United International Journal of Engineering and Sciences
(UIJES – A Peer-Reviewed Journal); ISSN:2582-5887 | Impact Factor:8.075(SJIF)
Volume 5 | Special Issue 1 | 2025 Edition
National Level Conference on "Advanced Trends in Engineering Science & Technology" – Organized by RKCE

Multi-nominal Approach on Conditional Gaussian Distribution for Speech Recognition in Machine Learning Algorithms

Anil Kumar Maddali[1], P Rajani Kumari[2], Ramakoteswara Rao S[3],

S Lakshmi Deepika[4]
Associate Professor
Department of Electronics & Communication Engineering,
DVR & Dr. HS MIC College of Technology,
Andhra Pradesh, India

Abstract :
Our design improvises on the custom model of the speech recognition for different audio dataset with (XLSX) that are utilized based on the dataset that governs the different features of frequency calculation parametric conditions that implements the given criteria. We improvise a decision tree machine learning approach with Multinomial feature on CGD (conditional Gaussian distribution) In order to highlight the feature labels must be categorized as MALE or FEMALE. The dataset feature with 3168 samples are modelled with proposed algorithm and represented the training accuracy with 1.5% improved, and tenting accuracy of more than 20 % difference from the existing DT classifier.

Keywords : DNN (Deep neural Net), SRT (speech recognition technology),
MNGD (Multinomial conditional Gaussian distribution), CGD (conditional Gaussian distribution)

I. INTRODUCTION

Speech recognition is the ability of a machine to perceive and comprehend human speech in order to respond appropriately to a statement or instruction. As a study object, the voice is used to enable the computer to instantly recognize and to interpret speaking language of humans using analysis of speeches and signals which tend to recognize different patterns.

Speaking Recognition Technology (SRT) is a technology that permits the computer to convert a vocal signal to the right words or commands by undergoing a procedure of classification and comprehension. Voice analysis with recognition is a multidisciplinary field that encompasses a wide range of technologies. It is particularly linked to acoustics, phonetics, linguistics, information theory, pattern recognition theory, and neurosciences.

The technology related to speech recognition and its verification factors plays an important role in improvising a solution analysis for different real-time applications related to the security, forensics, and authentication models to the development of Artificial Intelligence (AI) modeling [16]. In a world of automation where every operation is handled by sound or voice, speech recognition improves the real-time analytical necessity and importance for a variety of applications, from daily tasks to business operations. [17]. Amid the preparation stage, the trademark highlights of the speeches are either classified into male or female voices in databases, for each such scenario we tend to provide stage vise to improve the capabilities, we provide TEST and TRAIN framework [18]. Feature training of the dataset will be done using vector quantization and Feature testing of the dataset will be done using Viterbi algorithm. Home automation will be completely based on voice recognition system [19].

SRT is steadily establishing itself as a key advancement in computer processing technology, because of the fast growth of computing devices, as well as information technology, in recent years. The products used to create speech recognition technology have an impact on almost every aspect of society and human careers, including voice-activated telecommunications networks, communications network queries, financial, industrial, medical, and almost all other areas of civilization and everyday life. Several specialists feel that voice control was most significant scientific and technical advancements in the information technology industry between 2000 and 2010.

II. RELATEDWORK

The acknowledgment of a speaker has been noted by scientists for quite some time. Conventionally, tasks like this include methods like MFCC highlight extraction and HMM presentation. Nevertheless, professionals in the two regions have always been used to benefiting from one another. Performing DNNs in talks has inspired neural models in speaker acknowledgment [3], [4].

Scientists also know that using the understanding of one zone typically enhances the other. Speaker acknowledgement I-vectors have been employed in speech acknowledgment [5] to help with it, and speaker acknowledgment [6] has also been built using telephone rear ends. This combination of these two frameworks has suddenly gained a new lease of life. For example, speech and speaker joint derivation was presented in [8], and an LSTM-based does various tasks in [9] though interesting, the aforementioned studies aren't qualified as multi-task learning, as they are designed, planned, and executed autonomously.

