Hindawi
Journal of Sensors
Volume 2024, Article ID 9843217, 1 page
https://doi.org/10.1155/2024/9843217

Retraction

Retracted: *SmHeSol* (*IoT-BC*): Smart Healthcare Solution for Future Development Using Speech Feature Extraction Integration Approach with IoT and Blockchain

Journal of Sensors

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Upadhyay, M. Kumar, A. Kumar, K. Z. Ghafoor, and S. Manoharan, "*SmHeSol* (*IoT-BC*): Smart Healthcare Solution for Future Development Using Speech Feature Extraction Integration Approach with IoT and Blockchain," *Journal of Sensors*, vol. 2022, Article ID 3862860, 13 pages, 2022.

Hindawi
Journal of Sensors
Volume 2022, Article ID 3862860, 13 pages
https://doi.org/10.1155/2022/3862860

Research Article

SmHeSol (*IoT-BC*): Smart Healthcare Solution for Future Development Using Speech Feature Extraction Integration Approach with IoT and Blockchain

Shrikant Upadhyay,[1] **Mohit Kumar**,[2] **Ashwani Kumar**,[3] **Kayhan Zrar Ghafoor**,[4,5] **and S. Manoharan**[6]

[1]*Department of Electronics & Communication Engineering, Cambridge Institute of Technology, Ranchi 834001, India*
[2]*Department of Computer Science & Engineering, Cambridge Institute of Technology, Ranchi 834001, India*
[3]*Department of Computer Science and Engineering, Sreyas Institute of Engineering and Technology, Hyderabad 500068, India*
[4]*Department of Computer Science, Knowledge University, University Park, Kirkuk Road, 44001 Erbil, Iraq*
[5]*Department of Software & Informatics Engineering, Salahaddin University-Erbil, Erbil 44001, Iraq*
[6]*Department of Computer Science, School of Informatics and Electrical Engineering, Hachalu Hundesa Campus, Ambo University, Ambo, Ethiopia*

Correspondence should be addressed to S. Manoharan; manoharan.subramanian@ambou.edu.et

Received 22 February 2022; Accepted 21 April 2022; Published 30 May 2022

Academic Editor: Paulo Jorge Sequeira Gonçalves

Voice of any human plays an important role in communication and sharing information among each other. Through voice, internal behavior can be identified as to whether the person is happy or angry which is reflected. A person's behavior is not exactly reflected by their face; variation in his/her voice reflects somehow their behavior as there will be variation in voice and variation in frequency and pitch. Feelings and natural behaviors are important features, and there are many biological aspects through which they can be identified. Therefore, in this paper, we consider a Hindi speech specimen of different groups to identify the person's behavior and natural feelings under different acoustic conditions. Many research papers show emotion recognition based on neural networks with different models using speech signals to identify the present status of any patient, and it helps to build a way for a smart healthcare system. Enabling service in terms of Blockchain means the sufferer does not require communicating with complex and failed tasks for collecting information from various sources to send to their expert. Blockchain provides experts access to systems and enables entry to the dataset section. Patients have total control over their data, and they no longer require monitoring to keep their data managed and up to date. Also, manually coordinating with data is required for multiple visitors, which can be a very tedious one. In this paper, we focused on feature removal of speech using different extraction approaches which were used to know the quality or state of voice specimens and also understand which feature extraction plays a vital role in gaining a close state of speech. Internet of Things-based learning platforms are used to gather the voice sample, and also, a deep gaining method was followed to reach and achieve the best accuracy and identify the error rate which will help to gather close behavior and state of mind. Finally, a proposed model based on the Gaussian mixture model as a classifier was used for its spotting and attestation.

1. Introduction

In the past few years, effort has been put into the domain of speech recognition. An attempt has been made for small, medium, and large vocabularies. But the optimum result was found to be valid if recorded data from the database for testing was performed with the same scenario against the training dataset. The speech signal system is affected mostly when background distortion rises [1].

The addresser identification method involves automatically recognizing the person who is talking depending upon the individual's speech wave information and its quality. The technique follows the talk of a person to recognize its original quality and provides private execution of any facility which includes dataset access, dialing through voice, medical emergency inputs, mail through voice, and secured and confidential data which also provides execution of computer information in remote areas and other many facilities associated with speech.

Addressing through is complicated as it is associated with speech, and there will be various transformations that will arise at several stages which include interpretation, auditory, verbal, and lingual. Transformation differences are found in the verbal quality of talked speech; also, there is an addresser-related issue which involves the combination of corporeal dissimilarity inborn in the track of vocal and trained talking practice of any individual. In addresser identifications, all their dissimilarity is considered and used to differentiate two addressers, and it must be taken into account [2, 3]. Any behavior or mental status in terms of emotion identification using talked voice attracts more to detect the present status of a person. Through speech, various thoughts like behavior, emotions, and nature can be detected [4]. Smart healthcare development needs such an approach to make its structure strong using speech to realize the feeling of humans and reflect it to disclose sentiments [5, 6]. A smart health unit uses an IoT-based system that can solve such issues, and it is a challenging one and can provide solutions for problems arising in real time [7].

2. IoT-Based Smart Healthcare System

Embedded technology that contains a diversity of physical gadgets is an IoT used for detection and communication purposes. IoT provides technology to make daily life easy as its devices contain smart phones, detectors, actuators, etc. Development in the communication sector with the growth of mobile customers increases the strength of cloud-based technology for smart health systems and cities to enhance people's life. A global ecological platform created by IoT where more than one device can be allocated and divide data with a cloud-based server completes their goal without any human interlinkage and support to generate new operations. The network requirements' different communication protocols like wide-fidelity, Bluetooth, and 5G can be used as user demand to fulfill the satisfaction level in a defined communication range. The related and delivered data used for classification can be analyzed with proper planning and resolution-making where the created dataset from the raw sample is sent to the cloud network or any third party parts of the network system. IoT brings flexibility and luxury as it creates an easy path to handshake with various kinds of smart gadgets for observing different conditions and for the purpose of monitoring and communication [8–11].

Encapsulated sensors and IoT gadgets help to measure different body parameters like oxygen level, blood pressure, body motion, and ECG signal in home and hospital premises. Emotions and behavior are also one of the human features which can be diagnosed as it is one of the important aspects of every human being. So, by capturing all the important features of speech and by reviewing the patient short- and long-term medical records, his/her treatment is possible at low cost and with minimum clinical cost [12]. Previous many research articles only focus on the emotion identification by collecting specimens of one group which was done that was not sufficient as the vocal track, resonance, and pitch are not considered as all these are important features of any human voice. So, such issues are taken into consideration for analysis. In medical analysis of voice, their variations, tone of voice, age factor, and quality are some of the physiological parameters that will also be taken into account for better performance [13–15]. In smart healthcare systems, the WSN, ultrahigh frequency, radio frequency identification, smart mobile, and GSM technologies were used for implementation [16]. The main goal of this research was to maximize early disease detection and minimize the diagnosis error and better prognosis [17]. Kumar et al. gave many solutions for detecting the object from the images using machine learning algorithms [18–21]. Figure 1 demonstrates the structure of the basic smart healthcare system [16].

3. Related Work

Feature recovery of any talk is important for any addresser. With the help of different feature recovery approaches, the pattern can be matched and modeling of signals is possible. Gathering of desired parameters which are crucial for those used by a person hearing structure is an immutable restriction which is crucial for the addresser and converter to maintain robustness in fluctuations while sending to the channel.

