IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 June 2022, accepted 30 July 2022, date of publication 10 August 2022, date of current version 9 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3197878

SURVEY

Intelligent and Interactive Healthcare System (I^2HS) Using Machine Learning

RASHIKA RAINA[1] AND RAKESH KUMAR JHA[2], (Senior Member, IEEE)
[1]School of Electronics and Communication Engineering, Shri Mata Vaishno Devi University, Jammu and Kashmir 182320, India
[2]Department of Electronics and Communication Engineering, Indian Institute of Information Technology Design and Manufacturing, Jabalpur, Madhya Pradesh 482005, India

Corresponding author: Rakesh Kumar Jha (jharakesh.45@gmail.com)

ABSTRACT There has been a gigantic stir in the world's healthcare sector for the past couple of years with the advent of the Covid-19 pandemic. The healthcare system has suffered a major setback and, with the lack of doctors, nurses, and healthcare facilities the need for an intelligent healthcare system has come to the fore more than ever before. Smart healthcare technologies and AI/ML algorithms provide encouraging and favorable solutions to the healthcare sector's challenges. An Intelligent Human-Machine Interactive system is the need of the hour. This paper proposes a novel architecture for an Intelligent and Interactive Healthcare System that incorporates edge/fog/cloud computing techniques and focuses on Speech Recognition and its extensive application in an interactive system. The focal reason for using speech in the healthcare sector is that it is easily available and can easily predict any physical or psychological discomfort. Simply put, human speech is the most natural form of communication. The Hidden Markov Model is applied to process the proposed approach as using the probabilistic approach is more realistic for prediction purposes. Ongoing projects and directions for future work along with challenges/issues are also addressed.

INDEX TERMS Interactive system, human-machine interaction, smart healthcare, artificial intelligence, machine learning, hidden Markov model, C-RAN.

I. INTRODUCTION

With an increase in the ratio between ill people and the number of healthcare facilities, it has become a necessity to develop a smart healthcare system that will counter the needs of the people and increased demand [1]. The medical expenses and healthcare costs keep on soaring high, making it difficult for an average citizen to cover such expenses as the necessity of healthcare facilities has also increased in line with the growing demand. Therefore, through the use of various Information and Communication technologies and AI, we can deliver technical modalities at reasonable prices without compromising the quality of care. With many patients and people falling ill we have lots of data generated from medical records, which can be processed by ML and DL models to give better results [2]. Thus, developing such smart healthcare systems helps in not only looking out for the patient and easing their pain, but also giving medications, or detecting the disease.

AI/ML has several applications apart from healthcare such as Software, Spam Detection, Stock trading, Robotics, Advertising, Retail and E-commerce, IoT, Gaming Analytics, Voice Recognition, and many more. Voice recognition or speech recognition is one of the major applications of AI. Amazon's Alexa, Microsoft's Cortana, and Apple's Siri are all AI-driven, are very popular and have changed the way of interaction among various environments [3]. The report on Voice Assistant technology, released in 2018, reports that 27% of the world's population online make use of voice search on their mobile phones [3]. These voice assistants are trained with lots of data. Tons of data varying from dialects, languages, and even accents are fed to train the model for it to be efficient and reliable. Human-Machine Interaction (HMI) is the interaction or communication between machines

The associate editor coordinating the review of this manuscript and approving it for publication was Maurizio Tucci.

and humans via an interface also called the user interface. Nowadays, Human-Machine Interaction has led way for new advancements [3], [4].

A. CONTRIBUTIONS

The major contributions of this paper are addressed below.

1) It addresses the advancements in the healthcare sector using AI/ML Algorithms. It presents an extensive analysis of the work done so far in the related field and discusses the need to develop such a smart healthcare system.
2) A novel architecture is presented using the concept of Human-Machine interaction, Virtual Voice assistants (using speech recognition) integrated with the concept of edge/fog/cloud computing. The concept of new technology like C-RAN has also been incorporated into our architecture for energy efficiency.
3) Comprehensive review of different algorithms used in the implementation of speech recognition along with several resource allocations, power optimization, and energy efficiency techniques is presented. Security issues and their countermeasures are also discussed.
4) Mathematical Analysis of the proposed system is also presented.