Speech-handling enhancement enables better multi-tasking learning. From 2011 to the present, a new generation of repeated neural networks (RNNs) has rapidly grown to be the new cutting edge in voice recognition. Profound learning gives us two critical tidbits: First, supplemental profundity (a different level) divides the accomplished features, and second, worldwide profundity (intermittent connections) assembles dynamic evidence. Just because two assignments' structures resemble one other, may we employ a single model to play out both assignments

Undoubtedly, this "Perform Various Tasks Learning" has worked well to aid related errands [13]. It has been discovered that pooling low-level DNN layers improves language execution across the board [14]. Also, telephone and grapheme acknowledgement were considered as two interconnected endeavors [15]. While performing related chores, students will find that they have comparable components extraction and, thus, the minimal layers of DNNs may be shared. Regardless, this component sharing design has no impact on speech and speaker recognition.

The dual assigns are "conversely related," according to the hypothesis. Acknowledgement of speech needs highlights that include as much substance data as possible, whereas phonetic substance is ejected when speaking is acknowledged. These assignments do not allow sharing. Disappointingly, several endeavours are intertwined. For example, language designates proof and acknowledgement for speaking. Finding a method of accomplishing a variety of things simultaneously without sacrificing focus is enticing

III. EXISTING MODEL

3.1.HMM:

HMMs are a kind of predictive model used to forecast future occurrences. In the 1970s and 1980s, the Hidden Markov Model (HMM) analytic method was successfully utilized for the analysis of complex acoustic signals. computerized speech identification integrated telecom technology for the

identification of numerous users was also made possible by HMM throughout the 1980s and 1990s [9]. However, until the development of a more efficient, Rapid and precise voice identification system is refined, this approach is considered the most successful. The parameters of the HMM model are utilized to represent the voice signal's time-dependent properties.

This model contains dual stochastic processes that are linked to each other and which have the capability of defining the the signal's analytical properties. A finite-state Markov chain and a random observation vector are both studied under the Markov chain process. Its hidden properties are contingent on the signal qualities, which are exposed. A technique is presented to describe time-dependent signal, such as speech properties, using a arbitrary procedure [8]. Markov chain signal may vary with time, as is seen in the graph. This observation was described as the discrete HMM in a specific state (j) and by a group of probabilities (k = 1, 2, M).

The term discrete HMM is a helpful metaphor to define the concept, and is also known as an M discrete countable observation. If a continual random variable X has a value that is observed by a probability density function bj (X), then the state that the variable is in is observed by a continuous HMM. In the specification of the value of the j (X) parameter, continuous HMM has been used, using the Baum-Welch method for parameter estimation [7].

3.2 ANN

The machine learning algorithm is similar to how organic nervous systems process information in the case of an artificial neural networks (based on Artificial Neural Networks), where a huge quantity of basic processing facilities is linked together to form a smart information extraction system [1]. Using speech signal processing, we can achieve high parallelism and rapid judgement while also providing fault tolerance in the system under test. An auditory neural network model is often split into two types: first type introduces the standard HMM, DP combination of hybrid networks, and, the another, the setup of an auditory neural network model. [10]

Among the most frequently used and most promising speech recognition models are the singular perception model, the multi-layer perceiving model, the Kohonen self-organizing feature map model, the radial basis function neural network, the predictive linear regression, and other neural network models. It is necessary to use additional methods like delay neural networks and recurrent neural networks to be able to guarantee that the neural network correctly provides the voice input's characteristics across duration. When it comes to speech identification applications, below are the main components of artificial neural network technology.

The reduction of the modelling unit, which is frequently done during the phoneme modelling process, is a common practice to increase system-wide detection rate by increasing the recognition rates of phonemes. to be able to increase system-wide detection rate by increasing the recognition rates of phonemes, it is common practice to reduce the modelling unit.

It is necessary to thoroughly investigate the sound model, the sensory model, the neural function system, and the integration of reference info to be able to lessen the effect of differences in speech that are important than changes in the spoken signals. MFCC and linear predictive coding (LPC) with different other features have proven better results [20].