Modeling of signals was done in this research paper which involves feature removal, shaping of spectra, transformation invariable, and its modeling. Variation in the spectrum with respect to time is one of the important issues that are also taken into consideration. The IoT platform is used to deliver the control inputs gathered by the authorized doctors or medical team, and the server detects the authenticated person's voice and gives feedback again to authenticate one's credential. It is found true that it will send the data for further processing. The IoT platform consists of hardware with its core component called raspberry PI. Tiwari et al. proposed a hybrid-cascaded framework for image reconstruction [22–25]. The basic structure consists of IoT gadgets, dataset, cell phone, cloud service, monitoring unit, etc. IoT gadgets in the form of wrist watch, wearable

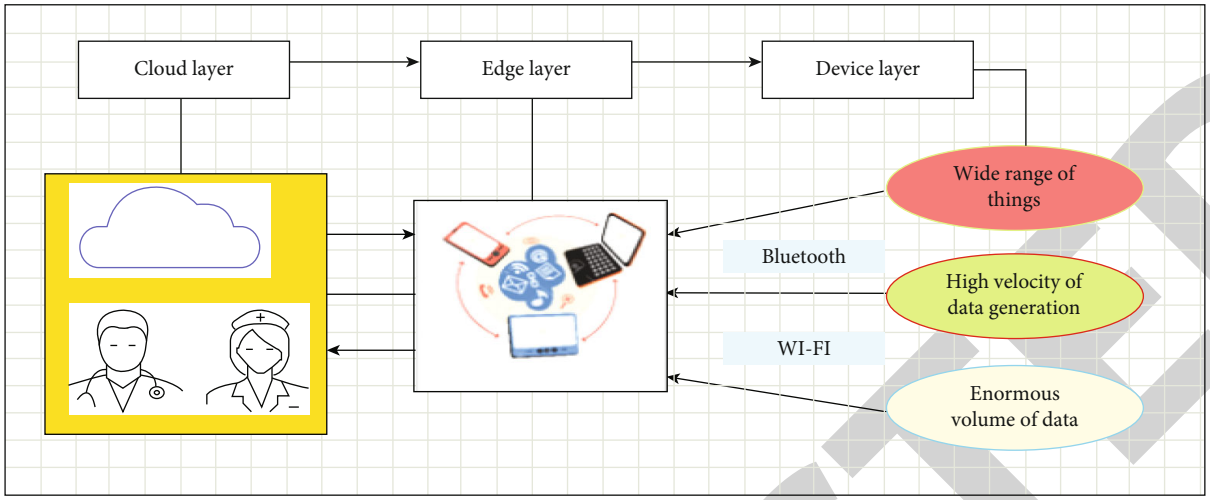


Figure 1: Structure of basic smart healthcare system [16].

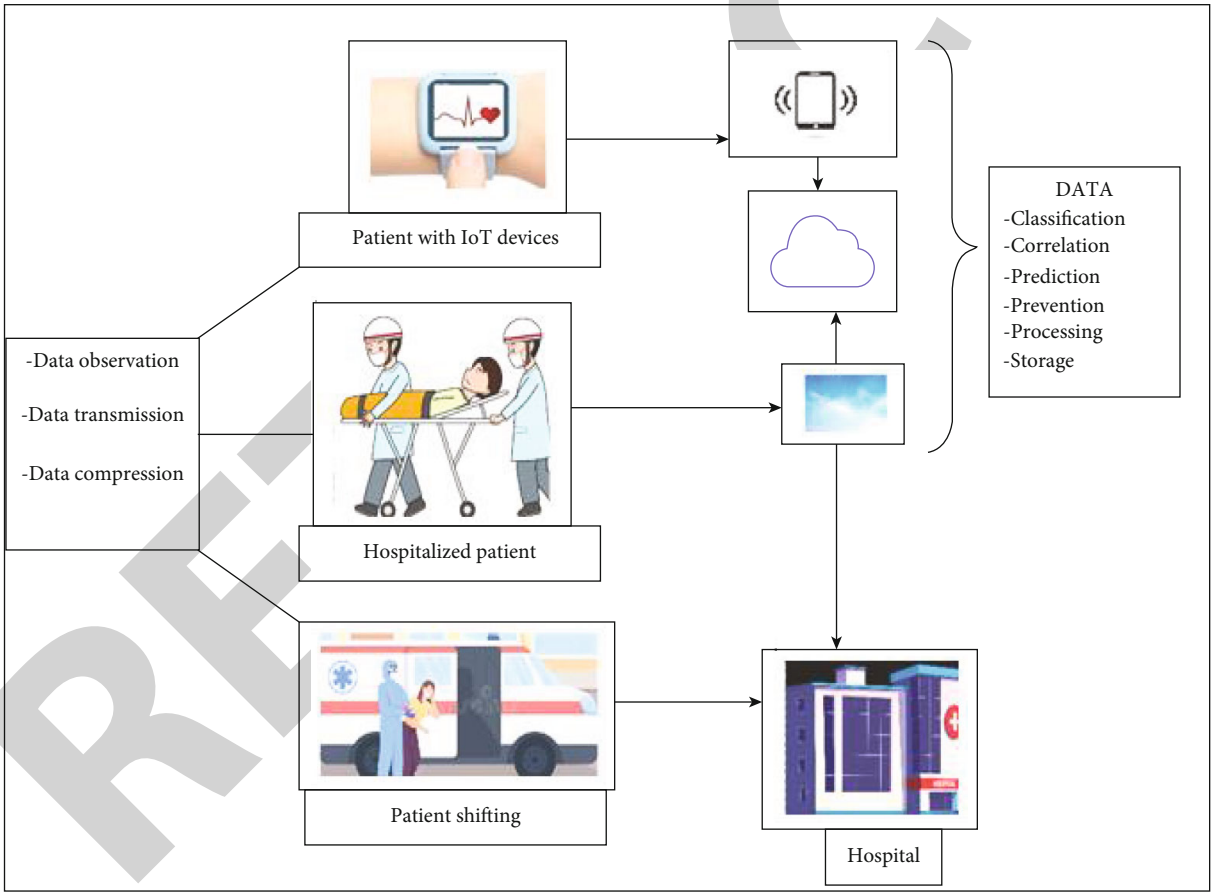


Figure2: Requirements of smart healthcare system [22].

shirt with sensors, shoes, etc., interfaced with a cell phone which contains an app to read the inputs coming from different parts of the body. A smart cell phone contains a recognizable operating function and makes it possible to deliver information for its prediction, processing, and storage of data. The monitoring unit is directly interfaced with the healthcare unit, IoT gadgets, and activity done in hospitals. This unit deals with data, observation, transmission, and compression of large information to minimize the storage capacity. Figure 2 shows the various requirements of an effective smart healthcare system [22].

The healthcare system when integrating speech with IoT enabled will be challenging for the system to take out the complete feature and its facts, and it is crucial to gather the complete information to compete with real-time devices. A quick response might be an essential requirement for the

emergency need, and through speech, it is quite easy for the person to react and respond easily. Fact transmission, gathering of facts, and their compression are very important requirements for the IoT system. Keeping all these challenges, the integration and features that are removed must match the system at the maximum threshold to authenticate the person and identify its present mental status for quick response. Salamon and Bello [26] presented a second convolution neural community of three convolutions, accompanied by max-pooling and two hidden layers for environmental sound classification. For enhancing the accuracy of results, they applied the fact augmentation approach. Our approach and results show good accuracy in emotions and behavior predictions through the process of feature extraction and normalization, and predictions show better results. Castillo et al. [27] delivered clever technologies for detecting emotion in aged care. They in particular used cameras and frame sensors together with environmental elements such as tune, colour, and mild, inside the detection of older people's feelings. We have advanced a live audio sentiment analysis device with the usage of audio sensors to locate human beings' feelings via an audio class technique. Figure 3 shows a real-time speech identification system for emotions and behavior based on IoT and deep learning with speech emotions with its classification and workflow.

4. Blockchain Role for Smart System

IoT brings new evolution to the healthcare sector with plenty of healthcare applications and health devices in markets. A bulk amount of information associated with healthcare are measured and delivered every hour inside or outside the healthcare unit for proper observations of patients. This huge amount of patient observation medical data with important information requires better management in terms of data security, privacy, and its availability. Patients' medical status required complete monitoring by hospitals as well as doctors for better treatment while maintaining privacy and data security of sensitive data of patients to share them with medical institutes and leading hospitals for expert consultation for gathering better information about related cases. Accountability law and health insurance law enforcement and many other public agencies access medical information legally, and it is approximated that around 200 to 500 individuals may have the right to read the health records of any patient without any authentication and permission [28]. When information is distributed extensively and kept in multiple outbreaks, securing information is one of the more crucial issues. As per the Ponemon Institute, in the year 2016, about 112 million information related to medical was negotiated, and such breaks of data tampered with and attacks raised by 162% in the year 2017 [29, 30].

