International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

Speech Classification using Logical ART Deep Mechanism of Machine Learning

Pooja Nayak S, S G Hiremath, Arun Biradar

Abstract*: Apart from this there are many domains including medical, voice synthesis, hate speech classification and other custom applications where classification of speech plays an important role. The conventional techniques of speech processing and classification works on a small data set also provide lower accuracy of the classification. This paper introduces a learning model using neural network (NN) for the large dataset machine training and classification using critical feature analysis for the pattern of speech spectrogram and waveforms. The performance evaluation of the proposed training model for the speech classification is validated on a single CPU and found to achieve (12-82) % of accuracy in just 5-epochs and also continuously decreases the loss at successive iteration of the epochs. This method provides learning model framework for the speech processing and classification for a very large dataset.

Keywords*: Speech, Classification, Machine learning, Large speech data.

I. INTRODUCTION

The second generation of the artificial intelligence after machine to machine communication is human to machine communication (H2MC) [1]. There are many ways by which a human communicates with another human such as either by a sign language or through speech-based communication. Traditionally, various approaches are being researched to establish a system of H2MC [2] and out of many available methods a speech-based synchronization processes to establish a H2MC has various advantages as well as challenges. The typical essential property of good design of H2MC must consist of the parameters like a) recognition, b) Interaction, c) Affinity, d) Maneuverability, as shown in the fig.1.

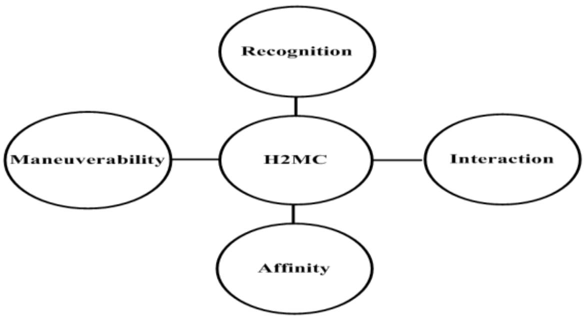

Fig.1 Essential design characteristics of H2MC

Revised Manuscript Received on December 30, 2019.

***** Correspondence Author

Pooja Nayak S*, Research Scholar, EWIT, Bangalore, India

S G Hiremath, Professor and Head, Department of ECE, EWIT, Bangalore, India

Arun Birader, Professor, CMR University, Bangalore, India

Form the fig.1 the component recognition can be described as an attribute which refers to identification of conception of perceptibility in the context of H2M interface. Interaction entity refers to a prime attribute of corresponding visualization of perceptibility interface from H2M. The backbone of maneuverability refers to the operational contents H2M interface object. Finally Affinity signifies the better understanding of H2M interface [3].

There several applications of H2M in existence considering the speech synthesis and analysis such as language recognition from speech based systems, quality assessment of speech signal, speech recognition from embedding attributes etc. The exploratory analysis of research trend in speech signal classification paradigm shows that various approaches are modeled considering reference modeling of neural networking (NN) features and attributes and it is highly adopted into many studies as compared to the learning based models of SVM, k-means logistic regression , random forest etc. But this NN model training functional design considered very limited dataset of speech signal attributes which is not sufficient to yield better classification and prediction accuracy as the validation model justifies this. Thereby it is realised that the processing of conventional more or less all machine learning approaches require large dataset extracted labelled features for efficient classification and to get better accuracy but here also the constraint is handling large number of feature and class attributes through limited memory size of conventional computing units. This scenario creates a bottleneck condition in realising the speech signal features and confine from better classification procedure. The formulated study addresses this problem by introducing a robust and stable learning model which is designed taking the reference of deep neural network for the purpose of efficient classification of speech signal attributes. The modeling introduces a novel add-on feature to store large dataset in an object which do not poses higher dependency over primary memory. Finally the promising outcome of the formulated system is illustrated to justify the modeling. Section II describes related work as a review of literature, section III describes Implementation strategy with data processing and learning, section IV describes results and analysis followed by section V as conclusion.