In order to comprehend the studies and enhance system quality, a hybrid network model (HMM + NN), as well as the use of a range of information sources are used for speech identification. Artificial neural network technology is used to achieve speech recognition, as depicted in fig. 2.

This covers both e-learning procedure and the speech identification process, among other things. This method, which includes the use of self-learning neural networks and the development of an array of connection weights and biases, is based on previously recorded speech signals and involves the use of well-known speech signals. [12]

A voice signal that is utilized as a network input is subjected to a speech recognition procedure that checks to determine whether the signal can be recognized. The outcomes of recognition are produced via the use of a network of relationships. It is essential to achieve a balance between speech characteristic parameters and neural network learning for these two processes to be successful.

Artificial neural networks (ANNs) have experienced a substantial rise in usage in the field of voice recognition over the recent years, according to recent research. The application of convolutional neural networks in speaker recognition transforming could also be classified as follows: voice recognition, speech synthesis (including speech recognition software). First and foremost, it is necessary to enhance the achievement of artificial neural networks [21]-[23].

An artificial neural network has been used to develop a technique for integrating fusion systems, which is second step in the process (ANN). Investigate, in the third step, using recently discovered or well-accepted mathematical models that contribute to the distinctive character of the neural network and that may be employed to the area of speech signal processing [24]-[25]. Using artificial neural networks in speech recognition has climbed to the top of the list of developing technologies, and it is expected to continue to grow. Given that artificial neural network technology has been utilized successfully to the issues with pattern classification and has been depicted to consume a significant amount of energy, with in the next ten years, we can anticipate the release of products using artificial neural networks for speech identification. And that individuals will modify their speech patterns to accommodate voice recognition systems based on artificial neural networks.

IV. PROPOSED MODEL

For every conceptual representation, we first begin the pretreatment step with windowing approaches, which provide improved frames. Since we are aware of this, we designed the previous design and its execution study, that has a impact on the present model and especially leads to analysis of the various blocks including the transformations, Mel-Spectrum, and eventually optimizing and clustering.

In order to enhance the capacity of the optimization and clustering modules, we have now proposed specific changes to the algorithms used in the present model. GPVQ and Modified Weight Based Euclidian Distance Calculation are two such planned optimization techniques used in this work. Gradient Predictive Vector Quantization would modify the present architecture in this case depend on the weights produced from the distances computed for each sample of the dataset and the user speech sample to be observed.

In order to examine the weights measured and predicted for each iteration and improve sample categorization, these weights must be predictive in nature. This will impact the design attributes. Below are more thorough explanations of each of these created blocks.

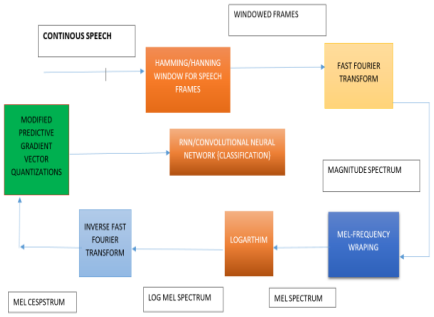

Fig.1. Design Block Diagram

4.1 MFCC TO EXTRACT FEATURES

MFCC stands for Mel Frequency Cepstral Coefficients. The feature extraction technique involves windowing the signal, applying the Discrete Fourier Transform, taking the log of the magnitude, and after that using a Mel scale the frequencies are warped, followed by taking the inverse Discrete Cosine Transform.

4.2 FAST FOURIER TRANSFORM (FFT)

$$X_w(k) = \sum_{n=0}^{N-1} x_w(n)\, e^{-\frac{j2\pi kn}{N}}, \quad k = 0,1,2 \ldots \ldots N-1$$

Here, $x_w(n)$ processed speech frame before and gives the length of the DFT. FFT provides fast calculation for performing the DFT to incorporate in domain modifications noticed in frequency.

4.3 MEL FREQUENCY WARPING

To convert the signals from the frequency domain to the Mell range, the Mel scale is applied to the frequency-based power spectrum range to change the range from which the signal is extracted to the Mell range. To achieve this goal, a series of filter processes are performed on either the spoken signal's power spectra, each using a different set of frequencies and depending on sound-related band channel capabilities and focal frequencies.