5. Methodology and Modeling

Modeling of signals can be done in the following ways. First is to shape the voice spectra; in the second step, the extraction of features will be done; in the third step, it involves modification of parameters involved with speech, and lastly, modeling will be done using analysis.

5.1. Shaping of Voice Spectra. Shaping of talked spectra requires two important functioning called filtering and its digitization [31]. In digitization, transformation of analog signals received in the form of sound waves can be converted into its digital form where filtering focuses on the crucial element, i.e., frequency ingredient present in signals. The shaping process is depicted in Figure 4. Digitization is one approach that is used to generate specimen data for the representation of talk signals with a high degree climb for signal intensity to noise intensity as possible. Conversion of signals over the final stage involves digital lookout cleaning and is accomplished using filters of impulse response given as

$$D_{\text{emp}}(y) = \sum_{k=0}^{n_{\text{emp}}} b_{\text{emp}}(k) y^{-k}. \tag{1}$$

Normally, prior attention filtrations used are called coefficient filter which is of digital type:

$$D_{\text{emp}}(y) = 1 + b_{\text{emp}} y^{-1}. \tag{2}$$

The classical value for b_{emp} is $[(-0.1)\text{-}(-0.4)]$, and this will help to strengthen the spectrum of signals nearly to about 20 dB decade.

Prior attention filtration offers various advantages as naturally generated signals have an unfavorable negative slope nearly 20 dB/decade due to the physiology effect of the talked system [32–34]. And filtration turns down this natural slope prior to spectral investigation [35]. Also, aural is more delicate on the above 1-kilohertz locality of spectra which amplifies the nearby spectrum and helps the spectral algorithm construction and modeling of the crucial feature of the talked spectrum.

5.2. Removal of Feature. In speech identification where the addresser is independent, a superior offer is put down on feature removal that is somewhat undeviating to variation in the addresser which involves analysis of talked or spoken signal which is further classified into spectral investigation and temporal investigation.

5.3. Investigation of Decisive and Filter Margin. This investigation is basic and essential in talked processing. It is considered an unprocessed model of the starting phase of the natural process in the person hearing system.

The filter margin as per "place theory" is the spot of higher shift along the primitive layer for incitation such as unmixed tones which is proportional to the frequency of logarithm of the tone [36].

Human voice perception consists of a complex sound frequency within definite bandwidth of some minimal frequency which cannot be identified individually unless one of the quantities of this sound comes above the bandwidth which is known as decisive bandwidth [37].

The approach with a combination of filter and decisive forms a theory where decisive and filter bank involves

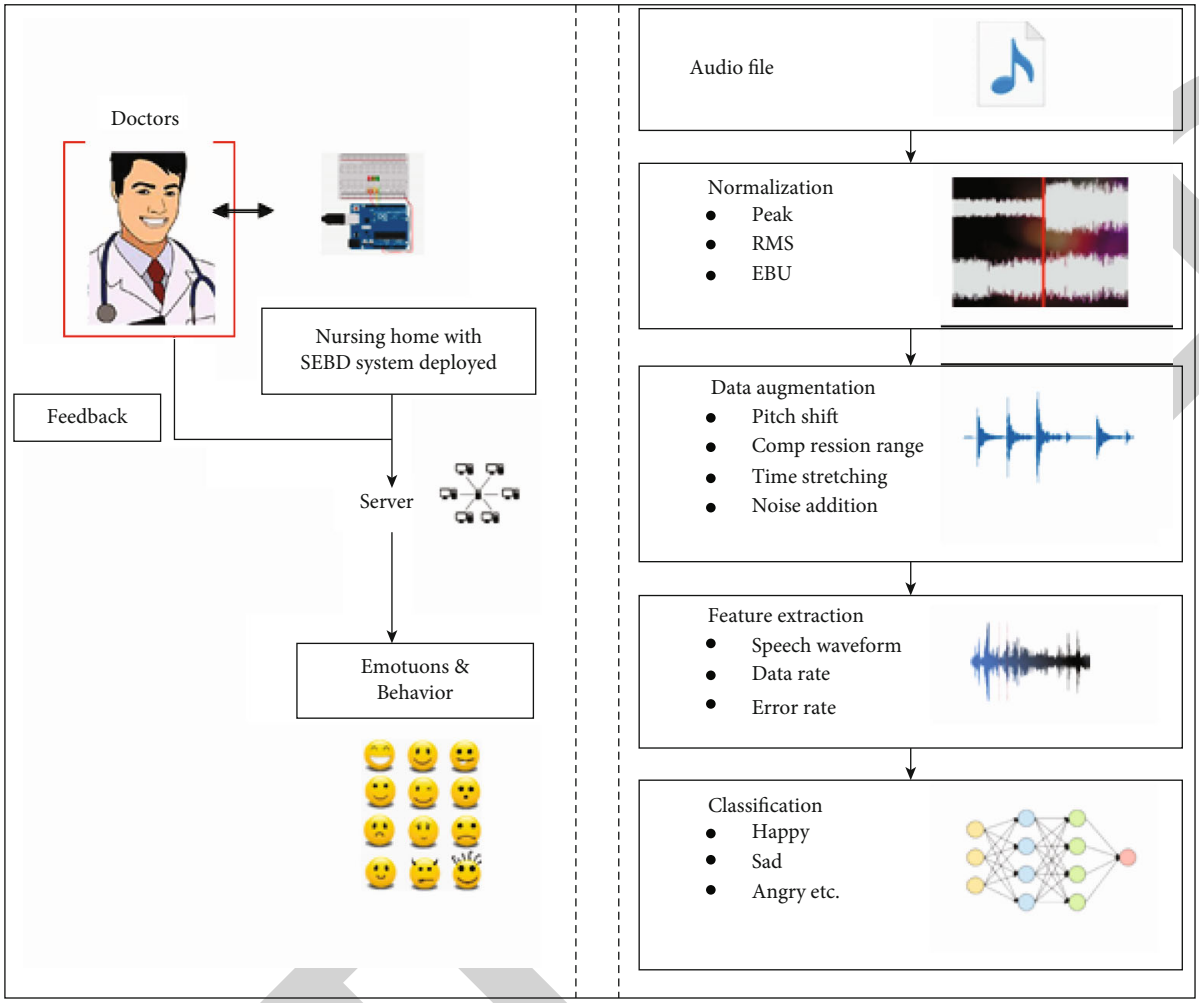


Figure 3: Real-time speech feature extraction integration based on IoT and Blockchain. (b) Speech emotions with its classification and workflow [27].