II. RELATED WORK

The authors (C. Yu *et al.*,2016) participated into one of the prestigious challenges held in Dallas for the language recognition for the speech system. Their model includes various modules using Support vector machine with linear discriminant (SVM-LD) and deep learning (DL) for the detection of the language for direct and indirect detection of the language from the unlabelled dataset. The method for indirect detection outperforms due to variant fusion and improvises the performance as compared to the standard cosine distance scoring (CDS) system [1]. The authors R. (Wang *et al, 2018),* studies the quality measurement of the speech signal by proposing a modification on the machine learning model using a no-reference technique on a large degrade speech signal database by developing a feature descriptor for high dimension waveform. The correlation among the feature descriptor and the average visual score is computed using a special type of boosting decision tree based on the gradients. They compared their results with respect to the P.563 method with the performance metric like Pearson coefficient, root mean and absolute mean errors and found it promising comparatively [2]. The exploratory and learning efforts by the authors (**Dixit, U. S. Shanthamallu, A. Spanias, V. Berisha and M. Banavar, 2018**) as a software to apply machine learning for speech signal embedded into HTML5 and extract the features like coefficients and frequencies out of it. The objectives are to perform extraction of features and clustering by speech processing using K means learning algorithm for the recognition task [3].The context where exactly the speech is recorded plays a very crucial role for speech processing and analytics, therefore a new vector namely feature-vector is introduced in the domain for the purpose of classification. The authros (**C. Papayiannis, C. Evers and P. A. Naylor, 2017)** performs the speech recognition task in the various distinguished environment to understand the effect and robustness of the feature selection and machine learning models and found experimentally that the classification accuracy is highly appreciable by reducing the world error rate(WER) very significantly due to appropriate feature selection[4].

The problem of mask estimation is used for modeling the speech signal classification. In this process a machine learning approach is used to map a relationship between the features of the speech signal and the mask. The issues of the noise are handled by means of the computation of the time and frequency domain of the background noises to improvises the quality of the signal. The conventional methods to classify the speech minimizes the induced error from the training portion of the dataset. In the work of (**X. Li, X. Wu and J. Chen, 2019**) proposes a loss-function which has ability to know the changes in the spectrum of the speech signal that ensures the classification performance to gain the higher SNR value [5].There is a mechanical device used in the dentistry named articulator and produces a different kind of noise which can be used to simulate the problem of the laryngectomy. The authors (J. A. Gonzalez *et al 2017)*in their work simulate the articulator, magnetic articulography with lips and tongue and train the system to training the estimated speech acoustics using direct synthesis. They explain the different learning algorithms including Gaussian mixture, Deep Learning and recurrent network and validate that a recurrent network performs well. Speech generated out of this method is found to be closely matching to the natural level of quality. This is a promising application by the speech synthesis useful to the people whose larynx is removed for the natural communication process [6].In the vision of human and machine interaction the role of automatic recognition of speech (ARS) is very critical and important. The authors (**M. S. Elmahdy and A. A. Morsy, 2017)** propose a sub-vocal recognition system using deep learning approach suing surface electromyogram and a closely placed microphone. The system is tested with a small corpus of words and works at 9.44-word error rate [7].There are two methods of evaluating sound, 1) instrumental measure and 2) perceptual measure, out of which perceptual measure is more powerful but it is very much time exhaustive process as well as not very cost effective, therefore, measuring these is an open research challenge. The authors, **(S. Safavi, A. Pearce, W. Wang and M. Plumbley, 2018)** have developed a method to perform prediction of sound from the signals related to the perceptions from a leaning model. They took N set of human listeners from the varied audio sources then trained the machine with the variety of features such as time for echo, echo ratio and proves that the machine learning models yield an accurate prediction score [8].

To meet the customer interaction and customer satisfaction, the speech emotion recognition is an open research problem in the domain of speech signal processing. The use of learning models is very much relevant and learning methods like SVM is widely used due their function characteristic of non-linearity to map the data to a very high dimension but at the same time it is not effective to handle many properties of dataset of speech emotion. The **authors (C. Zha, P. Yang, X. Zhang and L. Zhao,2016)** handles this issue of uni-functionality by introducing multiple kernels and tested on the Aibo dataset which exhibits better result as compared to the SVM [9].

In the continuation of handling the speech conversion for a patient whose articulators is remove, the authors **(S. Fu, P. Li, Y. Lai, C. Yang, L. Hsieh and Y. Tsao, 2017),** proposes a leaning model using dictionary and matrix factorization. It provides better result as compared to the tradition one but validated only on the small training dataset [10]. In the digital era there are many incidents where people record a speech which contains harmful or hate emotion which cause some disturbance in the civil society. These contents are shared through the social media platforms, therefore there is a need of analysis the content emotion in a speech and automatically stop the delivery or sharing. This research is at very nascent stage and lacks the standard dataset. The authors **(I. Alfina, R. Mulia, M. I. Fanany and Y. Ekanata, 2017)** contributed a step towards building such dataset with features including {General Speech, Religious hate speech, Racism Speech…etc} and analysed the scope machine learning usability for the classification with the popular method of SVM, Logistic regression and random forest etc. With feature set of n and character gram along with the negative semantic and validated with F-measure and found better performance with random forest decision learning algorithm [11].