A signal spectrum, which is similar to the sensitivity of the human ear, is known as a filter when it is translated into a representation such as the Mel scale. Compared to shorter wavelengths, higher wavelengths are far more sensitive to it. (below 1 kHz). As said, this is on the MEL scale, and that is why the filters have been made to mimic this (more filters in the low frequency range than the high frequency range). When a collection of these filters is used together, they are called a medium-scale filter bank. M represents the microphone, while f represents the frequency of the signal.

4.4 CEPSTRUM

The inverse transformation of the Fourier transform of an anticipated spectrum of a signal is referred to as a cepstrum. Some features of the cepstral highlighter's tonal qualities are more apparent with a basic spectrogram of the spoken input. In the cepstral space, a convolution of two signals expands to their complex cepstra, making it easier to go straight to the design details. A feature of special importance in this case is the fact that the observed signal, which is made up of the input signal and the vocal tract's impulse response, is a convolution of the two signals.

To measure and note the changes in the signal spectrum over time, the residual advance is used. The characteristics of the genuine cepstrum can be used to describe the opposite DFT, such as the Discrete Cosine Transform (DCT). The DCT is capable of communicating an inclusive succession of restricted information that focuses on whole-frequency cosine capabilities that are currently being refined. DFT employs complex exponential expressions (cosines and sines). To obtain the log meter

coefficient, the log meter scale is DCT. In this study, we get the speech components in the form of MFCCs. These coefficients are the component vectors that are used in this report to depict an individual's speech that is kept in databases and specified user scenarios.

4.5 VECTOR QUANTIZATION

The technique of reducing a sizable vector space in computer science to a limited number of areas in that space is known as vector quantization (VQ). Every area is referred to as a cluster, and it can be portrayed by a code word that represents its center/centroid. A codebook is a collection of all of the terms that are used in coding.

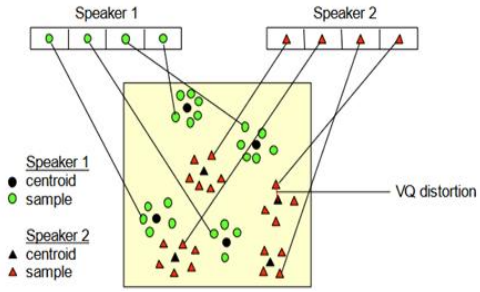

Fig 2. Conceptual diagram that illustrates the process of recognition in action.

In the illustration, speaker1 and speaker 2 and 2 components of the acoustic space vectors can be seen, as well as their interactions. Speaking of acoustic (including) vectors, the circles and triangles represent those of Speaker 1, respectively, and the circles and triangles represent those of Speaker 2. By clustering each speaker's training attribute vectors during the preparation step, a speaker's explicit codebook is generated for each speaker during the execution stage.

When the following code phrases (centroids) are spoken, dark circles (Speaker 1) and dark triangles (Speaker 2) show up on the screen, respectively. In computer science, this separation between a component vector and its nearest modified code word (centroid), as well as the ability to recognize better nodes in areas where the datum is particularly important, is known as a GP-VQ contortion.

To begin the angle foresight gradient predictive of the code word, we suggest a real-time study of the separation anticipated and observed in the Modified Euclidian separations to serve as a starting point. When it comes to contortion, the degree of similarity between the testing information (testing voice test) and the present preparation information is generally what you're looking for (prepared voice test). A smaller number of contortions indicates that the test is a more accurate match for the preparatory test. As part of the testing procedure, the information voice test is subjected to vector quantization using the previously produced codebooks, and the G-VQ twisting is performed. It is possible to identify the speaker who refers to the codebook with the lowest amount of contortion.

V. MACHINE LEARNING APPROACHES

As a result of its success, ASR has been used to develop numerous applications in a variety of different languages. For the purpose of developing an efficient system, it follows a standard approach that includes steps such as data gathering, speech segmentation, feature extraction, and model development. During the testing process, these models are utilized as benchmarks. A number of decision-making techniques are employed in the recognition phase to identify spoken words using characteristics collected from testing data. As a feature extraction approach, we made use of the MFCC. Machine learning paradigms such as HMM, ANN, and SVM are employed in this application

for pattern categorization. A more in-depth discussion of the general technique is provided in the following sections.