B. RESEARCH GAP

According to the review of literature and research available in the field of AI and its applications in the healthcare sector, a lot of work has been done to cater to the growing demands. To count a few such as Voice Pathology detection [6], Covid-19 detection using voice signals [7], heart disease diagnosis, Diabetes diagnosis [1], and several other health monitoring systems in IoT using Wearable devices and smartphones. A large number of health monitoring systems based on wearable technology are in development [1], [3].

Although such wearable health monitoring systems are good, but with the increasing number of communicable diseases, it is not a good pick for patients in dire need of some healthcare facility, as it will only increase the risk of them getting affected. These health monitoring prototypes developed using wearable devices are not adequately manufactured for application and are not fully fabricated for real-time.

Not enough data is available for the training of models for efficient and reliable results. Furthermore, many difficulties arise because of latency problems, as the responses are not in real-time. A lot of work is required to personalize the user's needs, which includes integrating various sensors, actuators, and user interfaces [2], [3]. Cloud servers are burdened with ensuring real-time communication and providing real-time healthcare solutions.

C. RESEARCH MOTIVATION

The smart healthcare sector is currently growing at a rapid pace. The need for intelligent healthcare has emerged more than ever, particularly in the period after the outbreak of COVID-19. A lot of wearable health monitoring systems have been developed before but they are not beneficial in today's times because of the increasing risk of communicable diseases. Thus, speech recognition [10], [11], [12], [13], another huge application of AI can be incorporated into developing an interactive healthcare system. Thanks to voice assistants, people have now explored new benefits and experiences for communication with the outer world. The whole way of communication is changing, human-to-human and even human-to-machine [4], [5]. Voice assistants are becoming increasingly important in supporting persons with impairments, especially now, as the technology becomes more accessible. This has motivated us to use speech recognition in our research to make our health care system interactive.

Cloud servers have a huge burden on them for ensuring all the real-time responses as fast as possible, which motivated us to include edge/fog/cloud computing into our architecture, which enhanced the proposed interactive system. Edge and fog computing acts pivotal in receding the burden from the cloud servers and improving the healthcare system's response time. The primary idea of fog computing is to shift local data centers to fog nodes or servers, which allows for substantially faster data transport and response times[1], [3].

We included C-RAN in our novel architecture as it has several advantages like an increase in throughput, decrease in delay, latency, and many more. The C-RAN architecture, which includes a number of other co-located BBUs, aids in network maintenance immensely [14]. Moreover, its capacity increase but in case of any failure also it reconfigures automatically minimizing the need for any sort of human intervention. Only a few BBU pool locations are chosen in case of any updates.

D. ORGANIZATION

Section I gives an introduction to the world of smart healthcare systems along with the growing application of speech recognition in today's time. It also describes the major contributions of this paper along with the Research Gap and the right motivation to carry this work forward.

More precisely, in Section II, the context of speech recognition is presented with its many applications in the field of the healthcare sector. A table of the related work is presented as well. The extraction of features is a vital step in speech processing. MFCC is one of the most popular feature extraction techniques [10] and is explained in detail. Several applications of Speech Recognition systems in healthcare like clinical documentation, disease diagnosis, etc. are also discussed.

Section III gives an overview of the algorithms applied for speech recognition.

Section IV consists of the various techniques, regarding resource allocation, energy efficiency, power optimization, and more. Section V comprises Security issues and

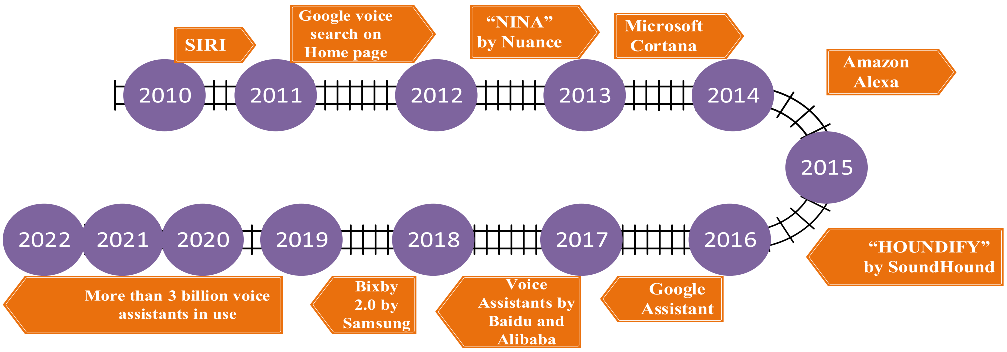

FIGURE 1. Voice assistant revolution.