One of the major challenges in the building leaning model for speech-based processing is the availability of the standard data set as well as it should in the large quantity with proper levelling. The more challenges exist in the medical sector where getting large amount of levelled data is difficult because of the privacy issue it is not accessible, therefore the method of speech-based learning in medical application shall support to provide higher accuracy with small dataset only.

The authors (**Y. Jiao, M. Tu, V. Berisha and J. Liss, 2018**) proses an adversarial approach of training for clinical data of dysarthric and archives almost 10% better result [12].

There is an essentiality of many potential applications where human and machine interaction is required. The current research trend has realised that proper analysis of speech signals can facilitate users to have a better communication experience in the context of any types of spoken dialogue systems. In the study of **(Z. Wang and I. Tashev, 2017),** a multi-layered deep neural network is introduced by exploiting their significant utility factors in the context of emotion detection from speech dialogue models. The computational modeling is mechanized based on an effective training-classification paradigm. The experimental analysis performed clearly shows that the system can effectively recognize and classify the speech signal prime entities as compared to existing baselines [13]. There exist several applications which process lower operational features to synthesize speech signals. But proper analysis of speech signal to recognize critical event till now a computationally challenging task due to higher-cost of signal annotations. There is a scope that implementation of learning model based approaches can assist to resolve this issue to an higher extent. Thereby **(H. Dubey, A. Sangwan and J. H. L. Hansen, 2018)** has introduced a learning based paradigm to properly segment the features extracted from the speech signal and also significantly solved the distortion problem. The outcome exhibited that the system can attain superiority as compared to existing approached of GMM [14]. The enhancement of speech signals spectral quality can be realised through exploiting the prior knowledge exploration which is a common feature in traditional learning models. However, the requirements of balancing the complexity and memory consumpsion trade-of are at bottle-neck stage. The study by **(R. Rehr and T. Gerkmann, 2018**), presented a theoretical and experimental study to evaluate the speech signal spectral estimation with a predictive design approach. The outcome shows that proper design of predictive approach not only solve the complexity problem but also enhances signal quality with effective estimators and outperforms Gaussian prior models [15]. A similar direction of research approach also attempted in the study of **(Y. Liao and Y. Wang, 2018)** towards exploring the potential aspects of learning paradigms and their key-feature to enhance the spectral quality of speech signals. For this purpose, the study introduces a novel computational design of multi-structured learning strategy to enhance the spectral quality of a speech signal which is required in most of the potential speech recognition and processing systems. The promising outcome exhibits its superiority over other existing techniques [16].The philosophical advent of Blizzard Machine Learning Challenge (BMLC) created a platform where participants can provide several contributions towards speech signal processing and that can bring assured outcome in the context of proper synthesis of speech signals. The study of **(Y. Hu, L. Liu, C. Ding, Z. Ling and L. Dai, 2017)** investigated the aspects of different waveform designs in terms of a novel feature extraction modeling. This has assisted the system to determine the predictive classification of potential waveforms in practice. The investigational outcome provides better insight into operational features which shows the effectiveness of the formulated concept [17]. The effective synthesis of speech signal attributes can assist in early detection of several disease conational regulations. The study of **(A. Agarwal, S. Chandrayan and S. S. Sahu , 2018)** has worked in this direction of research analysis and investigated whether a computational model can be introduced to combat the Parkinson's disease (PD) disease by detecting it as early as possible. The system modeling is done in a way where it utilizes a specific training dataset comprising the labeled features associated with Parkinson diseased patients. The quantified outcome shows that it obtains an accuracy of approximately 81.55% and that is superior as well as reliable than other learning models [18].