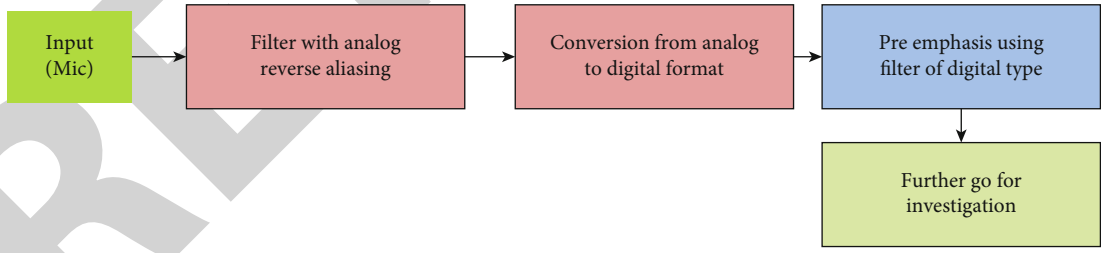


Figure 4: Shaping of spectra with its operation.

bandpass filtrate of linear aspect finite impulse which is arranged straightaway along the perceptual scale where bandwidths are considered to be equal to decisive for communicated centralized frequency [38]:

$$\begin{aligned} \text{Mel} &= 13\,\text{atan}\left(\frac{0.76f}{1000}\right) + 3.51, \\ \text{Mel} &= 3.5\,\text{atan}\,\frac{(f)^2}{(7500)^2} + 13\text{atan}\left(0.76\frac{f}{1000}\right). \end{aligned} \tag{3}$$

Therefore,

$$\text{frequency} = 2595\log_{10}\left(1 + \frac{f}{700}\right). \tag{4}$$

Then, decisive bandwidth can be written as

$$\text{BW}_{\text{dec}} = 25 + 25\left\{1 + \frac{1.4f}{1000}\right)^2\Bigg]^{0.69}. \tag{5}$$

To make the group delay uniform for all delays, a linear phase filter is generally used and filter output signals are properly synchronized with respect to time. Then, the filter equation may be written as

$$\frac{(P_i - 1)}{2},$$
$$R_i(n) = \sum \beta_i(k) r(p + k), \quad (6)$$
$$k = \frac{(P_i - 1)}{2},$$

where $\beta_i(k)$ indicates k^{th} coefficient of i^{th} decisive filter band.

The analysis in form of output is a power value in vector for every supporting data. They are usually combined with parameters like average power to form vector measurement of signals. The bank of filter then attempts to decay the signal into a discrete level of spectrum specimens that must have similar information that is presented at a higher level for the auditory system. It is highly robust to the nose as this approach follows a linear processing rule [39].

5.4. Celestial Analysis. The celestial analysis is very crucial as it supplies the methodology to isolate the vocal activity and its stretching [2]. In speech generation for linear acoustic templates, the mixed spectrum of speech contains activity of filtered signals using time-assorted filter linear in characteristics for denoting the shape of vocal stretching as depicted in Figure 5.

The talked signal is represented as

$$v(n) = h(n) \cdot s(n), \quad (7)$$

where $s(n)$ is the impulse response of the talked route and $h(n)$ is the agitation signal.

The representation of equation (7) in the frequency domain is

$$V(f) = H(f) + S(F),$$
$$\log V(f) = \log [H(f) = \log S,$$
$$V(K) = \sum_{K=0}^{K-1} vv(n) \exp\left(-\frac{j2\pi}{K}\right)k, \quad (8)$$
$$\widehat{v}(k) = \frac{1}{K}\sum_{n=0}^{K-1} \widehat{V}(k) \exp\left(-\frac{j2\pi}{K}\right)k.$$

In talked identification, cepstrum verification is used for pitch and tracking of format. Sample $\widehat{v}(k)$ in the first 3 milliseconds is removed from the agitation.

5.5. Analysis of Mel-Scale Cepstral. Mel-scale analysis follows nonlinear quantity in terms of the frequency axis that uses cepstral. To get the mel cepstral, speech signal $v(k)$ is recovered using filtration called window $w(n)$, and its discrete Fourier $V(k)$ is calculated.

The height of $V(k)$ is then summed with a series of filter frequencies (mel) whose BW and center frequency matched with audible bandpass filters.

In the next further step, the energy is evaluated in this loaded sequence. If $S_l(j)$ is the response of the j^{th} filter scale, the resulting energy for every talked frame at given time "n" will be

$$F_{\text{me}}(n, j) = \frac{1}{C_i}, \quad (9)$$

where M_j and U_j denote higher and lower frequencies where every filter is nonzero and C_i is the power of filter to normalize as per BW varying to give equal power for a flat range.

6. Voice Sample for Audio Analytics

The different voice samples of different ages and sexes expressing his/her feeling in Hindi text are collected using the above device samples collected using Cool Software, and samples are collected for different emotions and behavior performance in the form of spectrogram features as shown in Figure 6. To increase classification, we have designed a green version of the usage of data normalization [9] and fact augmentation techniques in deep gaining knowledge of workflow for the classification of emotions and behaviors.

7. Feature Extraction of Speech

When the incidents have unspecified random phenomena, then a stochastic prototype is used to exhibit the expression of probability density functions (pdf_s). Here, the given frame corresponding to each observation vector is considered to be random, so each spoken word generated by the speaker is considered to be a random sequence of feature vectors. This model creates a perfect model for the sequence attending to the random sequence statistics such as its variance, mean, or probability distribution. The template of the feature vector probability distribution corresponding to a given speaker differs from the other speakers. Thus, the main objective of this prototype is to calculate the prospect point of a spoken phrase for every speaker model [40–42]. Hidden Markov and Gaussian mixture models are the illustration of stochastic structure complementary algorithms. Figure 7 represents the pattern matching system [40].

The Hidden Markov Model (HMM) can be designed as a succession of feature vectors fully accurate where GMM (Gaussian mixture model) takes only a single feature vector corresponding to a single frame. HMM is efficient for text-dependent tasks, and GMM is good for text-independent tasks. Pattern matching is shown in Figure 8. A dissimilar number of classes specified for each speech shape are first like voiced or unvoiced; then, it will use vector quantization in order to group feature vectors according to their similarity, and the last phase will use speech sound information. In the teaching phase, a pattern complementary algorithm will use whole training feature vectors to make speaker models.

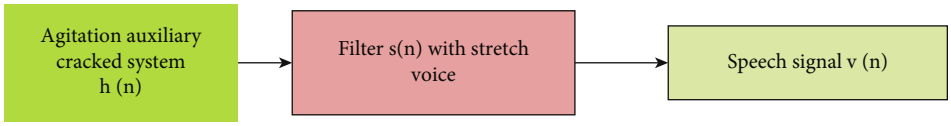


Figure 5: Model for speech generation.

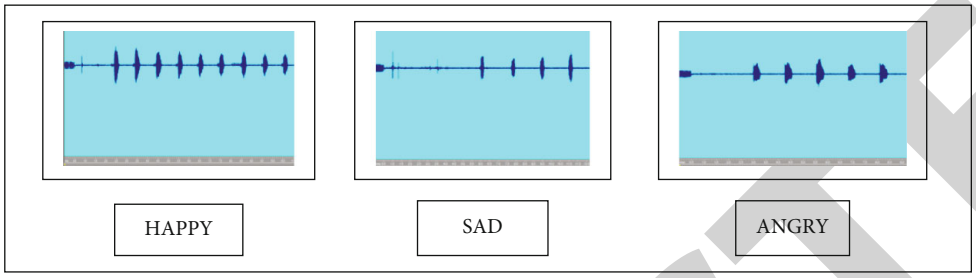


Figure 6: Voice spectrum of different emotions using Hindi text.

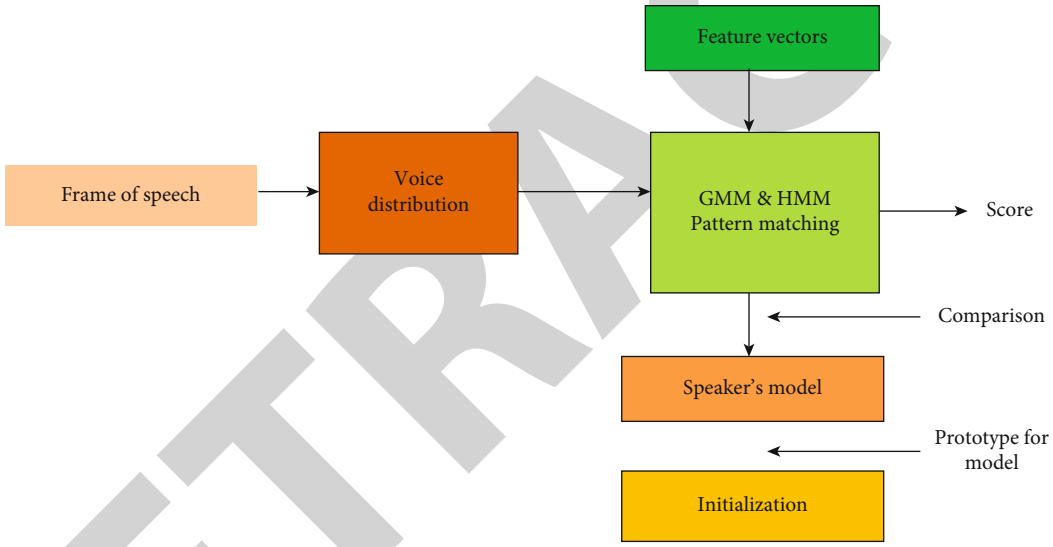


Figure 7: Pattern matching system [40].

