictive coding (LPC), Linear Predictive Cepstral Coefficients (LPCC) and Perceptual Linear Prediction (PLP).

2.1 Mel frequency cepstral coefficient (MFCC)

MFCC is one of the most frequently used spectral based feature extraction techniques in the domain of speech recognition. It is one of the frequency domain techniques where features are derived from the perception of human hearing as Mel scale. We can achieve much more accurate features in respect to time domain features in this technique. The block diagram of MFCC is demonstrated in fig. 1.

The first step in the MFCC method is pre-emphasis. This filtering method is used to generate energy in a high frequency. It also decreases the low frequency energy [70]. FIR filtering method is generally used for this purpose. It is described in the Eq. (1).

$$y_n = x_n - a x_{n-1} \qquad 0 < a < 1 \tag{1}$$

To obtain transfer function, Z-transform on Eq. (1) is applied as described in Eq. (2).

$$P(z) = 1 - a z^{-1} \qquad 0 < a < 1 \tag{2}$$

Speech signal is divided into smaller section usually between 20 ms to 40.

ms in frame blocking [37]. There may be some problems like discontinuities of signals and spectral distortion in the framing process. Thereafter windowing is applied for minimization of those problems. Generally Hamming window concept is applied for that purpose. Output of windowing for a signal X(n) is as illustrated in eq. (3).

$$Y(n) = X(n).h(n) \qquad 0 \le n \le N-1 \tag{3}$$

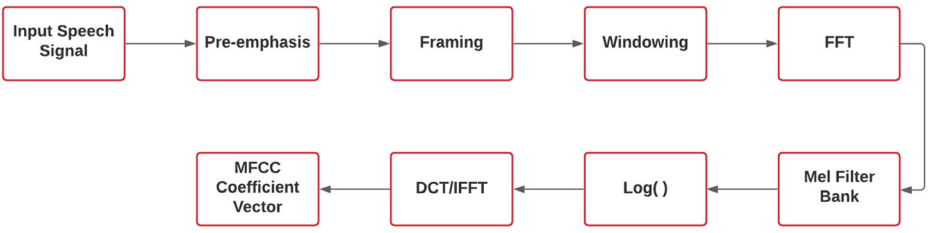

Fig. 1 Block Diagram of Conventional MFCC Feature Extraction Process

Here h(n) represents window function of n^{th} coefficient & N is the total number of samples in every frame [23]. Corresponding window function is illustrated in Eq. (4).

$$h(n) = 0.54-0.46\cos\left[\frac{2\pi n}{N-1}\right] \qquad 0\leq n\leq N-1 \tag{4}$$

Fast Fourier Transform (FFT) is opted for the windowing signal to convert it from time domain to frequency domain. Filter bank is the band pass filter that has overlapped triangular filter. Mel scale is imposed. Based on this, it is linear up to 1 KHz and logarithmic at greater frequencies. [10, 18]. MEL value is generated in Hz for a certain frequency f from the eq. (5) given bellow.

$$f[MEL] = 2595 \times \log_{10}\left[1 + {}^{f}/_{700}\right] \tag{5}$$

Logarithm is taken of the output of the Mel filter bank followed by the last step, i.e., Discrete Cosine Transform (DCT) for generating Mel Frequency Cepstral Coefficient (MFCC) features [5, 46]. It is possible to obtain derived MFCC features like delta MFCC (DMFCCs) and delta-delta MFCC (DDMFCCs). The first order derivatives are performed on MFCC to obtain delta MFCC (DMFCCs). Whereas the second order derivatives are required to derive delta-delta MFCC (DDMFCCs) [24].

2.2 Linear predictive coding and linear predictive cepstral coefficients

Linear predictive coding (LPC) is a well-known acoustic model widely used as low or medium bit rate coder in speech processing in noise free environment. The LPC works by calculating the power spectrum of signal. It is the well-known technique for formant estimation [6, 75]. In progression of feature extraction, the most popular LPCs act as different ways like: Coded-Excited LPC (CELP), Residual Excitation LPC, Pitch Excitation LPC, Multiple Excitation LPC (MPLPC) and Voice Excitation LPC [17].

Linear Predictive Cepstral Coefficients, abbreviated as LPCC is another established feature extraction method in speech signal processing. LPCC features can be derived from the cepstrum coefficients (CCs) computed in LPC analysis [57, 61]. The steps in LPC and LPCC are illustrated in fig. 2.

Autocorrelation method is proceeding on the frames of windowing speech signal. Now LPC analysis is done to get the LPC parameter again with additional cepstral coefficient computation is opted to arrive at the LPCC parameter.

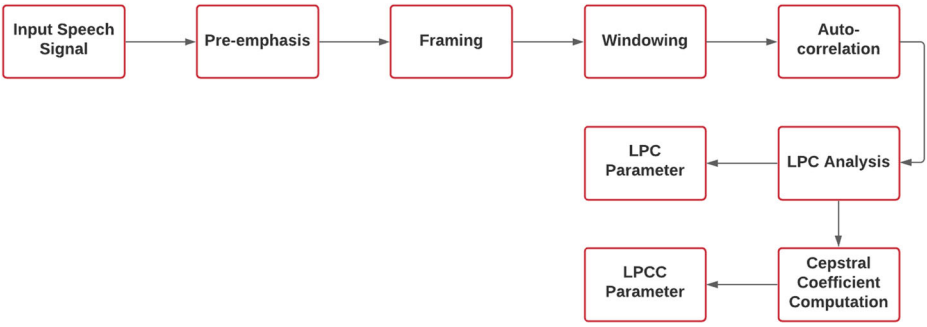

Fig. 2 Block Diagram of Conventional LPC and LPCC Feature Extraction Process

2.3 Perceptual linear prediction (PLP)

Perceptual Linear Prediction abbreviated as PLP model is developed considering the concept of psychophysics of hearing on human speech [22, 73]. PLP is used to get better speech recognition rate by discarding inappropriate speech information. PLP is nearly similar to LPC. Only difference is transformation of spectral characteristics that is employed correspond to the characteristics of human auditory system [10]. The step-by-step process is shown in fig. 3.

The speech spectrum is wrapped into the bark scale in critical band analysis phase [35]. The output spectrum is wrapped around with power spectrum and then resample. The simulated critical band filter is used for this purpose. In the next step, simulated equal loudness curve is introduced on the resulting sample to pre-emphasize. The cubic root amplitude compression is done in intensity loudness conversion. Finally, PLP parameters are generated after IFFT and autocorrelation step [22]. The relative spectra, known as RASTA, are often flowed with the PLP feature extraction to obtain the RASTA-PLP features [36]. The PLP and MFCC features are generally computed from a single-tapered spectrum estimate like Hamming-windowed periodogram spectrum estimate which has large variance. A set of different tapers, also called multi-taper spectral estimate, may be applied to minimize this large variance by averaging spectral estimates obtaining from those tapers [1].

3 Stuttered speech recognition: Traditional Machine Learning & Deep Learning based approaches

Finally, different classification and clustering methods are used to recognize stutter speech. Early studies on stuttered speech recognition mainly based on DTW score matching and traditional machine learning algorithm.

Before discussing different approaches to classify stuttering, we are going to recapitulate some basics of machine learning. Machine learning is basically a sub part of the broader family of Artificial Intelligence. Whereas deep learning is a next evolution of traditional machine learning. Architectural representation of Artificial Intelligence vs. Machine Learning vs. Deep Learning with different types of machine learning is given in the Fig. 4.