Design Procedure*:*

1. Data Acquisition
2. Feature Extraction
3. Speech feature segments
4. Machine learning Algorithms
5. Plotting Tabulating the outcome of the design feature.
6. Performance characteristics

5.1 DATA ACQUISITION

Our design implements the database modelling of 21 class labels each featuring the different classification criteria implementing the design metric for the data chosen. The implementation of the dataset considered from the design classes such as ["mean-frequency, standard deviation, median, Q25, Q75, IQR, skew, kurtosis, sp.ent, s-fm, mode centroid, mean-function, min-fun, max-fun, mean-dominance, min dominance, max dominance, d-frange, mod-index, label"]. These features are susceptible to different weight parametric changes from the proposed machine learning algorithm that we have implemented based on the formulation of multinomial approach for Gaussian distribution for Machine learning optimizer.

5.2 FEATURE EXTRACTION

To extract such features based on the either label or any other class columns that governs the important formulations and its solutions, resulting in extraction of different characteristic model ensuring different variables and methods for a class label. For such clusters we improvise the feature capabilities based on formulations mentioned in section 4.4. The feature extraction consists of features in governing the column vector as {mean frequency, centroid, label, mode, skew, kurtosis, Q25, Q75, and IQR}. Finally, these above 9 features are to be classified for Male and female classification features criteria for the design to be implemented.

VI. GAUSSIAN DIST WITH MULTINOMINAL APPROACH

One of the design features that governs on the weights and its implementation of surrogate optimization based on the formulation proposed as mention below:

$$P(v) = E\{\mu * \sum_{k=1}^{M}(\boldsymbol{x(k,j)})_k + \ \boldsymbol{\delta} * \ \sum_{j=1}^{L}\boldsymbol{x(k,j)}_j\}P(v) = E\{\mu * \sum_{k=1}^{M}(\boldsymbol{x(k,j)})_k + \ \boldsymbol{\delta} * \ \sum_{j=1}^{L}\boldsymbol{x(k,j)}_j\} \quad (1)$$

Here the P represents the probability of the governing class and its label that have be calculated by its weight parametric. With input feature for the label classification features and its number of outcomes are represented with $\boldsymbol{x(k,j)x(k,j)}$ with j and k variables of classification labels. Now with this formulation we apply the multinomial approach for Gaussian distribution for each label that will be classified with linear probability function as mentioned in equation 1. Let's consider the LPF would be the governing probability for which each conditional variable on Gaussian dist. as

$$F(x) = \sum_{k,j=1}^{M,N}\boldsymbol{x(k,j)} * e^{-\frac{(x-\mu)^2}{2*\pi*\delta^2}}F(x) = \sum_{k,j=1}^{M,N}\boldsymbol{x(k,j)} * e^{-\frac{(x-\mu)^2}{2*\pi*\delta^2}}$$

We provide an inequality condition for the function $F(x)F(x)$ based on the label features and its criteria for which we have implemented the overall condition resulting the multinomial approach as:

For the label 1 we have mentioned the condition as

$if\,(var\{i,3\} > 0.15 \,||\, var\{i,4\} > 0.18 \,\&\&\, var\{i,6\} > 10 \,\&\&\, var\{i,7\} > 10)$
$if\,(var\{i,3\} > 0.15 \,||\, var\{i,4\} > 0.18 \,\&\&\, var\{i,6\} > 10 \,\&\&\, var\{i,7\} > 10)$ (3)
$var\{i,3-7\} and\ var\{19\} are\ the\ test\ features\ that\ are\ utilzied\ for\ the\ design$
$var\{i,3-7\} and\ var\{19\} are\ the\ test\ features\ that\ are\ utilzied\ for\ the\ design.$

Here with above condition in equation (3) we tend to provide the design model and its feature with class labels for 1-1578 and 1579-3168 for male and female features

VII. EXPERIMENTAL DETAILS, RESULTS AND DISCUSSION

Our design implicates on the feature recognitions on frequency parameter that govern the male or female classification model utilizing the kaggle dataset in [11]. Figure 1 represents the class distribution of the dataset sample for 1568 features of male and female classes separately. With realized case we have implicated the MNGD solution weights to the Decision tree classification for each class comparison based in the columns utilized as in equation 3.