It is essential to perform proper speech signal quality improvement by means of monitoring the performance metric in the context of hearing aid (HA). Thereby the reference signal should be modelled with properly aligned time-frequency vectors which are well-synchronized to generate the enhanced quality output of HA. The formulated model of **(H. Salehi, D. Suelzle, P. Folkeard and V. Parsa ,2016)** explored different potential learning features of linear prediction and can enhance the outcome of HA in different test conditions which makes it more robust as compared to the conventional system [19]. A similar sort of study also performed by **(R. Gupta, T. Chaspari, J. Kim, N. Kumar, D. Bone and S. Narayanan, 2016)** to enhance the speech signal attributes through learning aided proper analysis. This study basically performed an investigational study to explore the possibility of applicability of learning models and signal processing to significantly improve the quality of speech signal synthesis. It has also revived various ML based knowledge and data driven models and their practicability in realistic speech pathology [20]. It can be seen that the advancement of various multimedia applications can be realized by exploring their computational features and how they interact with the user modules. To enhance the level of interaction text-to-speech applications poses higher utility. The study of **(M. A. Kutlugün and Y. Şirin, 2018)** explored the learning paradigms of machines to perform efficient way of speech signal synthesis through cost-effective classification model. The procedure basically synthesizes textual documents on the basis of a classification and produces respective sound formats. The accuracy of text-to-speech learning in realized though a machine learning model. The outcome shows it attain better classification accuracy and ensure promising aspect in document classification problem [21].A classification problem of human speech analysis and the possibility to extract efficient speech attributes towards providing training to an auditory system is explored in the study of **(H. Phan, L. Hertel, M. Maass, R. Mazur and A. Mertins, 2016).** A mapping of non-speech signal attributes with speech signal attributes is made based on a robust classification and prediction modeling. The algorithm is modeled in a way where approximation of dataset pertaining to different speech patterns is explored. The experimental outcome on the basis of event classification shows that it can challenge the state-of-the-works [22]. The advancement in the field of computational intelligence systems has given tremendous improvement in speech recognition systems with an aid of learning model architectural paradigms.

Thereby, hierarchical clustering algorithms gained lots of attention from research community towards improving the speech recognition paradigm. The novel clustering design is introduced in the study of **(A. Sarkar, S. Dasgupta, S. K. Naskar and S. Bandyopadhyay, 2018)** with sophisticated variants. The segmentation of speech samples is realised in the environment of Libri speech corpus. The experimental analysis eventually attained comparable outcome in the end [23]. The study of **(P. D. Çelen and F. Hardalaç, 2017**) also focused on faster information processing in the context of speech signal synthesis. The segmentation of the speech signal samples is obtained through speech-to-text program which is designed based on the learning aided classifier. The accuracy of recognition is obtained approximately 75% [24]. Another automated speech recognition system is modeled taking the reference mechanisms of signal processing and acoustic models by **(B. Wu et al., 2017).** The study introduces 2-fold techniques for speech quality improvement and multi-conditional training requirements respectively finally it undergoes thorough a joint optimization approach which leverages the objective measures under the simulated datasets. The study also incorporated as neural network based learning model which attain better classification accuracy with lower error rate [25].

III. IMPLEMENTATION

A. Speech Data Pre-processing and Preparation

The preliminary stage of the design and analysis considers the dataset to be loaded and after that it performs data synthesis in the form of pre-processing of different speech signal attributes to make it more significant and suitable as input to the learning model as compared to the raw signal attributes.

B. Speech Signal Memory constraint object

The growing size of the speech signal samples require a Meacham of holding a large dataset beyond the capacity of the memory constraints. The system defines a speech object $(S_o) \leftarrow (N,V)$, as an ordered pair of name(N) and the value(V) with 'm' classes of categories. The model takes 'k' categories such that $k \subset m$. The labels of the 'k' categories is matched with the labels in S_o, for each matching labels (l) mark '1' for true. The mask (Ms) is computed by comparing 'm' pseudo random values with a threshold introduced 'Fr a fraction as binary mask using equ(1)

$$Ms \leftarrow [\ \varphi\ (m) < Fr) \ \ldots\ldots\ (1)$$

Where, φ() generates pseudo random values.

C. Noise Labelling

The existence of the noise is very obvious especially the background noise. The $k \notin Bn($ a background noise in the set of 'm is assigned to a set of [0,1]. The logical conjunction is performed among $Bn \cap Ms \leftarrow Bn \wedge Ms$ as shown in the table 1.

Table 1 logical conjunction between Bn and Ms

Bn	Ms	Is= Bn ^ Ms
1	1	1
1	0	0
.	.	.
.	.	.
0	1	0

The intersected data points in the 'm' is marked as a special label as insignificant signal (Is). The interests of speech class final object (Sfo) = $So \subset (m \vee Is)$ the operation of the m V ls is shown in the table 2.

Table 2 Intersection between m and Is

m	Is	Sfo= m V ls
1	1	1
1	0	1
.	.	.
.	.	.
0	1	1

This operation performs signal dimension reduction with the number of classes (p) <m. The fig.2 illustrates the relaxation of the m class/category reduction to p class.