From the characteristics perspective of training data, Machine learning can be broadly categorized as supervised learning, unsupervised learning & reinforcement learning. These types of machine learning techniques are briefly discussed below [44, 45, 76].

Supervised Learning: Most widely used type of machine learning technique that learns from labeled training data to make predictions about learning targets.

Unsupervised Learning: It is a learning methodology that learns from training data that is neither classified nor labeled to group unsorted information according to similarities, patterns, and differences without any prior training of data.

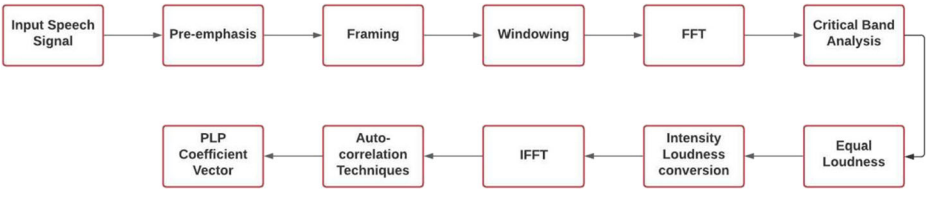

Fig. 3 Block Diagram of Conventional PLP Feature Extraction Process

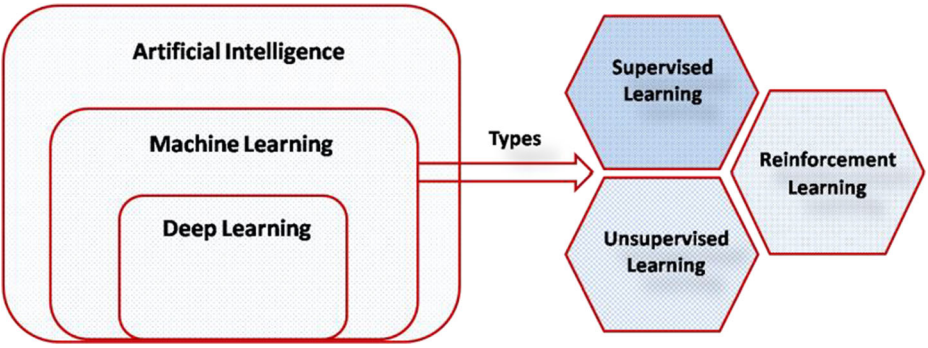

Fig. 4 Architectural Representation of Various Types of Learning Methods

Reinforcement Learning: A learning method that learns on its own feedback from the data which are in the form of sequences of actions, observations, rewards produced by interaction with a specific environment to determine the ideal behavior in order to maximize its performance.

There are different types of methods to solve different types of problem in Machine learning. Some of these are as follows:

Classification: It is a process to predict the categorization data in a specific manner from a given data set like 'YES' or 'NO'; 'TRUE' or 'FALSE'; 'A' or 'B' or 'C' etc. Algorithms falls under this method are KNN, Naïve Bayes, Decision Tree, Random Forest, Logistic Regression, Support Vector Machine.

Regression: It is a process to establish an equation among target variable and a set of independent variables. Linear regression is one of this type of algorithm.

Clustering: A process to group unsorted information according to similarities, patterns, and differences of a given data. K-means algorithm falls under this category.

Deep learning is part of an extensive family of artificial neural networks is also a rising field in the classification domain. Mostly, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) types of neural network are followed to design deep learning algorithms. Convolution neural networks, deep belief networks, deep neural networks and recurrent neural networks are in the wider family of deep learning architectures. These types of architectures have been structured to recognize stuttered speech [9, 59]. Different approaches for Stuttering recognition based on traditional machine learning algorithms like k-NN, LDA, k-means, SVM, HMM and deep learning algorithms are discussed in the next section.

3.1 Dynamic time warping (DTW)-based stuttered speech recognition

Dynamic time warping is one of the mostly used similarity measure algorithm between the two sequences those may differ in time or speed. Generally, similarities for temporal sequences are measured using DTW score matching algorithm. For time series Q and T of length n and m respectively are arranged as n-by-m matrix [3, 54]. Here

$$Q = q_1, q_2, q_3 \ldots\ldots q_n$$
$$T = t_1, t_2, t_3 \ldots\ldots\ldots\ldots\ldots t_m$$

The distance between the two points q_i and t_j is $d(q_i, t_j)$ in the (i^{th}, j^{th}) elements of the matrix. The absolute distance in terms of the Euclidean distance for the two sequences is calculated:

$$d(q_i, t_j) = \sqrt{(q_i - t_j)^2} \tag{6}$$

Now, the optimal match M(i, j) between i^{th} & j^{th} elements of the matrix is computed from the following equation:

$$M(i,j) = d(i,j) + \min[M(i-1,j-1), M(i-1,j), M(i,j-1)] \tag{7}$$

Dynamic time warping has been introduced by some researchers in context of stuttering speech recognition.12, 13, 26 and 39 dimensional MFCC features had been analyzed by Ramkumar KM and Ganesa in 2011 in their study. They concluded that 39 dimensional MFCC provides 84.58% accuracy in stutter speech assessment which is better compared to other dimensional MFCC. DTW score matching algorithm was used to identify stutter speech among those multidimensional features. They prepared their database from a group of people with age group 25 to 30 years by reading standard 150 words English passage and recorded. For training and testing purpose ten samples from the database were used [32]. An approach to identify repetition in stuttered speech was presented where the speech signal is segregated into isolated unit based on energy. The score from extracted Spectral (MFCC) and prosodic features (like formants and shimmer) from the isolated unit were analyzed to set a threshold using Dynamic Time Warping (DTW). Using this threshold value repeated events was identified [54]. Similarity matching using DTW classifier was employed to achieve 86% efficiency to assess automatic word repetition from UCLASS database [16]. Prolongation and repetition of stuttered speech was automatically recognized by concentrating of their epoch features. Glottal closure instants had been identified from the stuttering speech by considering the harmonic of the phase of ZFF speech. The voice onset time (VOT) for the stop consonants (/BA/, /DA/, /GA/, /PA/ and /TA/) was detected using ZFF to recognize and repetition of speech [42].

3.2 Stuttered speech recognition based on linear discriminant analysis and k-nearest neighbors algorithm

K-nearest neighbor (k-NN) algorithm, referred as lazy learning or instance-based learning, is utilized not only for classification but also for regression problems in aspect of pattern recognition. KNN is the easiest non-parametric learning calculation to group tests based on nearest training examples in the element space. The test's class is anticipated by the K training samples those are the closest neighbor to the test [64]. This strategy deals with two stages: a) deciding K close neighbors, b) utilizing these nearby neighbors, discovering class type [62]. Linear Discriminant Analysis classifier (LDA) has been broadly utilized for data classification and feature selection by determining hyperplanes in the area of speech recognition, face recognition and image retrieval. LDA chips away at by characterizing a solitary new composite variable, the discriminant score that is a mix of the original predictors by satisfying the boosting distinction between the predefined groups as for the new variable. Groupings are done in the conviction that each class will have a typical conveyance of discriminant scores however with the biggest conceivable distinction in mean scores for the classes [63]. In 2009, Lim Sin Chee et al. employed an approach to automatically detect two types of stuttered speech i.e., prolongations and repetitions. In the work, researchers annotated 10 samples of speech taken from the UCLASS chronicle manually as prolongation and repetition stuttered. LPCC features were extracted on this database. Two classifiers, LDA and k-NN were applied

to achieve detection accuracy of 89.77% [8]. LPC and Weighted LPCC (WLPCC) with addition to LPCC were engaged in the same classifiers to improve the recognition accuracy. It is seen that a superior recognition accuracy of 97.06% was accomplished for WLPCC compared to LPCC with 95.69% and LPC with 93.14% using k-NN classifier for 20 ms frames with sampling rates equal to 16 kHz. Whereas LDA classifier gives some much better accuracy i.e., 97.45% for WLPCC compared to LPCC with 95.10% and LPC with 90.39% for the same length frames to identify the stuttered events [19].