countermeasures. Section VI explains the proposed novel architecture in detail. Mathematical analysis is presented in Section VII. A table presenting the ongoing projects is presented in Appendix.

II. BACKGROUND OF SPEECH RECOGNITION

Speech is an integral part of the way communication takes place among people and hence Speech Recognition becomes an important domain that needs to be well addressed and worked upon. There are several speech-related applications, to name a few: automatic speech recognition, speech enhancement, speaker identification, emotional speech recognition, etc. Speech signals provide a lot of information like content or the message intended by the speaker, the speaker's identity, the speaker's emotional state, accent, gender of the speaker, and the spoken language.

A. FEATURE EXTRACTION

Feature Extraction primarily involves the analysis of speech signals. Feature Extraction techniques can be broadly classified into spectral analysis and temporal analysis. Spectral analysis techniques include Critical Bank Filter Bank Analysis, Cepstral Analysis, Mel Cepstrum Analysis, Linear Predictive Coding (LPC) Analysis, perceptually based Linear Predictive Analysis whereas Temporal analysis techniques include power estimation, fundamental frequency estimation, and a few others. Commonly used features are MFCC, LDA (Linear Discriminate Analysis), HLDA Transform, STFT, MLLT, LPS (Log Power Spectral), BFCC (Bark-Frequency Cepstrum Coefficients), MLLT (Maximum Likelihood Linear Transform). MFCC (Mel Frequency Cepstrum Coefficients) is the most popular feature technique in extracting speech signals [10]. MFCC performs a series of steps in feature extraction.

1) ANALOG TO DIGITAL CONVERSION

An audio signal is taken as input and sampled into digital form with a sampling frequency in this first step.

2) PRE-EMPHASIS

Pre-emphasis focuses on filtering the signals of a higher frequency range. An audio signal is a continuous-time signal that varies in the frequency range, certain voice segments have a higher frequency and thus higher energy than the lower frequencies [24]. This step is implemented by a high pass filter of the first order. The widely used transfer function for a pre-emphasis filter is given by:

$$\mathrm{Y(h)} = 1 - \mathrm{xh}^{-1} \tag{1}$$

where x handles the slope of the filter and generally varies from 0.4 to 1.0.

3) FRAMING AND WINDOWING

In this step, audio signal waveforms are sliced into frames. A sliced frame is given as

$$\mathrm{r[k]} = \mathrm{x[k]c[k]} \tag{2}$$

where r[k] denotes the sliced frame, x[k] refers to the window applied and c[k] is the original audio signal clip. The windows can be Rectangular, Hamming, or Hanning Window. The audio signal is sliced when the amplitude of the signal decreases to the edge of the frame. So, Hamming and Hanning windowing techniques are used to chop the signals.

The windowing techniques are given by the equation

$$\mathrm{x[k]} = (1 - \varsigma) - \varsigma \cos\left(\frac{2\pi k}{M - 1}\right) \tag{3}$$

where M is the width of the window and ς differs for Hamming and Hanning windows. $\varsigma = 0.46164$ and 0.5 for Hamming and Hanning windows respectively. Every frame is usually 25 ms [24].

4) DISCRETE FOURIER TRANSFORM (DFT)

The time-domain signal is converted to a frequency-domain signal by applying the DFT which is given by

$$S_k = \sum_{n=0}^{J-1} \left(d_n e^{\frac{i2\pi}{J} kn}\right) \tag{4}$$

where $\mathrm{S_k}$ is the sequence of complex numbers for given $\mathrm{d_n}$, $0 \leq \mathrm{k} \leq \mathrm{J} - 1$ and J refers to the number of points used to calculate DFT [24].