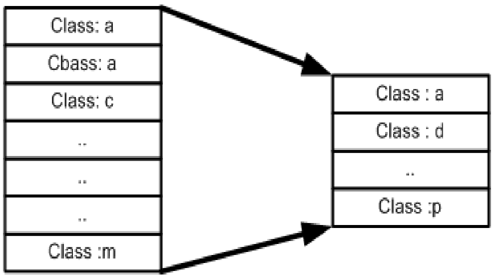

Fig.2 Relaxation of 'm' class reduction to 'p' class

D. Dataset preparation for machine Learning

The entire reduces data with class 'p' is divided into three portions such that the data portions (Dp) = {Training Data (Trd), Validation data (Vd), Testing Data (Td)}. To divide and split the class 'p' data into three different units Trd, Vd and Td, a computational procedure **$Func_1$([Trd, Vd, Td])** is designed. It basically processes three prime attributes With a set of computational steps and yield the outcome. Initially the numerical process invokes an associative functional module **$Func_{read}$ (Vfile)** to read the object that carries the list of validation files (**Vfile).** Considering the reference file Vfile, the updated version of Vd is obtained by splitting the list structure of Vfile on the occurrence of delimiters such as spaces, commas and consecutive dash tuples. Similarly, the similar process of line of execution flow is repeated to attain the numerical values of Td**.**

A mathematical computation is further initializing few core attributes to perform the validation schema on [Sfo → p] to verify that which file from the created object of Sfo referenced with p goes to Vd and which flies to Td.

After successful division and accurate distribution of Trd , Vd and Td , the spectrometry analysis of different features attributes from **Trd, Vd, Td** considers key components of segment interval (Si) , frame interval(fi) , hop interval (hi) and number of frequency bands (nf). Finally a core design module of $\theta_1(X)$ get activated to add background noisy data to Sfo → p and further computes the normalization metric in the form of probability density factor. The computation further also performs training class frequency dissemination in the form of normalization process prior performing the validation of class label attributes from the process of validation frequency data. The process further defines a learning model based on the conceptual line of philosophy of neural networks and start training the model with Trd.x and Trd.y**.**

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

The neural network based learning architecture is designed on referring the convolution neural network. Extract different speech based objects from capturing devices and classify the spectrogram component of that speech signal object. A matrix of probabilistic prediction (Mp) is created which stores the numerical outcome of the classification prediction. Finally another computational module of $\theta_2(X)$ is activated which ensures the classification prediction accuracy on the basis of a threshold operation and validation model referring the buffer which hold the probabilistic computed values at each and every round of classification procedure. The final process obtains the classification prediction probability of the predicted class variable based on thresholding. The following fig.3 shows overview of the core-execution flow of the formulated system.

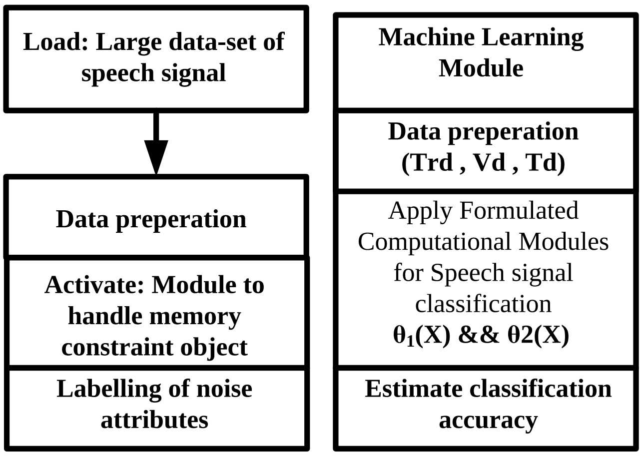

Fig.3 Block oriented representation of core-executional module

IV. RESULTS AND ANALYSIS

A. Neural network training and validation

The effectiveness of the training program is measured with the help of two different performance matrix: 1) %accuracy 2) Loss Vs each iteration. The training simulation parameter is shown in the table.3 below:

Table 3 Training and validation parameters

Serial number	Parameter	Values
1	Epoch	5
2	Iteration per epoch	219
3	Max iteration	5475
4	Frequency	219
5	Hardware resource	single CPU
6.	Learning rate schedule	Piece wise
7.	Learning rate	0.003

The fig.4 shows the outcome obtained after evaluating the core model to evaluate the training process. The training process is modeled on the basis of class frequency updated of m number of separated labeled classes. The number of classes considered stating from class-1 to class-m.