Fig 3. Representing the Overall class label distribution.

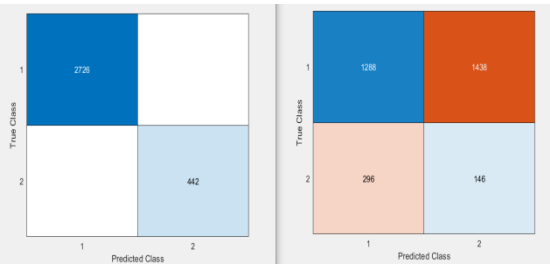

Fig 4. Representing the proposed Weighted MNGD- Decision tree classifier with Normal decision tree classifier confusion chart.

ALGORITHMS	TRAINING ACCURACY	TESTING ACCURACY	POSITIVE	NEGATIVE
MNGD-DT	85.74	72.9	1158*+1126*	884
DT	84.25	45.25	1426	1742

Table 1: Representing the proposed DT and existing DT model.

From the tabulation, table 1, we have represented the training and testing accuracy for the samples of 3168 for reference [11], KAGGLE website, as per the design training scenario we have seen significant change of 1.5% of the accuracy improved, and nearly of 27% difference of the testing accuracy. Figure 2 describes about the confusion chart estimated using the MNGD-DT with training

and testing feature. With aspect on the figure 2 the table1 values are represented with a true positives for the proposed are considered with respect to the values on basis of 1-1568 and 1569-3168 out which only 2284 sample are classified as male and female. Since the condition for male and female classification is depends on the feature model equation 3 resulting a 73% accurate for each male and female data classification. We have improvised the original DT with 27% accuracy on both positive and negative cases also.

7.1. CONCLUSIONS

Our design investigates on the Gender feature of the frequency values classification dataset from the reference 11. This data model with 8 features of classifications on the frequency parametric are modelled with MNGD and solution analysis on equations 1-3. With feature model we have trained with original DT and proposed DT (MNGD) with classification accuracy of 72.9 and 45.25 as mentioned in table1. The overall performance are being adjusted with different weight feature from 0-10 values where value calculated from MNGD is 9.3539. Hence with this weight we have implicated in the Decision tree conditions for each comparison of male and female.

7.2. SCOPE

Even though the design with accuracy value is 72.9, we would improvise the design feature with addition of different clustering algorithms and other boosting feature to implicate much better accuracy and other performance characteristics in confusion chart. A deep learning approach with MNGD layer parametric as custom layer implementation where the weights range from 0-1000.

VIII. REFERENCES

1. M. Gales, S. Watanabe and E. Fosler-Lussier, "Structured discriminative models for speech recognition", *IEEE Signal Process. Mag., vol. 29, no. 6, pp. 70-81, Nov. 2012.*
2. S. Zhang and M. Gales, "Structured SVMs for automatic speech recognition", *IEEE Trans. Audio Speech Lang. Process., vol. 21, no. 3, pp. 544-555, Mar. 2013.*
3. K. Wu and D. Childers, "Gdener recognition from speech. Part I: Coarse analysis", *J. Acoust. Soc. Am., Vol. 90, No. 4, pp. 1828-1840, 1991.*
4. H. Hermansky and N. Morgan, "RASTA Processing of Speech", *IEEE Trans. Speech and Audio Processing, Vol. 2, No. 4, pp. 578-589, Oct. 1994.*
5. N. Narang and T. Bourlai, "Gender and ethnicity classification using deep learning in heterogeneous face recognition", 2016 *International Conference on Biometrics (ICB), pp. 1-8, 2016.*
6. M. Duan, K. Li, C. Yang and K. Li, "A hybrid deep learning CNN-ELM for age and gender classification", *Neurocomputing, vol. 275, pp. 448-461, 2018, [online] Available: https://doi.org/10.1016/j.neucom.2017.08.062.*
7. L. Bui, D. Tran, X. Huang and Girija Chetty, "Classification of Gender and Face based on Gradient faces," *Visual Information Processing (EUVIP), pp. 269-272, 2011.*
8. A. Amberkar, P. Awasarmol, G. Deshmukh and P. Dave, "Speech Recognition using Recurrent Neural Networks", 2018 *International Conference on Current Trends towards Converging Technologies (ICCTCT), pp. 1-4, Mar. 2018.*
9. C. Shan, "Learning Local Binary Patterns for gender classification on real-world face images", *Pattern Recognition Letters, vol. 33, no. 4, pp. 431-437, Mar. 2012.*
10. L. A. Alexandre, "Gender recognition: A multiscale decision fusion approach", *Pattern Recognition Letters, vol 31, no. 11, pp. 1422-1427, August 2010.*