3.3 Stuttered speech recognition based on *K*-means clustering algorithm

K-means is in the category of Partitioning method of clustering techniques. This type of clustering is the simplest unsupervised learning algorithms for unlabeled data where n observations are partitioned into k cluster. The nearest mean of all the observations of a cluster is the prototype of the cluster. *K*-means clustering algorithm has been used in stuttering speech recognition. Speech format is corrected by removing the silent pause. Voiced speech has more energy in respect to unvoiced speech. *K*-means algorithm has been used to cluster those from MFCC features and VQ code book is generated. Score matching of dysfluent speech with the database was done by DTW algorithm [51].

3.4 Support vector machine (SVM) based classification of stuttered speech

Support Vector Machine (SVM) is a supervised machine learning model based on finding a hyperplane in N-dimensional space to classify the data points. Here N is represented as the total numbers of features. Though SVM is applied for both regression and classification it is extensively used for classification. The optimal hyperplane is derived by solving a quadratic optimization (QP) problem in order to perform linear classification. In addition to perform non-linear classification, a kernel function which is liable to engender linearly separable data in the feature space is introduced in SVMs to map the original input space into the higher dimensional feature space [52]. Support Vector Machine is widely used in speech recognition and very much significant in classification of stuttered speech. In 2009, KM Ravikumar et al. had worked on detection of syllable repetition to identify stuttered speech using SVM classifier. In their research, read speech of 15 stuttering people was used with MFCC as feature extraction method. Here, 93.45% of accuracy had been reached by this classifier that is better compared of their previous research work with HMM classifier [56]. SVM classifier with unimodal kernel function and multimodal kernel function had been used for speech of group of 16 stuttered speakers from UCLASS database to gain 96.133% and 96.4%accuracyrespectively. Two kernel functions with linear and RBF (Radial Basis Function) function and k-fold cross validation have been used [49]. Least Square Support Vector Machine (LS SVM) was applied to distinguish the repetition and prolongation speech in order to check the performance of the sample entropy feature for detection of the stammered event. UCLASS database has been used to get the accuracy of above 90% with maximum 96.84% in using db_2 wavelet packet filter in ERB scale [20]. It is seen that SVM is most suitable for recognition stuttering speech in respect to k-NN, LDA classification model. A comparison to classify speech disfluencies using k-NN, LDA and SVM classification was described and for that UCLASS database was used and obtained 10 ms, 20 ms and 30 ms frame length to get the best classification accuracies. MFCC, LPC, PLP feature extraction methods with 10-fold cross-validation was employed to get the best accuracy of 95% by SVM classifier [14]. Multi-class Support Vector Machine (SVM)

classifier was applied for identification of three types of stuttering such as syllable repetition, word repetition and prolongation. Researchers calculated an optimal hyper-plane using the concept of "one vs rest" method for dealing with multiple classes. They utilized LPC, LPCC and MFCC parameterization techniques to achieve recognition accuracy 75.00%, 92.00% and 88.00% respectively [39]. It is significant that by combining prosodic features and cepstral features, the performance of SVM classifier on recognition disfluency was improved. Prosodic features like pitch, energy, and duration and cepstral feature like MFCC, delta MFCC and delta–delta MFCC were used. The combination of DDMFCC and prosodic features were responsible to achieve the best disfluency recognition efficiency rate of 96.85% using SVM classifier among all these features [40]. SVM classifier had been used to find the percentage of disfluencies associated with stuttered speech. For that MFCC feature extraction method was employed upon 20 samples from UCLASS database to get the percentage of disfluency from different number of syllables [53]. SVM classification was applied to classify disfluent speech with accuracy 90% and fluent speech with accuracy 96.67% using from MFCC feature extraction method. Standard UCLASS database was used to get the result [33]. Arya A Surya and Surekha Mariam Varghese discussed three methods to recognize the stuttered speech. They proposed following three methods – Supervised model for stuttered speech recognition, Stuttered speech recognition by stuttering pruning and automated text-to-speech based stuttered speech recognition. Data samples from University College London Archive of Stuttered Speech (UCLASS) and National Institute of Speech and Hearing (NISH) had been employed to get the 76% accuracy from first method, the 62% accuracy from second method, the 80% accuracy from third method [65]. A supervised sparse feature learning approach was presented on functional near infrared spectroscopy (fNIRS) brain imaging data that is recorded during a speech to discover discriminative biomarkers. Traditional machine learning algorithms like support vector machine (SVM), k-nearest neighbor (kNN), decision tree, ensemble, and linear discriminant (LDA) was applied for classification. The model was capable of differentiating neural activation patterns between stuttered and non-stuttered with a precision of 87.5% predicated on a five-fold cross-validation procedure using support vector machine (SVM) [25].

3.5 Hidden Markov model (HMM) based classification of stuttered speech

The HMMs are stochastic models and are widely used in the field of pattern recognition, especially in speech recognition. The HMM can be represented as extension of the Markov Model. The main structure of this model is that the current state is hidden whereas only the output is observed. An HMM is defined by the following components: states (Q), transition probability matrix (A), observations (O), observation likelihoods, also called emission probabilities (B) and initial probability distribution (π). The probability of the model being in a given state can be resolute by visual examination of the output of the HMM. Influent speech was identified automatically using HMM by transferring the speech into a formal grammar. Speech of 37 patients with the text "Northwind and sun" was taken as database to consider the duration of the detected pause. As interval of stuttered speech, the sum of all the detected pause had been considered to distinct between non stuttered even, stuttered with many repetitions i.e., short pause and stuttered with few repetitions i.e., long pauses [47]. Hidden Markov Model (HMM) technique was utilized on MFCC features for evaluating children speech stuttering problem. 20 normal speeches and 15 artificial stuttered speeches were created by recording of Malay language word "Sembilan" from 7 males and 3 female speakers. The

samples were used in the model for training and testing purpose. By setting the threshold, their model gave 96% of recognition rate for normal speech and 90% for artificial stutter speech [68]. Marek WISNIEWSKI et al. presented an approach with Hidden Markov Model (HMM) to recognize speech disorders - prolonged fricative phonemes. MFCC was used for parameterization of the acoustic signal. To minimize the number of parameters, encoding with several codebooks with sizes 30, 38, 64, 128, 256, 512 were prepared. Prolonged fricative phonemes (^, s, z, x, ¥, v, •, f) were chosen and corresponding 5 fragments were prepared for every phoneme containing only the prolongation. Training vectors for this model were created by encoding every group of fragments with the earlier prepared codebook. Several HMM models with sizes of 5, 8, 10 and 15 states were used for test [71]. They also worked on recognition of blockades with repetition of stop phonemes along with prolongation of fricative phonemes using HMM [72].