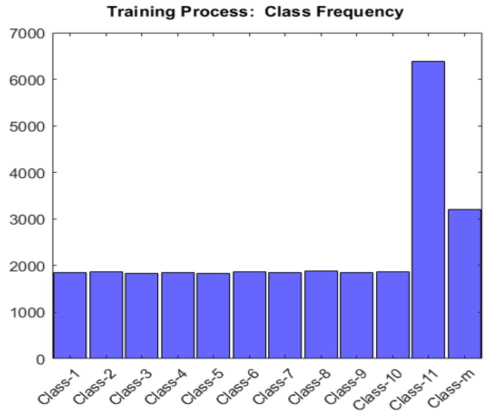

Fig.4 Training Process: Class Frequency

A closer observation of the above fig. 5 clearly shows that in the case of class-11 the corresponding frequency is much bigger that is 6500 and in class-m comparatively lowers value if frequency can be obtained. On the other hand, class-1 to class-10 poses similar sort of frequency values.

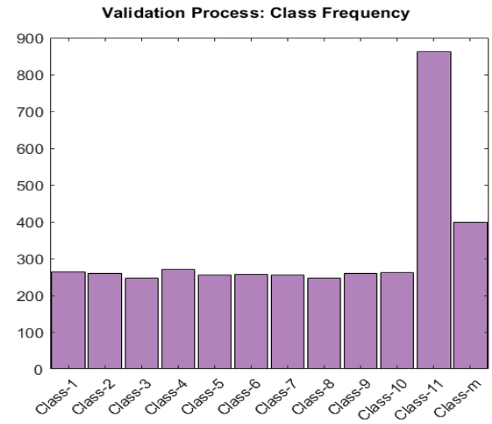

Fig.5 Validation process Class Frequency

After training process the synthesized speech signal extracted feature attributes are taken for the validation process on the basis of class frequency modeling. The outcome clearly shows that the class frequency much higher in this case also in the type of class-11.

B. Performance evaluation of % accuracy

The computational observation of performance accuracy with the parameter set in the table 3 is tabulated in table 4.

Table 4 Observation table of %accuracy Vs iteration for 5 epochs

	Epoch-1	Epoch-2	Epoch-3	Epoch-4	Epoch-5
%Accuracy	12	70	76	79	82
Iteration	0	220	430	500	890

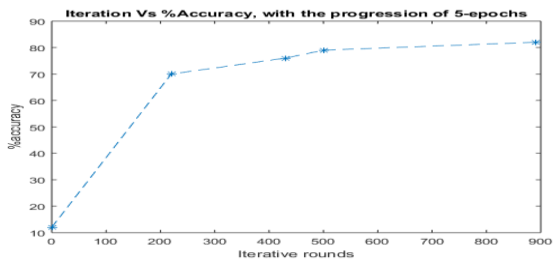

Fig.6 % accuracy Vs iteration in a single CPU for 5-epochs

C. Performance evaluation of loss

The computational observation of performance loss with the parameter set in the table 4 is tabulated in table 5.

Table 5 Observation table of loss Vs iteration for 5 epochs

	Epoch-1	Epoch-2	Epoch-3	Epoch-4	Epoch-5
%loss	2.8	0.9	0.6	0.4	0.3
Iteration	0	220	430	500	890

The fig 7 shows that the %accuracy increases along with the increasing number of epochs in the incremental iteration that validates the model that the accuracy will increase for the defined number of targeted epoch and iteration. The % accuracy is 10% in the 1st epoch, whereas in the 5th epoch at 890 iteration is found as 82%.

The fig 7 shows the graph of loss computation Vs iterative round in 5 epochs

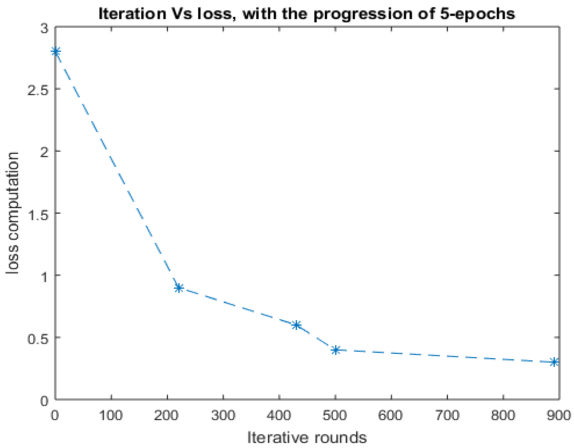

Fig.7 Loss computation Vs iteration with single CPU at 5 epoch

The trend of loss vs iteration in every iteration Is decremental where in the 1st epoch the loss value is 2.8 whereas the loss is 0.9, 06, 04, 0.3 in the successive epoch at 2, 3, 4, 5 in 220, 430, 500, 890 iterations respectively. This shows that with the incremental process of training the loss is reducing; therefore the model exhibits stable performance.