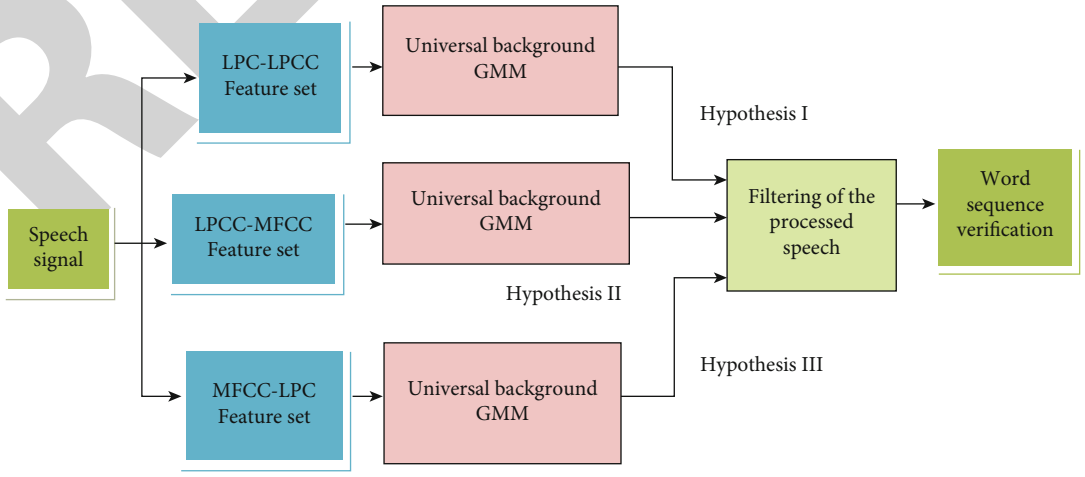


Figure 8: Proposed model for speech validation and identification.

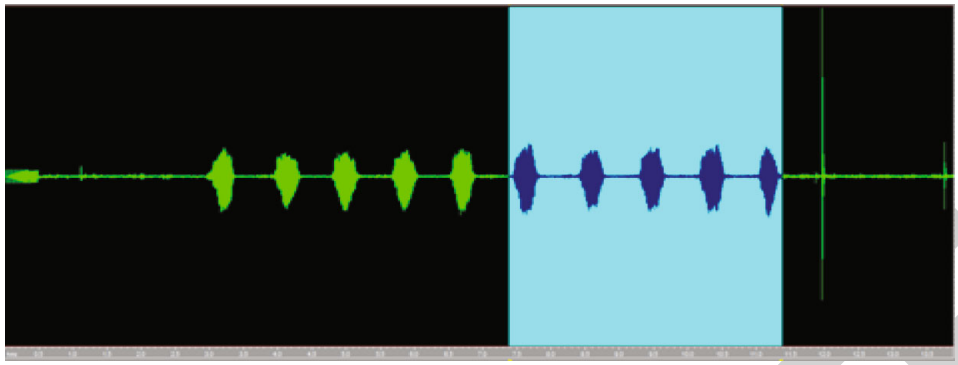


Figure 9: Collection of specimen of talked (sample taken during experiment).

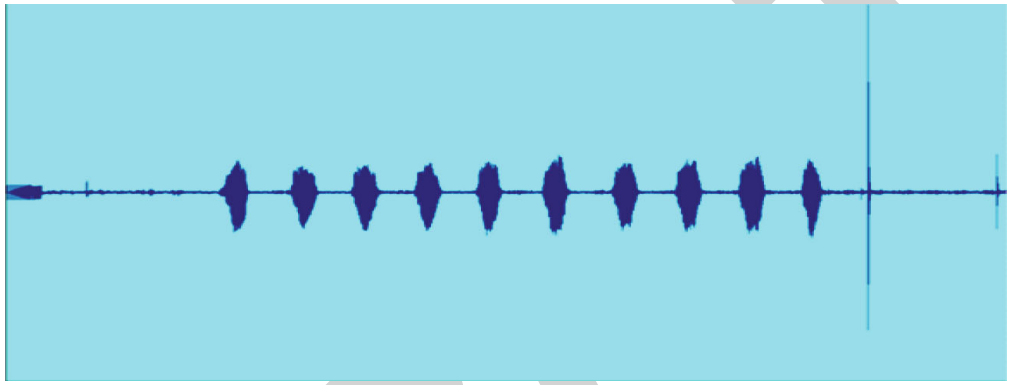


Figure 10: Specimen of voice samples (sample taken during experiment).

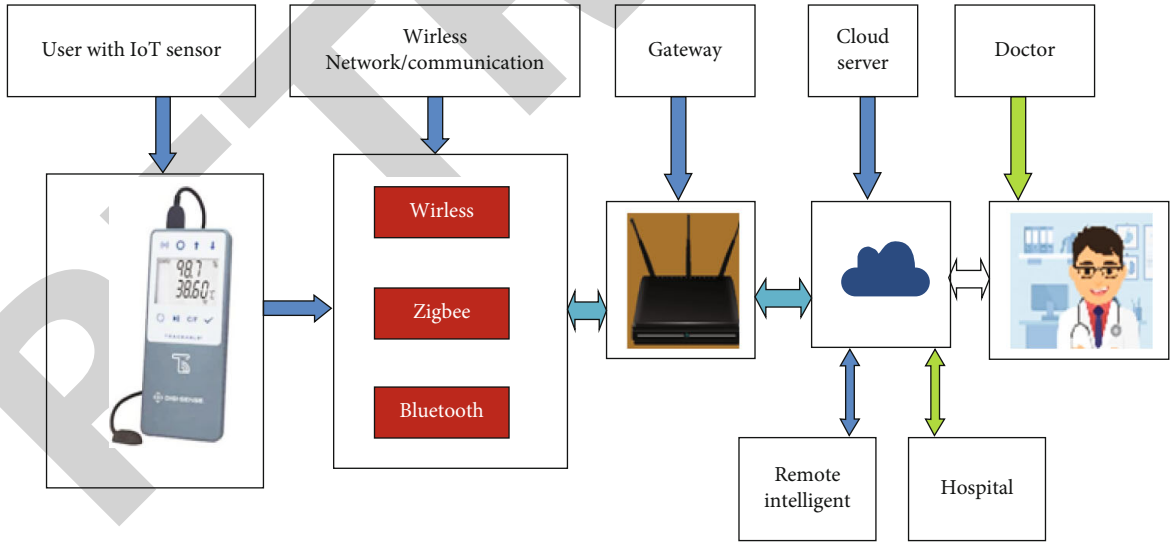
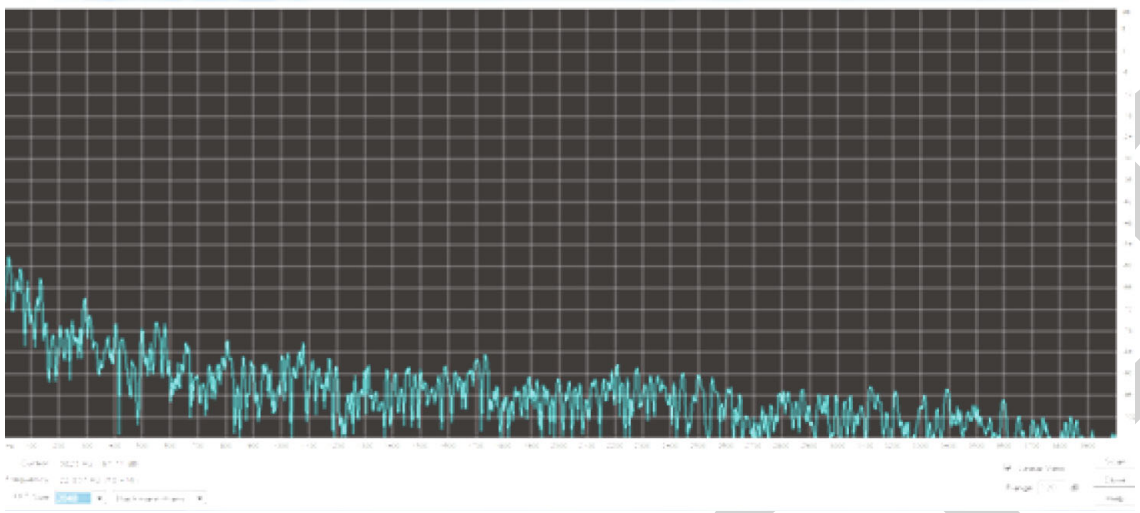