3.6 Artificial neural network (ANNs) and deep neural network based classification of stuttered speech

Focusing on the functioning of biological nervous systems, one mathematical model is Artificial Neural Networks (ANNs). These models are structured as three parts namely input layer, hidden layers and output layers. ANNs signify as weighted directed graph where artificial neurons represent as nodes and connections between neuron input and neuron output as directed edges with weights. Based on connection pattern ANN categorization has been cited as feed-forward networks and recurrent networks. If there is no loop in the graph then those networks are categorized as single layer perceptron, multilayer perceptron and radial basis function network of feed forward network types. Whereas if there is loop in the graph then those networks are called as recurrent types of networks such as Kohonen's SOM, Competitive networks, Hopfield network, ART models [4, 11, 29, 60]. Various deep learning architectures like convolutional neural network, deep neural network, deep belief network and recurrent neural networks are significant in stuttering speech recognition. Peter Howel and Stevie Sackin, in their study trained an artificial neural network model for recognizing repetitions and prolongations in stuttered speech. The network was fully interconnected. For differentiating repetition and prolongation, separate artificial neural networks were trained. Two types of acoustic inputs were taken where one contains combination of autocorrelation function (ACF) and spectral information and other contains Envelope parameters. The weights of networks were changed automatically to hold the mapping between input and output. Under this artificial neural network, the repetition and prolongation were distinguished as severe, moderate or mild with different hit/miss rate depending on different input parameters [26]. ANN was employed to predict classification of normal speech and stuttering with 92% of accuracy. In the study, two groups of data were engaged where Group-I data (involving 25 stuttered children) was employed to train the ANN and Group-II (involving 26 stuttered children) for predicting the model. Ages, sex, frequency, duration etc. types of same ten features from the two groups of data were exercised. The multilayer perceptron classifier was employed for classification [15]. K. M. Ravikumar et al. proposed a four-stage automated approach consisting of segmentation, feature extraction, score matching and decision logic for detecting repetition. MFCC features were extracted whereas DTW was implemented for score matching. The decision was made by using perceptron classifier based on the DTW score to get the result with 83% of accuracy [55]. Neural networks had been employed in two stages to classify speech as fluent and non-fluent. In the first stage, Kohonen network was applied

whereas in the second stage Multilayer Perceptron and Radial Basis Function network were utilized. 59 fluent speeches of 4 fluent speakers and 59 non-fluent speeches of 8 stuttering people were obtained as data samples. Using all networks classification, correctness was achieved in the range between 88.1% and 94.9% [66]. I. Szczurowska et al. considered the neural network to categorize non-fluent and fluent utterance in their study. Same fragment duration with same no of fluent speech were implemented in the network at two stages for 4-s fragments of 40 number having blockades in pronunciation in words stating with the consonants (p, b, t, d, k and g) and repetition of 1 to 11 stop consonants. For decreasing the dimension of the input signal, Kohonen network consisting of 21 input neurons and 25 output neurons was used first. Then multilayer perceptron including 171 input neurons, 53 hidden layers and one output layer was examined to achieve 76.67% classification accuracy [67]. In order to identify dysfluencies in stuttered speech of children, a two-stage technique was used to build an automatic recognition method. In the first stage, speech was segmented in words and words were classified as fluent or disfluent using ANN classifier. The approach for recognizing part and whole word repetition, prolongations and broken word repetition was addressed based on various features in two phases comprise lexical disfluency (LD) for single word and supralexical disfluency (SD) for group of words. In this approach, fluent words were classified with 95% of accuracy whereas dysfluent words were classified with 78% of accuracy [27]. Based on back propagation algorithm, a multilayer feed forward network was structured for identifying of repetition and prolongation type of stuttering. Test features as MFCC, formants, pitch, zero crossing rate and energy were chosen. The model achieved 87.39% accuracy for these types of recognition [58]. Repetitions, prolongations and blocks types of speech disorders were predicted by using an effective ANN approach namely Adaptive Optimization based Artificial Neural Network (AOANN). For training and testing purpose, recordings of 20 participants using PRAAT tool with 44 KHz sampling rate was employed to this model. MFCC features were employed to test the effectiveness of the model. A comparative study among their proposed model with other existing approaches like PSO-ANN (Particle Swarm Optimization-ANN), GA-ANN (Genetic Algorithm-ANN) and default ANN in terms of MSE (Mean Square Error) and RMSE (Root Mean Square Error) metric were explained [43]. DNN is one type of artificial neural network linked to a number of completely connected hidden layers to produce output class from input vector using probability distribution function. DNN can be modified as deep convolution neural networks (DCNNs) and deep local united neural networks (DLUNNs) by introducing different types of regularization. Performance and comparison of DNNs by introducing several metrics and final speech recognizer word error rates were explained in the research. Training DNNs with discriminative loss functions for speech tasks using DNN optimization were also served [38]. Combination of four pre-trained Bernoulli-Bernoulli restricted Boltzman machine (RBMs) and a decision layer type of deep belief network (DBN) were used to create a DBN-DNN classifier. Stacey Oue et al. investigated different types of input features used to feed into this DBN-DNN classifier to detect stuttering in their research. Mainly MFCCs and LPCCs features were used to achieve 86% recognition accuracies in dysarthric speech and 84% for non-dysarthric speech [48]. A deep neural network (NN) of 18 convolution layers along with residual layers followed Bi-LSTM units was proposed in order to detect and identify various forms of stutter. Spectrogram features were used to train the model. Each recording from UCLASS dataset was annotated manually as different types of stuttering. Two recurrent layers, each of 512 bidirectional LSTM units, had been added to attain a typical miss rate of 10.03% [34].

4 Discussion and analysis

4.1 Discussion

In this study, some of the recent and sophisticated techniques used in stuttering speech recognition and classification are discussed. It is found that the significance of machine learning and deep learning models are prominent in the context of classification of stuttering speech. Mostly LPC, LPCC, PLP and MFCC feature extraction methods were employed in the previous studies. Whereas appearance of three major classifier HMMs, SVM and ANNs was notable in recognition of stuttering speech. Most of the researchers employed UCLASS (University College London's Archive of Stuttered Speech) dataset to train and test their model. Three UCLASS data set (Release One, Release Two and FSF) are available in the archive. Release One dataset contains 139 samples of monologues form. Participants were in the age group of 8 to 18 years. On the other hand, recordings of monologs, readings and conversation are in the release two dataset. 43 different participants contributed 107 recordings to create this dataset [28, 69]. In addition, it is also observed that out of the most common 6 types of stuttering (i.e., interjection, revisions, incomplete phrases, repetition, prolonged sounds, broken words) only two or three types of stuttering were recognized in most of the previous studies. More focus on these problems, we may get better recognition with better accuracy in the field of stuttering speech recognition in future. With addition, to incorporate the intelligent model in mobile communication, may consider data hiding and watermarking types of security approaches [2, 12, 13, 30]. Based on different types of stuttering, summary of several research works related to stuttering speech recognition are depicted in tabular form.

Accuracy of different researches considering different features, classifier and database are depicted in Table 1 for recognizing stutter and non-stutter type of stuttering.

Corresponding author vs accuracy graph is presented in Fig. 5. It can be seen that maximum accuracy for identifying stutter and non-stutter type of stuttering is 96%. Tian-Swee Tan et al. applied HMM model and MFCC features to achieve the recognition accuracy. But they worked with small no of data; among some data are artificial stutter speech samples. They did not focus on the different types of stuttering identification.