V. CONCLUSION

The aggressive use of digital platforms for sharing and communicating multimedia with voice generates huge dataset. In a speech therapy vocal control mechanism etc requires a very effective in terms of higher accuracy and lower time for the speech classification. The trend of data generation and storage and availability of high-performance computing provides a suitable eco-system to innovate learning mechanism-based speech processing and classification models using large dataset datastore mechanism to hold data which is possible to be fit in memory. The neural network and variants including single perceptron neutron to multi-layer perceptron and further a series of neural networks placed in a layer along with adaptive resonance theory neural network (ART-NN) is modified by hybridizing it with the network of neural networks as a deep training is utilized in this paper. A very large dataset is taken for the feature extraction, training, testing and validation the model of training and classifies m-subset of samples of various classes. The model exhibits higher accuracy and lower loss with the successive number of iterations in each epoch of training. The model can be further improvised for the complex speech statements by developing feature descriptors based on the semantic of the language to meet the real time goal of the futuristic applications speech classification is considered as a core engine.

REFERENCES

1. Alan Cooper, Robert Reimann, The Essentials of Interaction Design, China: Electronics Industy, 2005.
2. Donald A.Norman, Emotional Design, China: Electronics Industy, 2006.
3. Donald ACite (Human-Machine Interface: Design Principles of Visual Information in HumanMachine Interface Design)
4. C. Yu et al., "UTD-CRSS system for the NIST 2015 language recognition i-vector machine learningchallenge," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016, pp. 5835-5839.doi: 10.1109/ICASSP.2016.7472796
5. R. Wang et al., "No-reference Speech Quality Assessment of SWB Signal Based on Machine Learning," 2018 15th International Symposium on Wireless Communication Systems (ISWCS), Lisbon, 2018, pp. 1-5.doi: 10.1109/ISWCS.2018.8491197
6. Dixit, U. S. Shanthamallu, A. Spanias, V. Berisha and M. Banavar, "Online Machine Learning Experiments in HTML5," 2018 IEEE Frontiers in Education Conference (FIE), San Jose, CA, USA, 2018, pp. 1-5.doi: 10.1109/FIE.2018.8659113
7. C. Papayiannis, C. Evers and P. A. Naylor, "Discriminative feature domains for reverberant acoustic environments," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017, pp. 756-760. doi: 10.1109/ICASSP.2017.7952257
8. X. Li, X. Wu and J. Chen, "A Spectral-change-aware Loss Function for DNN-based Speech Separation," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 6870-6874.doi: 10.1109/ICASSP.2019.8683850
9. J. A. Gonzalez et al., "Direct Speech Reconstruction From Articulatory Sensor Data by Machine Learning," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, pp. 2362-2374, Dec. 2017.doi: 10.1109/TASLP.2017.2757263
10. M. S. Elmahdy and A. A. Morsy, "Subvocal speech recognition via close-talk microphone and surface electromyogram using deep learning," 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), Prague, 2017, pp. 165-168. doi: 10.15439/2017F153
11. S.Safavi, A. Pearce, W. Wang and M. Plumbley, "Predicting the perceived level of reverberation using machine learning," 2018 52nd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2018, pp. 27-30.doi: 10.1109/ACSSC.2018.8645201
12. C. Zha, P. Yang, X. Zhang and L. Zhao, "Spontaneous Speech Emotion Recognition via Multiple Kernel Learning," 2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Macau, 2016, pp. 621-623. doi: 10.1109/ICMTMA.2016.152
13. S. Fu, P. Li, Y. Lai, C. Yang, L. Hsieh and Y. Tsao, "Joint Dictionary Learning-Based Non-Negative Matrix Factorization for Voice Conversion to Improve Speech Intelligibility After Oral Surgery," in IEEE Transactions on Biomedical Engineering, vol. 64, no. 11, pp. 2584-2594, Nov. 2017.doi: 10.1109/TBME.2016.2644258
14. I.Alfina, R. Mulia, M. I. Fanany and Y. Ekanata, "Hate speech detection in the Indonesian language: A dataset and preliminary study," 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Bali, 2017, pp. 233-238.doi:10.1109/ICACSIS.2017.8355039.
15. Y. Jiao, M. Tu, V. Berisha and J. Liss, "Simulating Dysarthric Speech for Training Data Augmentation in Clinical Speech Applications," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018, pp. 6009-6013.doi: 10.1109/ICASSP.2018.8462290
16. Z. Wang and I. Tashev, "Learning utterance-level representations for speech emotion and age/gender recognition using deep neural networks," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017, pp. 5150-5154.doi: 10.1109/ICASSP.2017.7953138