Figure 11: Proposed self-tracking using IoT enabled using speech feature for smart facility in hospital [34].

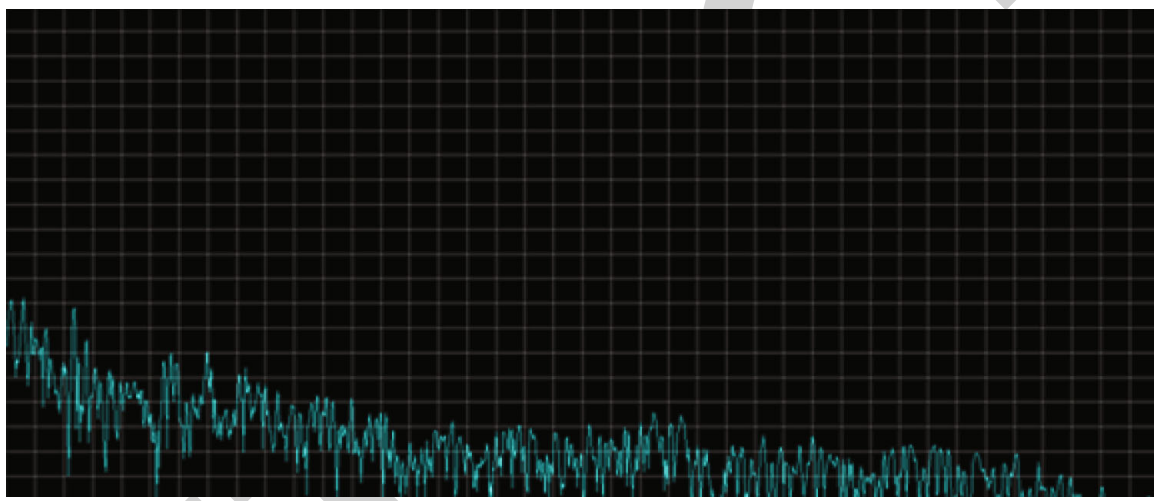
Per voice and per speaker for each model will be created [43]. The model will be called initial model architecture and reestimate and value them accordingly. The three feature extraction taken for behavior identification are Linear Prediction Cepstral Coefficients (LPCC), Linear Predictive Coding Analysis (LPC), and Mel-Scale Cepstrum Coefficient (MFCC).

8. Propose Model with Integration of Feature Integration

For the second model, we prepare different subsystems having their own feature set and acoustic phonetic modeling using a microphone for acoustic to phonetic realization including universal background GMM as one of the



(a)



(b)

Figure 12: Variations in frequency for different talks with emotions (a). Variations in frequency for different talks with emotions (b).

Table 1: Hindi spoken digit.

रोतेहुएबोलना (cry out)	गुस्साकबोलना (to speak angrily)
आईलवयू/i love you	द्रदभरीआवाज/painful sound
जोरसेबोलना/blurt	चल्लिाकरबोलना/shout out
घबराहटकेबादबोलना/speaking after panic	हँसतेहुएकहना/laughing to say
मधुरगाना/singing sweetly	बनिासोचेसमझेबोलना/speaking without thinking

Table 2: Sound file.

Serial no.	Name & sex	Types of emotions	## file (original)	# accelerated file
1	Anuj_56F	रोतेहुएबोलना/cry out	96	1248
2	Sandeep_65M	आईलवयू/i love you	96	1248
3	Rajim_42F	जोरसेबोलना/blurt	96	1248
3	Sarojni_43F	घबराहटकेबादबोलना/speaking after panic	96	1248
5	Narang_39M	मधुरगाना/singing sweetly	96	1248
Total			480	6,240

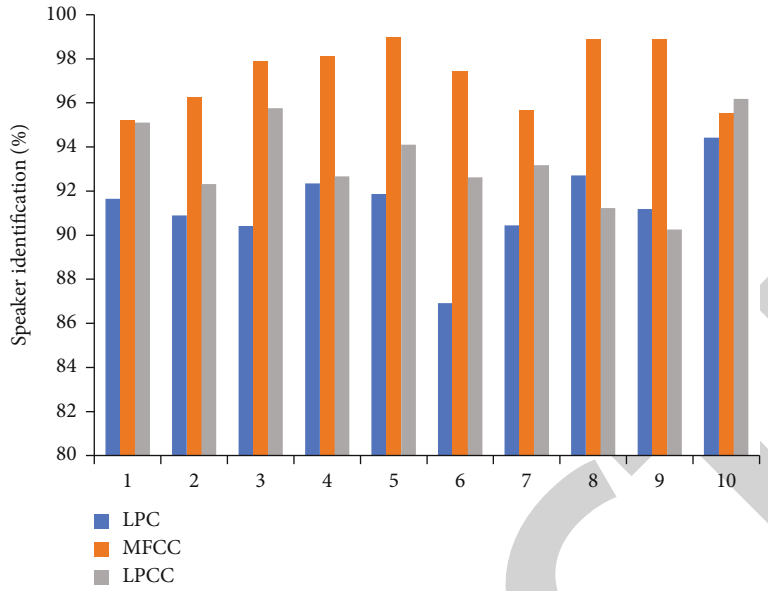


Figure 13: Behavior prediction using feature extraction.

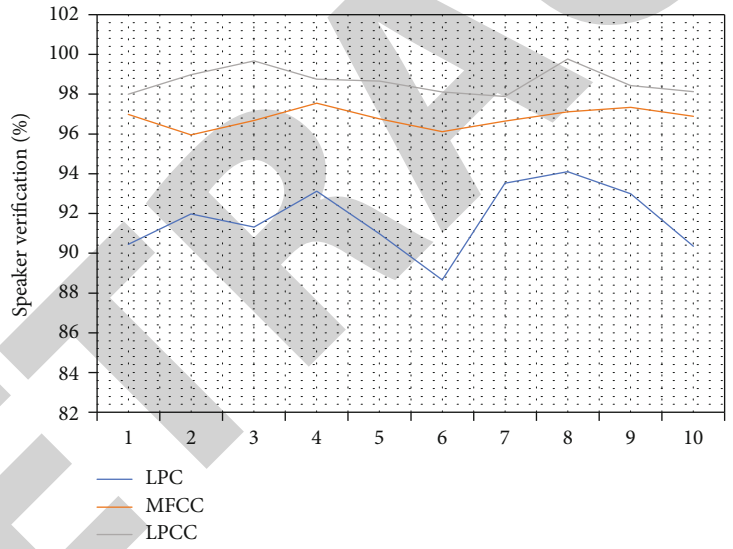


Figure 14: Efficiency for speaker verification (using different feature extraction).

Table 3: Efficiency rate in % of combined feature extraction.