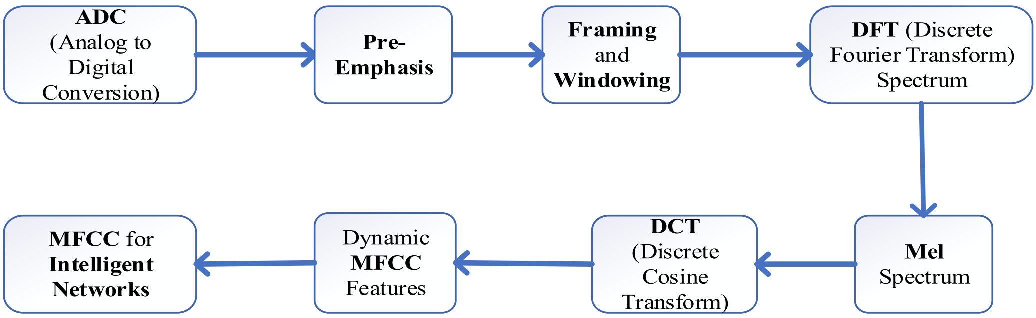

FIGURE 2. Next-generation MFCC block diagram.

5) MEL – SPECTRUM

The signal goes via the mel-filter bank, which is a collection of bandpass filters. Mel- Spectrum is obtained as the output of the mel-filter bank. Mel scale is a logarithmic scale. Mel-frequency is given as

$$b_{\mathrm{Mel}} = 2595 \log(1 + \frac{b}{500}) \tag{5}$$

Here b stands for the perceived frequency and b_{Mel} denotes the Mel-frequency. Filter banks are based on both time-domain and frequency-domain signals, whereas MFCC computations filter banks in general operate within the frequency domain. The output of a mel-filter bank is a power spectrum [24].

6) DISCRETE COSINE TRANSFORM

DCT transforms the signal into frequency components. It generates a set of cepstral coefficients when applied to the DCT matching-frequency coefficient. The first few MFCC coefficients extract the majority of the signal information. Minimizing the high-order components of DCT or ignoring the first few MFCC coefficients will make the system stronger. Therefore, MFCC can be computed by the following equation

$$q(k) = \sum\nolimits_{l=0}^{L-1} \log_{10}(j(l)) \cos(\frac{\pi k (l - 0.5)}{L}) \tag{6}$$

where q is the number of MFCC coefficients, q(k) are the cepstral coefficients and k varies from 0,1,2…., l-1 [24].

7) DYNAMIC MFCC FEATURES

Since cepstral coefficients contain only information about a given frame, they are called static properties. More information about the second-order can be obtained by calculating the first- and second-order derivatives of the cepstral coefficients. The first and second-order derivatives are called delta coefficients and delta-delta coefficients, respectively. Delta coefficients provide information about speech rate, while delta-delta coefficients inform us about the acceleration of speech [24].

8) MFCC FOR INTELLIGENT NETWORKS

For the training and processing of Intelligent Networks MFCC is fed to the training module. In the case of speech interaction systems, these MFCCs are fed to several DNN acoustic models. MFCC features are widely used in several speech interaction systems, by extracting audio features from audio samples or tracing people's lip movements while speaking [9].

B. APPLICATIONS OF SPEECH RECOGNITION IN HEALTHCARE

Speech recognition has a multitude of purposes in healthcare ranging from clinical documentation to social robots for the elderly and home care to diagnosis of diseases.

1) CLINICAL DOCUMENTATION

A speech interface that makes use of ASR is used to document clinical information. Clinical documentation takes up a lot of time for clinicians when compared to other forms of direct patient care. Nurses and practitioners use this interface for transcribing speech to text and text to speech which is more efficient in processing transcripts. It reduces the amount of time taken by the doctors to diagnose, using the readily available information. This speech interface makes use of voice-based assistants to keep track of Electronic Medical Records and give relevant information on demand [8].