17. H. Dubey, A. Sangwan and J. H. L. Hansen, "Robust Feature Clustering for Unsupervised Speech Activity Detection," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018, pp. 2726-2730. doi: 10.1109/ICASSP.2018.8461652
18. R.Rehr and T. Gerkmann, "On the Importance of Super-Gaussian Speech Priors for Machine-Learning Based Speech Enhancement," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 2, pp. 357-366, Feb. 2018. doi: 10.1109/TASLP.2017.2778151
19. Y. Liao and Y. Wang, "Some Experiences on Applying Deep Learning to Speech Signal and Natural Language Processing," 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), Kosice, 2018, pp. 83-94.doi: 10.1109/DISA.2018.8490638
20. Y. Hu, L. Liu, C. Ding, Z. Ling and L. Dai, "The USTC system for blizzard machine learning challenge 2017-ES2," 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, 2017, pp. 650-656.doi: 10.1109/ASRU.2017.8268998
21. A. Agarwal, S. Chandrayan and S. S. Sahu, "Prediction of Parkinson's disease using speech signal with Extreme Learning Machine," 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, 2016, pp. 3776-3779. doi: 10.1109/ICEEOT.2016.7755419
22. H. Salehi, D. Suelzle, P. Folkeard and V. Parsa, "Learning-Based Reference-Free Speech Quality Measures for Hearing Aid Applications," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 12, pp. 2277-2288, Dec. 2018.doi: 10.1109/TASLP.2018.2860786
23. R. Gupta, T. Chaspari, J. Kim, N. Kumar, D. Bone and S. Narayanan, "Pathological speech processing: State-of-the-art, current challenges, and future directions," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016, pp. 6470-6474.doi: 10.1109/ICASSP.2016.7472923
24.]M. A. Kutlugün and Y. Şirin, "A novel approach improvement framework for text to speech synthesis," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4doi: 10.1109/SIU.2018.8404828
25. H. Phan, L. Hertel, M. Maass, R. Mazur and A. Mertins, "Learning Representations for Nonspeech Audio Events Through Their Similarities to Speech Patterns," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 4, pp. 807-822, April 2016.doi: 10.1109/TASLP.2016.2530401
26. A. Sarkar, S. Dasgupta, S. K. Naskar and S. Bandyopadhyay, "Says Who? Deep Learning Models for Joint Speech Recognition, Segmentation and Diarization," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018, pp. 5229-5233.doi: 10.1109/ICASSP.2018.8462375
27. P. D. Çelen and F. Hardalaç, "Speech record speed up with machine learning technics," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4.doi: 10.1109/SIU.2017.7960316
28. B. Wu et al., "An End-to-End Deep Learning Approach to Simultaneous Speech Dereverberation and Acoustic Modeling for Robust Speech Recognition," in IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 8, pp. 1289-1300, Dec. 2017.doi: 10.1109/JSTSP.2017.2756439

AUTHORS PROFILE

	**Pooja Nayak S**, She has completed her BE in computer science and engineering in the year 2009 , MTech in Computer networks engineering in the year 2011, worked as assistant professor in SSIT, Tumkur from 2012-14, assistant professor at EWIT, Bangalore from 2014 to 2017, currently research scholar under VTU.
	**Dr. S G Hiremath,** is having over 20 years of experience in the field of engineering education. He has been awarded Ph D from Anna University in the faculty of ECE. His interest in research areas of Instrumentation, Digital Signal Processing, Neural Networks, Control Engineering, Hardware Design, Electronics Circuit Design, Instrumentation Sensors, Process Controls, PLC s and SCADA, Modeling & Simulation, Neural Networks, Telecom and related Teaching and promotional areas. He has published more than 70 research publication in the international journals and conferences.
	**Dr. Arun Biradar** is an academician having more than 22 years of experience in teaching, development, administration and leadership. He completed B.E. in Computer science and Engineering from Gulbarga University and M.Tech in Software Engineering from Mysore University and Ph.D. in Computer Science Engineering from Swami Ramanand Teerth Marathwada University, Nanded. His research interests are Wireless Ad-hoc Networks, Computer Networks, Software Engineering, Genetic Algorithms, Machine Learning, IoT and Cloud Computing. He is a recognised guide under Visvesvaraya Technological University, Belagavi, and has many students pursuing their Ph.D. under his guidance. He is a member of various professional bodies and has been a part of the NAAC and NBA Advisory Committee to many Colleges. He is an active member of IEEE