Accuracy of different researches considering different features, classifier and database are depicted in Table 2 for recognizing repetition type of stuttering. Corresponding author vs accuracy graph is presented in Fig. 6. It is seen that highest accuracy for identifying repetition type of stuttering is 96.4%. Juraj Palfy et al. employed SVM classifier and MFCC, PCA based features to achieve the recognition accuracy. But they worked with small no of data. They did not focus on the different other types of stuttering identification.

Accuracy of different researches considering different features, classifier and database are depicted in Table 3 for recognizing repetition and prolongation type of stuttering. Corresponding author vs accuracy graph is presented in Fig. 7. It is seen that highest accuracy for identifying repetition type of stuttering is 96.85%. P. Mahesha et al. employed SVM classifier and Cepstral and prosodic features to achieve the recognition accuracy. But they worked with small no of data. They did not focus on the different other types of stuttering identification.

Accuracy of different researches considering different features, classifier and database are depicted in Table 4 for recognizing repetition type of stuttering. It is seen that deep learning algorithm can be significant for classify different types of stuttering.

Table 1 Summary on Stuttered and Non-Stuttered Recognition

Database Type	Database Description	Approach	Features	Classifier	Best Outcome (Recognition Accuracy)
Standard	UCLASS database	Vikhyath Narayan K N et al. [33]	MFCC feature	SVM	90%
	The TORGO database: recordings from 7 speakers	Stacey Oue et al. [48]	MFCCs, LPCCs	DBN-DNN	86%
Derived	fNIRS data from the 32 children participants	Rahilsadat Hosseini et al. [25]	An extended set of features, identified by MISGL and MIL	LDA, decision tree, k-NN, SVM and ensemble	87.5%
	Malay Language data:20 normal speech & 15 artificial stutter speech samples	Tian-Swee Tan et al. [68]	MFCC	HMM	96%
	Recordings by two groups of disfluent children with good knowledge of Kannada	Y.V. Geetha et al. [15]	10 input variables like age, sex, frequency, duration, historical, attitudinal and behavioral scores etc.	ANN	92%
	59 fluent recordings of four fluent speakers and 59 non-fluent recordings of eight stuttering people	Izabela Swietlicka et al. [66]	Samples were analyzed by FFT 512 and an A-weighting filter	Kohonen network, Multilayer Perceptron and Radial Basis Function network	94.9%
	40 fragments each of 4-s	I. Szczurowska et al. [67]	Samples were analyzed by FFT 512 and an A-weighting filter	Kohonen network, Multilayer Perceptron	76.67%

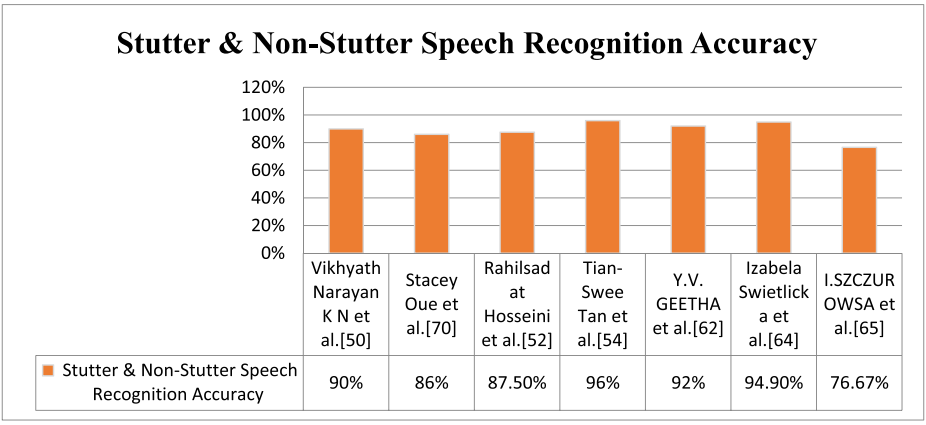

Fig. 5 Stutter & Non-Stutter Speech Recognition Accuracy

Table 2 Summary on Repetition Type of Stuttering

Database Type	Database Description	Approach	Features	Classifier	Best Outcome (Recognition Accuracy)
Standard	10 recordings from the UCLASS database	Girish M et al. [16]	MFCC features	DTW	86%
	16 recordings from UCLASS database	Juraj Palfy et al. [49]	MFCC features, PCA, kernel PCA and MFCC based derived features	SVM classifier with unimodal and multimodal kernel function	96.4%
Derived	10 recordings	Ramkumar KM et al. [32]	MFCC (12,13,26,39 dimensional)	Threshold value using DTW Score matching	84.58%
	Recordings of 27 s consist of 50 repetition events.	Pravin B. Ramteke et al. [54]	Spectral (MFCC) and prosodic features (like formants and shimmer)	Threshold value using DTW Score matching	94%
	Recordings of 15 stuttering speakers	KM Ravikumar et al. [56]	MFCC	SVM	93.45%
	Recordings of the text "Northwind and sun" by 37 patient	E. North et al. [47]	Duration and frequency of dysfluent portions, speaking rate	HMM	–
	Recordings of 150 English words by 10 speakers	K. M. Ravikumar et al. [55]	MFCC	ANN with DTW	83%

4.2 Analysis

Generally, six types of stuttering are observed such as interjection, revisions, incomplete phrases, repetition, prolongation and broken words among the stuttered speech. Repetition is also subcategorized as phrase–repetitions, word-repetitions, part-word repetitions. Automatic identification or assessment of all these category of stuttering is the main focused area among

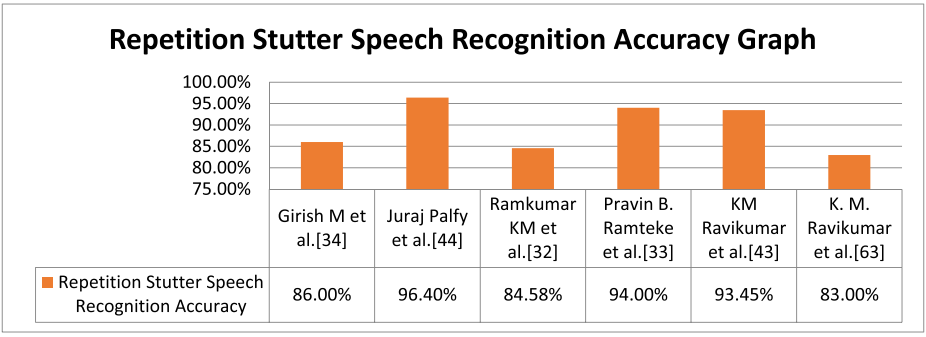

Fig. 6 Repetition Speech Recognition Accuracy

Table 3 Summary on Repetition and Prolongation Type of Stuttering Recognition

Database Type	Database Description	Approach	Features	Classifier	Best Outcome (Recognition Accuracy)
Standard	10 audio samples from UCLASS database	Lim Sin Chee et al. [8]	LPCC features	LDA and k-NN	89.77%
	10 audio samples from UCLASS database	M. Hariharan et al. [19]	LPC, LPCC,WLPCC (Weighted LPCC)	k-NN and LDA	97.45%
	39 audio samples from UCLASS database	M. Hariharan et al. [20]	sample entropy features using Mel scale, Bark scale and ERB scale	Least Square SVM	96.84%
	39 audio samples from UCLASS database	Chong Yen Fook et al. [14]	MFCC, LPC,PLP	k-NN, LDA and SVM	95%
	20 audio samples from UCLASS database	P. Mahesha et al. [39]	LPC, LPCC and MFCC	Multi-class SVM	92.00%
	30 audio samples from UCLASS database	P. Mahesha et al. [40]	Cepstral and prosodic features	SVM	96.85%
	audio samples from UCLASS and NISH	Arya A Surya et al. [65]	MFCC feature	SVM, Neural Network, ANN	80%
Derived	2 min recordings for training and with addition five further speaker's recordings for test purpose.	Peter Howell et al. [26]	Envelope parameters, Autocorrelation function (ACF) coef-ficients and spectral information	ANN	–
	Recordings (376 words) of English passage "Arthur the rat" from 12 children participants	Peter Howel et al. [27]	Features based on pattern of alternating energy and spectral change over time	ANN	78%
	Total 78 Hindi language segments: repetitions-32, prolongation-18 and normal speech-28	P.S. Savin et al. [58]	Formants, zero crossing rate (ZCR), pitch, Energy and Mel-frequency cepstral coefficients(MFCCs)	Artificial Neural Networks (ANN)	87.39