11. Kaggle, "Voice Frequency Values", link: https://www.kaggle.com/primaryobjects/voicegender
12. E. Sariyanidi, V. Dagli, S. Tek, B. Tunc, and M. Gokmen, "A novel face representation using local zernike moments", in *Signal Processing and Communications Applications Conference (SIU), 2012 20th, 2012, pp. 1-4.*
13. Denys Katerenchuk, "Age Group Classification with Speech and Metadata Multimodality Fusion", 2018.
14. J. Shor, D. Emanuel, O. Lang, O. Tuval, M. Brenner, J. Cattiau, et al., Personalizing ASR for dysarthric and accented speech with limited data, 2019.
15. M. Markitantov and O. Verkholyak, "Automatic Recognition of Speaker Age and Gender Based on Deep Neural Networks", *Proceedings of International Conference on Speech and Computer (SPECOM), 2019.*
16. Anil Kumar Maddali and Habibulla Khan, "Classification of disordered patient's voice by using pervasive computational algorithms" *International Journal of Pervasive Computing and Communications,2022.*
17. Anil Kumar Maddali et al., "Functional Analysis and Hybrid Optimal Cepstrum Approach for Gender Classification Using Machine Learning",*International Journal of Emerging Trends in Engineering Research, 8(6), June 2020, 2868 – 2877.*
18. M.Anil Kumar and Habibulla Khan, "User Voice management and Power spectrum analysis for Voice Recognition Systems" International Journal of Advanced Science and Technology Vol. 29, No.02, (2020), pp. 2325-2333.
19. M. Anil Kumar, Habibulla Khan, "An Unmanned Speech Cognizant for Medical Application", Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, 02-Special Issue, 2018.
20. Deena, S., Hasan, M., Doulaty, M., Saz, O. and Hain, T. "Recurrent neural network language model adaptation for multi-genre broadcast speech recognition and alignment", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 27 No. 3, 2019.
21. Schultz, B.G., Aditya Tarigoppula, V.S., Noffs, G., Rojas, S., van der Walt, A., Grayden, D.B. and Vogel, A.P., "Automatic speech recognition in neurodegenerative disease", International Journal of Speech Technology, Vol. 24 No. 3, pp. 771-779,2021.
22. Prasad, N., Galundia, K. and Daksh Gupta, P. "Complete perspective on speech recognition", International Journal of Scientific and Engineering Research, Vol. 12 No. 6, pp. 92-98, 2021.
23. Li, C., Du, J., Liu, Q.-F. and Lee, C.-H., "A cross-entropy-guided measure (CEGM) for assessing speech recognition performance and optimizing DNN-Based speech enhancement", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 29, pp. 106-117, 2021.
24. Khanna, R., Oh, D. and Kim, Y., "Through-wall remote human voice recognition using doppler radar with transfer learning", IEEE Sensors Journal, Vol. 19 No. 12,2019
25. Li, L., Wang, D., Chen, Y., Shi, Y., Tang, Z. and Zheng, T.F., "Deep factorization for speech signal", Proc. IEEE Int. Conf. Acoust., Speech Signal Process., pp. 5094-5098, 2018.