Hindi dialects	LPC+LPCC (%)	LPCC+MFCC (%)	MFCC+LPC (%)
HAA	91.21	97.76	98.98
TUM	92.89	96.54	98.65
JAO	92.12	95.99	97.22
NAHI	94.98	96.39	97.52
KAHAN	91.67	94.78	98.99
RAAT	89.22	95.54	98.54
JEENA	90.29	95.65	97.98
MAUSAM	95.87	93.92	99.10
JAISAY	92.65	92.32	97.87
KYA HUA	92.10	93.99	98.22

classifiers of this model and try to analyze the efficiency and error rate for the same text shown as depicted in Figure 8.

We also prepare different subsystems having their own feature set and acoustic modeling using a microphone for acoustic to phonetic realizations of voice signals. Context-dependent phonemes are used as a basic unit of phonetics in this modeling structure. Subsystems generate hypotheses independent using different mechanisms but are compatible with each other. Now, keeping combinations of the proposed combinations of features, we try to analyze the performance in terms of efficiency and error rate of Hindi text. Figures 9 and 10 are collections of specimens of talked and specimen of voice samples, respectively.

The technical aspect of the tracking mechanism depends upon both IoT devices and the interfacing unit that helps to send and receive the information that consists of IoT

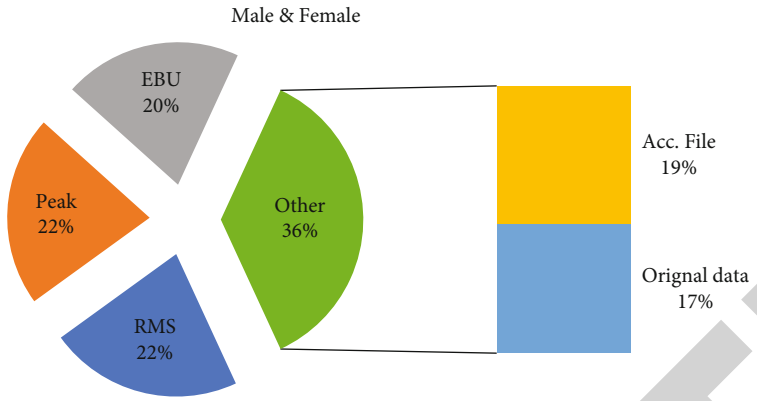


Figure 15: Prediction of male and female talk for different throws.

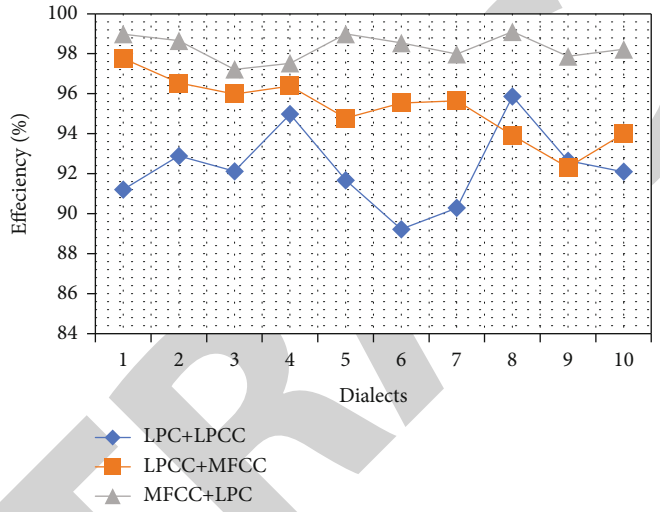


Figure 16: Efficiency rate in low acoustic condition.

gadgets, cloud server, datasets, and processing unit. The tracking mechanism is followed by the IoT sensor which is used to send the generated or gathered data to the wireless network using gateway service which is directly connected with the cloud server. This server directly transfers the information to the remote intelligent hospitals and the healthcare section (concern doctors) to share various information and prepare them to take some effective steps against any emergency. Figure 11 shows the proposed self-tracking using IoT enabled using speech feature for the smart facility in hospitals [34].

9. Results and Discussion

The dataset has recordings of 20 males and 20 females, and these recordings have extraordinary emotion categories, i.e., calm, satisfied, unhappy, indignant, worried, wonder, disgust, and impartial. We took into consideration seven emotion classes for experimentation. We decided to work with men and women facts one at a time.

The variation in the spectrum for different emotions and behavior is shown in Figures 12(a) and 12(b). The facts have been validated using the specimen of masculine and lassie of different age groups to evaluate the exact variations in their tone and get the various ingredients to analyze the performance of quality in terms of pitch, variation, clarity, etc. The above self-tracking model will help to access the remote information coming from rural areas and provide better precision in terms of identification of speech for better response and support healthcare to provide urgent help to people. The cloud acts as a server for sharing the facts to the nearest local healthcare unit to reach the people who require emergency support and grant medical facility for better health.

The various Hindi spoken digits considered for the analysis are represented in Table 1, and the sound file for different emotions with their original and accelerated file of different age groups is represented in Table 2.

The speaker identification performance for different feature extraction is represented in Figure 13 where the performance of MFCC is found to be efficient compared to LPCC and LPC.

The speaker verification performance for different feature extraction is represented in Figure 14 where the performance of MFCC is again observed to be efficient compared to LPCC and LPC.

9.1. Experiments with Combined Feature Extraction in Noise-Free Environment Using First Model. Here, the combination

of three feature extraction techniques, i.e., LPC+LPCC, LPCC+MFCC, and MFCC+LPC, has been analyzed for the same database of 50 speakers in ideal conditions. Integration of feature removal tries to analyze the various qualities of talked speech and is used to gather the collection of information having variations in speech with his/her feeling, and it covers its emotions also. Integrating the quality of feature with IoT is really a challenging one as the system requires continuous analysis in real time for monitoring and updating the status. The integration of feature extraction for different Hindi dialects and its performance is represented in Table 3.

The prediction of male and female talk throw performance using three faithful normalization approaches, i.e., EBU, peak, and RMS, which show the best accuracy for the female voice is depicted in Figure 15.

After integrating the considered feature extraction technique, the combination of MFCC+LPC is observed to be quite efficient; this is found to be nearly about 98.92% as shown in Figure 16.

In this experiment, the second model using combined feature extraction was performed in which the performance of MFCC+LPC has been found more efficient, i.e., 98.22%, compared to that of LPCC+MFCC and LPC+LPCC, i.e., 93.99% and 92.10%, under such scenario.

10. Conclusion and Future Scope

From the above result, we conclude that the prototype for real statistics when executed for feminine talk voice was found to have greater precision with 98% contrast to masculine talk/audio voice and found to be maximum at 95%. We also noticed that our model was performing excellent, but some audio was found to have low accuracy due to noise, time, and compression. Hence, we normalize the audio using promising approaches like RMS, peak, EBU, and accelerated file where the accuracy of the male voice is found to be 92% whereas for female, it was about 98%. The behavior of different feature extraction was analyzed, and three important features LPC, MFCC, and LPC are simulated with the standard and classical model of speech identification and verification following all the standards and regulations of the models. And it was observed that the feature extraction behavior of MFCC is quite efficient, i.e., approx 97%, as compared to LPC and LPCC in acoustic conditions. Integration of feature extraction gives rise to the efficiency rate of 98% which is quite impressive for precision and reflects better to implement for the system. In deep learning, challenging tasks for tutoring prototypes with small fact and observed fact data for training prototype increasing data samples may increase accuracy and real-time prediction. For augmentation (data), we follow pitch shifting, time stretching, and range compression without any external effect. We have applied 60-40% ratio tutoring and validation used to take out spectrogram of speech in terms of features which increase the number of samples in terms of audio and model representing better accuracy for female voice, i.e., 98% and 97% for male voice. The accuracy prediction for behaviors was also done for male and female voices where female voice was found to be more accurate compared to male voice due to pitch stretching in male voice. This can be adopted in the smart healthcare system to identify the emotions and behavior of any patient in the future. Blockchain provides access to facts given by the source and is suitable for speech as it extracts all its behavior of speech and makes it secure for future real-time applications.