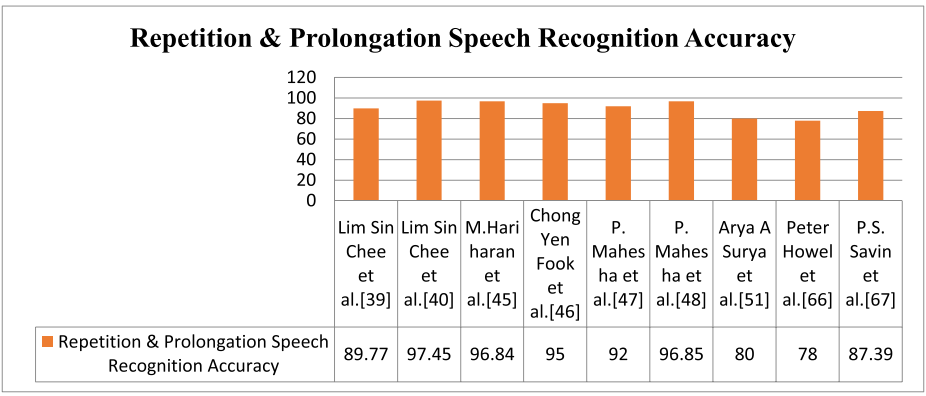

Fig. 7 Repetition & Prolongation Recognition Accuracy

Table 4 Summary on Stuttering (more than two types)

Database Type	Database Description	Approach	Features	Classifier	Stuttering Types	Best Outcome (Recognition Accuracy)
Standard	UCLASS and synthetic dataset from original non stuttered Libri Speech dataset.	Tedd Kourkounak et al. [34]	Spectrogram Features	A deep neural network: 18 convolution layers, residual layers and bidirectional long short term memory (Bi-LSTM) units	Sound Repetition Word Repetition Phrase Repetition Revision Interjection Prolongation	84.10 96.60 95.54 97.14 81.40 94.08
	UCLASS database: 20 audio recordings	Raghavendra M et al. [53]	MFCC feature	SVM	repetition, prolongation and interjection	–
Derived	Three speech samples, each of 54 s in size from three different speakers	Marek Wisniewski et al. [72]	MFCC	HMM	Prolongation, repetition and blockades	70%
	Recordings of 20 participants	G. Manjula et al. [43]	MFCC	AOANN (Adaptive Optimization based ANN)	Repetitions, Prolongations and blockades	–

the researchers over past few years. Some researches on stuttering speech recognition are based on only identification of stutter and non-stutter speech. It is seen that highest accuracy for identifying stutter and non-stutter speech is 96%. Tian-Swee Tan et al. applied HMM model and MFCC features to achieve the recognition accuracy. But they worked with small no of derived data which contain of some amount of artificial stutter speech samples. Different types of stuttering were not distinguished by this model. Some researchers worked on identifying repetition type of stuttering. Among them, the model, designed by Juraj Palfy et al. by employing SVM classifier with unimodal and multimodal kernel function, was able to achieve 96.4% repetition accuracy. They employed MFCC features along with PCA, kernel PCA and MFCC based derived features into the model. But they also experimented with small dataset from standard UCLASS database. Lim Sin Chee et al. experimented with weighted LPCC features and were able to design a model to identify repetition and prolongation types of stuttering with 97.45% of accuracy. But the size of experimented dataset containing 10 audio samples from standard UCLASS database was very small. A deep NN with 18 convolution layers, residual layers and bidirectional long short term memory (Bi-LSTM) units based model using spectrogram features was designed by Tedd Kourkounak et al. to identify repetition (sound, word, phrase), revision, interjection and prolongation types of stuttering. They applied synthetic data with some data from standard UCLASS database. So deep learning based model with derived features can be applied to identify all types of stuttering. In the previous works, there may be some reliability issues of the outcome due to small amount of data. The amount of data for training purpose should be increased to increase the accuracy. To increase the database size, synthetic data can be created by using some algorithms like Generative Adversarial Networks (GANs).

5 Conclusion and future work

Speech is the communication carrier to express human thoughts, feelings and ideas. Stuttering, or stammering is a disorder of speech which affects millions of people in the glove. In the field of stuttered speech recognition, different machine learning models were applied for analysis and classification over the last few decades. In this study, different machine learning and deep learning models with their application in stuttered speech recognition are discussed. The 3 major classifiers i.e., ANNs, HMMs and SVM have been used to classify different types of stutterers. Deep learning algorithms have become very popular nowadays over traditional machine learning algorithms for stuttering speech recognition, discussed briefly in this study. The major challenges like small volume unlabeled data, similarity between different stuttering classes are observed. Moreover, an input speech file sometimes contains more than one types of stuttering which creates difficulties on labeling. Most of the research had been concentrated on prolongation and repetition types of stuttering. Some work on Interjection types of stuttering was also done but work on classification of broken words, revisions, incomplete phrases types of stuttering is almost nil. Most of the researchers labeled different no of stuttered speech from UClASS database manually in order to train their model. Different features like LPC, LPCC, PLP and MFCC were used in the previous researches to train and test the models among them MFCC features was extensively used. Reviews and comparisons of earlier researches have been highlighted in this paper. Accuracy in respect to recognition and correction of stuttering speech may be improved by employment of modified feature extraction algorithm and different deep learning based algorithms on large database.

Deep neural network can be employed to classify different types of stuttering with better accuracy. There are very few researches on removing of stuttering of different types from a speech signal. Identification of stuttering is required but main focused should be on removal of stuttering. Interjection, prolongation type of stuttering and unvoiced speech can be removed by different ways. Threshold amplitude is one of the ways to remove those types of stuttering. Considering lower energy of a speech signal and by removing those parts, one speech can be interjection and prolongation types of stuttering free. Natural language processing can be introduced to get repetition types of stuttering free speech. For that different existing Text-to-Speech (TTS) system can be used. Silent pause within a word is broken words types of stuttering. Silent pause can be removed by considering amplitude thresholding, lower energy. But identifying silent pause within a word is the main challenge. Natural language processing (NLP) concept must be introduced for that purpose. NLP must be a major part to remove revisions, incomplete phrases types of stuttering. Accuracy of stuttering recognition may be improved by using modified feature extraction algorithm. By using Generative Adversarial Networks (GANs) algorithm, synthetic data can be created to increase the database size. Then different deep learning based algorithms may be employed to achieve better stuttered speech recognition accuracy. The model can be designed for multilingual system.

Funding There is no funding for this research work.

Declarations

Conflicts of interests/competing interests The authors want to declare that, there are no conflicts of interests / competing interests in this research work.

References

1. Alam MJ, Kinnunen T, Kenny P, Ouellet P, O'Shaughnessy D (2013) Multitaper MFCC and PLP features for speaker verification using i-vectors. Speech Comm 55(2):237–251
2. Alanazi F, Elhadad A, Hamad S, Ghareeb A (2019) Sensors data collection framework using mobile identification with secure data sharing model. Int J Electrical Comput Eng 9(5):4258
3. Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. In *KDD workshop* (Vol. 10, no. 16, pp. 359-370).
4. Bhattacharya S, Das N, Sahu S, Mondal A, & Borah S. (2020). Deep classification of sound: A concise review. First doctoral symposium on natural computing research(DANCER-2020), Springer, India.
5. Boulmaiz A, Messadeg D, Doghmane N, Taleb-Ahmed A (2017) Design and implementation of a robust acoustic recognition system for waterbird species using TMS320C6713 DSK. Int J Ambient Comput Intell (IJACI) 8(1):98–118
6. Buza O, Toderean G, Nica A, Caruntu A (2006) Voice signal processing for speech synthesis. In 2006 IEEE international conference on automation, quality and testing, robotics (Vol. 2, pp. 360-364). IEEE.
7. Chee LS, Ai OC, Yaacob S (2009) Overview of automatic stuttering recognition system. In proc. international conference on man-machine systems, no. October, Batu Ferringhi, Penang Malaysia (pp. 1-6).
8. Chee LS, Ai OC, Hariharan M, Yaacob S (2009) Automatic detection of prolongations and repetitions using LPCC. In 2009 international conference for technical postgraduates (TECHPOS) (pp. 1-4). IEEE.
9. Das N, Chakraborty S, Chaki J, Padhy N, Dey N (2020) Fundamentals, present and future perspectives of speech enhancement. IntJ Speech Technol. 1-19.
10. Dave N (2013) Feature extraction methods LPC, PLP and MFCC in speech recognition. Int J Advan Res Eng Technol 1(6):1–4
11. Dey N (2019) Intelligent speech signal processing, 1st edn. Academic Press

12. Elhadad A, Hamad S, Khalifa A, Ghareeb A (2017) High capacity information hiding for privacy protection in digital video files. Neural Comput Applic 28(1):91–95
13. Elhadad A, Ghareeb A, Abbas S (2021) A blind and high-capacity data hiding of DICOM medical images based on fuzzification concepts. Alexandria Eng J 60(2):2471–2482
14. Fook CY, Muthusamy H, Chee LS, Yaacob SB, Adom AHB (2013) Comparison of speech parameterization techniques for the classification of speech disfluencies. Turkish J Electrical Eng Comput sci 21(sup. 1): 1983–1994
15. Geetha YV, Pratibha K, Ashok R, Ravindra SK (2000) Classification of childhood disfluencies using neural networks. J Fluen Disord 25(2):99–117
16. Girish M, Anil R, Ahmed A, & Hithaish Kumar M (2017). Word repetition analysis in stuttered speech using MFCC and dynamic time warping. National Conference on Communication and Image Processing TJIT, Bangalore.
17. Gupta H, Gupta D (2016) LPC and LPCC method of feature extraction in speech recognition system. In 2016 6th international conference-cloud system and big data engineering (confluence) (pp. 498-502). IEEE.
18. Gupta S, Jaafar J, Ahmad WW, Bansal A (2013) Feature extraction using MFCC. Signal Image Process: Int J (SIPIJ) 4(4):101–108
19. Hariharan M, Chee LS, Ai OC, Yaacob S (2012) Classification of speech dysfluencies using LPC based parameterization techniques. J Med Syst 36(3):1821–1830
20. Hariharan M, Vijean V, Fook CY, Yaacob S (2012) Speech stuttering assessment using sample entropy and Least Square support vector machine. In 2012 IEEE 8th international colloquium on signal processing and its applications (pp. 240-245). IEEE.
21. Healey EC (2010) What the literature tells us about listeners' reactions to stuttering: implications for the clinical management of stuttering. Sem Speech Language 31, no. 04, pp. 227-235). © Thieme Medical Publishers.
22. Hermansky H (1990) Perceptual linear predictive (PLP) analysis of speech. J Acoust Soc Am 87(4):1738–1752
23. Hidayat R, Bejo A, Sumaryono S, Winursito A (2018) Denoising speech for MFCC feature extraction using wavelet transformation in speech recognition system. In 2018 10th international conference on information technology and electrical engineering (ICITEE) (pp. 280-284). IEEE.
24. Hossan MA, Memon S, Gregory MA (2010) A novel approach for MFCC feature extraction. In 2010 4th international conference on signal processing and communication systems (pp. 1-5). IEEE.
25. Hosseini R, Walsh B, Tian F, Wang S (2018) An fNIRS-based feature learning and classification framework to distinguish hemodynamic patterns in children who stutter. IEEE Trans Neural Syst Rehabil Eng 26(6):1254–1263
26. Howell P, Sackin S (1995) Automatic recognition of repetitions and prolongations in stuttered speech. In proceedings of the first world congress on fluency disorders (Vol. 2, pp. 372-374). Nijmegen, the Netherlands: university press Nijmegen.
27. Howell P, Sackin S, Glenn K (1997) Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II. ANN recognition of repetitions and prolongations with supplied word segment markers. J Speech, Language, Hearing Res 40(5):1085–1096
28. Howell P, Davis S, Bartrip J, Wormald L (2004) Effectiveness of frequency shifted feedback at reducing disfluency for linguistically easy, and difficult, sections of speech (original audio recordings included). Stammer Res: On-Line J Publish Brit Stamm Assoc 1(3):309
29. Jain AK, Mao J, Mohiuddin KM (1996) Artificial neural networks: A tutorial. Computer 29(3):31–44
30. Khalil OH, Elhadad A, Ghareeb A (2020) A blind proposed 3D mesh watermarking technique for copyright protection. Imaging Sci J 68(2):90–99
31. Khan N (2015) The effect of stuttering on speech and learning process, A case study. Int J Stud English Language Literature (IJSELL) 3(4):89–103
32. Km RK, Ganesan S (2011) Comparison of multidimensional MFCC feature vectors for objective assessment of stuttered disfluencies. Int J Adv Netw Appl 2(05):854–860
33. KN VN, Meharunnisa SP (2016) Detection and analysis of stuttered speech. Int J Adv Res Electronics Comm Eng (IJARECE) 5(4):2278–909X
34. Kourkounakis T, Hajavi A & Etemad A (2020). FluentNet: end-to-end detection of speech disfluency with deep learning. *arXiv preprint arXiv:2009.11394*.
35. Kumar P, Biswas A, Mishra AN, Chandra M (2010) Spoken language identification using hybrid feature extraction methods. arXiv preprint arXiv:1003.5623.
36. Li Q, Huang Y (2010) An auditory-based feature extraction algorithm for robust speaker identification under mismatched conditions. IEEE Trans Audio Speech Lang Process 19(6):1791–1801