Data Availability

Data will be provided upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Furui, "Speaker-independent isolated word recognition using dynamic features of speech spectrum," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 1, pp. 52–59, 1986.

[2] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, no. 4, pp. 1334–1345, 2007.

[3] S. Z. Bong, K. Wan, M. Murugappan, N. M. Ibrahim, Y. Rajamanickam, and K. Mohamad, "Implementation of wavelet packet transform and non linear analysis for emotion classification in stroke patient using brain signals," *Biomedical Signal Processing and Control*, vol. 36, pp. 102–112, 2017.

[4] M. Kotti and F. Paterno, "Speaker-independent emotion recognition exploiting a psychologically-inspired binary cascade classification schema," *International Journal of Speech Technology*, vol. 15, no. 2, pp. 131–150, 2012.

[5] M. Kumar, P. Mukherjee, K. Verma, S. Verma, and D. B. Rawat, "Improved Deep Convolutional Neural Network based Malicious Node Detection and Energy-Efficient Data Transmission in Wireless Sensor Networks," *IEEE Transactions on Network Science and Engineering*, pp. 1–1, 2021.

[6] S. Vidya Priya Darcini, D. P. Isravel, and S. Silas, "A comprehensive review on the emerging iot-cloud based technologies for smart healthcare," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 606–611, Coimbatore, India, 2020.

[7] M. L. Johns, "HIPAA privacy and security: a practical course of action," *Topics in Health Information Management*, vol. 22, no. 4, pp. 40–48, 2002.

[8] T.-T. Kuo, H.-E. Kim, and L. Ohno-Machado, "Blockchain distributed ledger technologies for biomedical and health care applications," *Journal of the American Medical Informatics Association*, vol. 24, no. 6, pp. 1211–1220, 2017.

[9] J. W. Picone, "Signal modeling techniques in speech recognition," *Proceedings of the IEEE*, vol. 81, no. 9, pp. 1215–1247, 1993.

[10] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[11] D. O. Shaughnessy, *Speech Communication: Human and Machine*, University Press, India, 2001.

[12] S. Upadhyay, S. K. Sharma, and A. Upadhyay, "Speaker identification and verification using different model for text

dependent," *International Journal of Applied Engineering Research*, vol. 12, no. 8, pp. 1633–1638, 2017.

[13] H. Hermansky, B. A. Hanson, and H. Wakita, "Perceptually based processing in automatic speech recognition," in *Proc. IEEE Int. Conf. on Acoustic, speech, and Signal Processing*, pp. 1971–1974, Tokyo, Japan, 1986.

[14] L. Roderer, *The Physics and Psychophysics of Music: An Introduction*, Springer Verlag, New York, 1995.

[15] J. Gałka and M. Ziółko, "Wavelet speech feature extraction using best basis algorithm," in *Advances in Nonlinear Speech Processing*, pp. 128–135, Springer-Verlag, Berlin Heidelberg, 2010.

[16] P. A. Catherwood, D. Steele, M. Little, S. Mccomb, and J. Mclaughlin, "A community-based IoT personalized wireless healthcare solution trial," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1–13, 2018.

[17] J. Lloret, J. Tomas, A. Canovas, and L. Parra, "An integrated IoT architecture for smart metering," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 50–57, 2016.

[18] A. Kumar, "A cloud-based buyer-seller watermarking protocol (CB-BSWP) using semi-trusted third party for copy deterrence and privacy preserving," *Multimedia Tools and Applications*, pp. 1–32, 2022.

[19] A. Kumar, "Design of secure image fusion technique using cloud for privacy-preserving and copyright protection," *International Journal of Cloud Applications and Computing*, vol. 9, no. 3, pp. 22–36, 2019.

[20] A. Kumar, "A review on implementation of digital image watermarking techniques using LSB and DWT," in *the Third International Conference on Information and Communication Technology for Sustainable Development (ICT4SD 2018)*, Hotel Vivanta by Taj, Goa, India, August 2018.

[21] A. Kumar, Z. J. Zhang, and H. Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, Article ID 204, 2020.

[22] M. Devi, S. Singh, S. Tiwari, S. Chandra Patel, and M. T. Ayana, "A survey of soft computing approaches in biomedical imaging," *Engineering*, vol. 2021, article 1563844, 2021.

[23] S. Tiwari, "A variational framework for low-dose sinogram restoration," *International Journal of Biomedical Engineering and Technology*, vol. 24, no. 4, pp. 356–367, 2017.

[24] V. P. Singh, R. Srivastava, Y. Pathak, S. Tiwari, and K. Kaur, "Content-based image retrieval based on supervised learning and statistical-based moments," *Modern Physics Letters B*, vol. 33, no. 19, p. 1950213, 2019.

[25] S. Tiwari and R. Srivastava, "An OSEM-based hybrid-cascaded framework for PET/SPECT image reconstruction," *International Journal of Biomedical Engineering and Technology*, vol. 18, no. 4, pp. 310–332, 2015.

[26] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.

[27] J. C. Castillo, Á. Castro-González, A. Fernández-Caballero et al., "Software architecture for smart emotion recognition and regulation of the ageing adult," *Cognitive Computation*, vol. 8, no. 2, pp. 357–367, 2016.

[28] M. Vyas, "A Gaussian mixture model based speech recognition system using matlab," *Signal & Image Processing: An International Journal*, vol. 4, no. 4, pp. 109–118, 2013.

[29] R. Sarikaya and J. H. L. Hansen, "High resolution speech feature parametrization for monophone-based stressed speech recognition," *IEEE Signal Processing Letters*, vol. 7, no. 7, pp. 182–185, 2000.

[30] F. Alias, J. C. Socoro, and X. Sevillano, "A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds," *Journal of Applied Science*, vol. 6, no. 5, pp. 143-144, 2016.

[31] I. Mporas, T. Ganchev, M. Siafarikas, and N. Fakotakis, "Comparison of speech features on the speech recognition task," *Journal of Computer Science*, vol. 3, no. 8, pp. 608–616, 2007.

[32] H. A. Murthy, F. Beufays, L. P. Heck, and M. Weintraub, "Robust text-independent speaker identification over telephone channels," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 5, pp. 554–568, 1999.

[33] B. McFee, C. Raffel, D. Liang et al., "Librosa: audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, pp. 8–25, Austin, Texas, 2015.

[34] L. Y. Mano, B. S. Faiçal, L. H. Nakamura et al., "Exploiting IoT technologies for enhancing health smart homes through patient identification and emotion recognition," *Computer Communications*, vol. 89-90, pp. 178–190, 2016.

[35] M. M. Dhanvijay and S. C. Patil, "Internet of things: a survey of enabling technologies in healthcare and its applications," *Computer Networks*, vol. 153, pp. 113–131, 2019.

[36] M. Cobos, J. Perez-Solano, and L. Berger, "Acoustic-based technologies for ambient assisted living," in *Introduction to Smart eHealth and eCare Technologies*, pp. 159–180, Taylor & Francis Group, Boca Raton, FL, USA, 2016.

[37] A. Ahad, M. Tahir, and K. A. Yau, "5g-based smart healthcare network: architecture, taxonomy, challenges and future research directions," *IEEE Access*, vol. 7, pp. 100747–100762, 2019.

[38] N. G. Peters, "Normalization of ambient higher order ambisonic audio data," 2018, US Patent 9, 875, 745, 2018.

[39] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1533–1545, 2014.

[40] I. Rebai, Y. Ben Ayed, W. Mahdi, and J.-P. Lorre, "Improving speech recognition using data augmentation and acoustic model fusion," *Procedia Computer Science*, vol. 112, pp. 316–322, 2017.

[41] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, 1980.

[42] F. Zheng, G. Zhang, and Z. Song, "Comparison of different implementations of MFCC," *Journal of Computer Science and Technology*, vol. 16, no. 6, pp. 582–589, 2001.

[43] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, 1995.