37. Likitha MS, Gupta SRR, Hasitha K, Raju AU (2017) Speech based human emotion recognition using MFCC. In 2017 international conference on wireless communications, signal processing and networking (WiSPNET) (pp. 2257-2260). IEEE.
38. Maas AL, Qi P, Xie Z, Hannun AY, Lengerich CT, Jurafsky D, Ng AY (2017) Building DNN acoustic models for large vocabulary speech recognition. Comput Speech Lang 41:195–213
39. Mahesha P, Vinod DS (2013) Classification of speech dysfluencies using speech parameterization techniques and multiclass SVM. In international conference on heterogeneous networking for quality, reliability, security and robustness (pp. 298-308). Springer, Berlin, Heidelberg.
40. Mahesha P, Vinod DS (2015) Combining cepstral and prosodic features for classification of disfluencies in stuttered speech. In intelligent computing, communication and devices (pp. 623–633). Springer, New Delhi
41. Manjula G, Kumar S (2016) Overview of Analysis and Classification of Stuttered Speech Proceed 11th IRF Int Conf
42. Manjula G, Kumar MS, Geetha YV, Kasar T (2017) Identification and validation of repetitions/ prolongations in stuttering speech using epoch features. Int J Appl Eng Res 12(22):11976–11980
43. Manjula G, Shivakumar M, Geetha YV (2019) Adaptive optimization based neural network for classification of stuttered speech. In Proceedings of the 3rd international Conference on Cryptography, Security and Privacy (pp. 93-98).
44. Meenakshi M (2020) Machine learning algorithms and their real-life applications: A survey. Available at SSRN 3595299
45. Mirri S, Delnevo G, Roccetti M (2020) Is a COVID-19 second wave possible in Emilia-Romagna (Italy)? Forecasting a future outbreak with particulate pollution and machine learning. Computation 8(3):74
46. Mohan BJ (2014) Speech recognition using MFCC and DTW. In 2014 international conference on advances in electrical engineering (ICAEE) (pp. 1-4). IEEE.
47. Nöth E, Niemann H, Haderlein T, Decher M, Eysholdt U, Rosanowski F, Wittenberg T (2000) Automatic stuttering recognition using hidden Markov models In Sixth International Conference on Spoken Language Processing
48. Oue S, Marxer R, Rudzicz F (2015) Automatic dysfluency detection in dysarthric speech using deep belief networks. In proceedings of SLPAT 2015: 6th workshop on speech and language processing for assistive technologies (pp. 60-64).
49. Pálfy J, Pospíchal J (2011) Recognition of repetitions using support vector machines. In signal processing algorithms, architectures, *arrangements, and applications SPA 2011* (pp. 1-6). IEEE.
50. Pinelli P (1992) Neurophysiology in the science of speech. Curr Opinion Neurol Neurosurg 5(5):744–755
51. Prakash CO, Sai YP, Kumar VN (2018) Design and implementation of silent pause stuttered speech recognition system
52. Qi F, Bao C, Liu Y (2004, December) A novel two-step SVM classifier for voiced/unvoiced/silence classification of speech. In *2004* international symposium on Chinese spoken language processing (pp. 77-80). IEEE.
53. Raghavendra M, Rajeswari P (2016) Determination of disfluencies associated in stuttered speech using MFCC feature extraction. *Comput*. Speech Lang, IJEDR 4(2):2321–9939
54. Ramteke PB, Koolagudi SG, Afroz F (2016). Repetition detection in stuttered speech. In Proceedings of 3rd international conference on advanced computing, networking and informatics (pp. 611–617). Springer, New Delhi
55. Ravikumar KM, Reddy B, Rajagopal R, Nagaraj H (2008) Automatic detection of syllable repetition in read speech for objective assessment of stuttered disfluencies. Proceed World Acad Sci, Eng Technol *36*:270–273
56. Ravikumar KM, Rajagopal R, Nagaraj HC (2009) An approach for objective assessment of stuttered speech using MFCC features. ICGST Int J Digital Signal Process, DSP 9(1):19–24
57. Revada LKV, Rambatla VK, Ande KVN (2011) A novel approach to speech recognition by using generalized regression neural networks. Int J Comput Sci Issues (IJCSI) 8(2):484
58. Savin PS, Ramteke PB & Koolagudi SG (2016). Recognition of repetition and prolongation in stuttered speech using ANN. In proceedings of 3rd international conference on advanced computing, networking and informatics (pp. 65–71). Springer, New Delhi
59. Sen S, Dutta A, Dey N (2019) Audio processing and speech recognition: *concepts*. Springer, Techniques and Research Overviews
60. Sen S, Dutta A, Dey N (2019) Speech processing and recognition system. Audio Processing and Speech Recognition. Springer Briefs in Applied Sciences and Technology. Springer, Singapore
61. Sharma U, Maheshkar S, Mishra AN (2015) Study of robust feature extraction techniques for speech recognition system. In 2015 international conference on futuristic trends on computational analysis and knowledge management (ABLAZE) (pp. 654-658). IEEE.

62. Shirvan RA, Tahami E (2011) Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method. In 2011 18th Iranian conference of biomedical engineering (ICBME*)* (pp. 278-283). IEEE.
63. Subasi A, Gursoy MI (2010) EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Syst Appl 37(12):8659–8666
64. Suguna N, Thanushkodi K (2010) An improved k-nearest neighbor classification using genetic algorithm. Int J Comp Sci 7(2):18–21
65. Surya AA, Varghese SM (2016) Automatic speech recognition system for stuttering disabled persons. Int J Control Theory Appl 9(43):16–20
66. Świetlicka I, Kuniszyk-Jóźkowiak W, & Smołka E (2009). Artificial neural networks in the disabled speech analysis. In computer recognition systems *3* (pp. 347–354). Springer, Berlin, Heidelberg
67. Szczurowska I, Kuniszyk-Jóźkowiak W, Smołka E (2014) The application of Kohonen and multilayer perceptron networks in the speech nonfluency analysis. Arch Acoust 31(4 (S)):205–210
68. Tan TS, Ariff AK, Ting CM, Salleh SH (2007) Application of Malay speech technology in Malay speech therapy assistance tools. In 2007 International Conference on Intelligent and Advanced Systems (pp. 330-334). IEEE.
69. UCLASS DATABASE, URL:https://www.uclass.psychol.ucl.ac.uk/ [last access date: 01/01/2021]
70. Wahyuni ES (2017) Arabic speech recognition using MFCC feature extraction and ANN classification. In 2017 2nd international conferences on information technology, information systems and electrical engineering (ICITISEE) (pp. 22-25). IEEE.
71. Wiśniewski M, Kuniszyk-Jóźkowiak W, Smołka E, Suszyński W (2007) Automatic detection of prolonged fricative phonemes with the hidden Markov models approach. J Med Inform Technol:11
72. Wiśniewski, M., Kuniszyk-Jóźkowiak, W., Smołka, E., & Suszyński, W. (2007). Automatic detection of disorders in a continuous speech with the hidden Markov models approach. In computer recognition systems 2 (pp. 445–453). Springer, Berlin, Heidelberg
73. Xie L, Liu ZQ (2006) A comparative study of audio features for audio-to-visual conversion in mpeg-4 compliant facial animation. In 2006 international conference on machine Learni ng and cybernetics (pp. 4359-4364). IEEE.
74. Yairi E (2007) Subtyping stuttering I: A review. J Fluen Disord 32(3):165–196
75. Yuhas BP, Goldstein MH, Sejnowski TJ, Jenkins RE (1990) Neural network models of sensory integration for improved vowel recognition. Proc IEEE 78(10):1658–1668
76. Zhang JM, Harman M, Ma L, Liu Y (2020) Machine learning testing: survey, landscapes and horizons. IEEE Trans Softw Eng