2) HEARING AND SPEAKING IMPAIRMENTS

Speech can provide a lot of information about the emotional state, gender, age, and others but it also provides information about voice disorders as well. Speech recognition helps in the diagnosis of voice disorders like voice pathology detection, Dysarthria, stroke survivors, Alzheimer's, Parkinson's, Amyotrophic Lateral Sclerosis, etc. These disorders produce difficulty in speaking, and weakness as well. Hearing and Speaking impairments can be fixed using ASR. Individuals with various speech, voice, or language disorders can be assisted with speech technology for effective communication [8].

FIGURE 3. **Applications of Speech recognition in healthcare.**

3) SOCIAL ROBOTS

Various social robots have been developed to provide care to the elderly and at home, which make use of several human-machine or human-robot interactions like Speech Interaction, Gesture Control, and Touch screen user interface to name a few. The platform for such interactive systems is designed using MySQL database and using speech to text functionality. Data sharing protocols provided in the system are REST API and MQTT [1].

4) HEALTHCARE SYSTEMS

The use of AI in health services has grown dramatically in recent years, with several models being developed to detect various diseases or disorders like Covid-19 detection, Voice pathology detection systems to name some. These services are not useful for people who don't know much about computers, in other words, computer illiterate and visually impaired people.

The authors of [1] have reviewed several healthcare systems that use AI/ML algorithms to predict several diseases and develop some interactive healthcare systems. Several Covid-19 detection frameworks developed made use of CNN, SVM, ResNet-50, Random Forest, Naïve Bayes, K-means Algorithm, etc.

Ultra-Low Latency-based Healthcare systems developed and reviewed by the authors of [1] make use of software integration architectures and device layer/fog layer/cloud layer because these healthcare networks demand data or information that needs to be transmitted quickly and efficiently. IoMT networks have utilized similar software integrated architectures. For an ultra-low latency-based network several frameworks have been developed that make use of multiple layers for processing and transmission of data and are flexible. Incorporating multiple layers like edge/fog/cloud in the architectures presents unlimited storage space and computational power. Cloud provides unlimited storage for information and analytics but it does have aftermath to it. Apart from the cloud, other techniques like Hadoop Map Reduce technique are also used to process a huge amount of data simultaneously.

5) DIAGNOSIS OF PSYCHOLOGICAL DISORDERS

The most fundamental and widely used mode of communication is speech. Speech signals provide a lot of details about the content spoken and lots of speaker-related information including the frame of mind of the speaker. The speaker's emotional state can help identify, if the person is suffering from any psychological disorder. Speech processing can be used as an effective biomarker for diagnosing any mental or psychological disorder like anxiety, stress, depression, or any suicidal behaviour [8].

C. RELATED WORK

A table comprising related work classified on the basis of their contribution is presented below.

Authors of [1] have traveled through many cutting-edge systems, describing key areas of smart healthcare systems, including wearable and smartphone-based health monitoring systems, and ML techniques in predictive analysis for different diseases. Using several algorithms, many models have been developed for disease detection like heart disease and diabetes, Smart homes, and social robots incorporate eco-friendly living and software integration architectures like edge/fog/cloud computing. They discussed recent technological advancements, challenges/issues, and prospects of such healthcare systems are discussed.

Afterward, the authors of [4] explain how Alexa works for home automation. They explained that the affinity between the IoT system and the virtual voice assistant is generally set by the use of client-server frameworks such as REST. Alexa's voice interaction model comprises four components: wake word, starting phrase, invocation name, and utterance of intent with slots. The user's voice stream performs speech recognition i.e., speech to text and text to speech, and then transformed into JSON format which is then sent to the server using REST Application Peripheral Interface (API).

In [6] the authors have presented a dossier of a VPD system by converging the IoT with AI techniques. A bimodal input system is developed that takes EGG and voice signals as input. Spectrograms that are taken from EGG and voice signals are inserted in the pre-trained CNN. Characteristics collected from CNN are combined and processed using bi-LSTM. Three distinct pre-trained CNN models: Xception, ResNet50, and MobileNet were examined. The authors made use of the publicly available voice dataset Saarbruecken. Experimental results have shown an accuracy of 95.65% through the proposed system.

In [7] Covid-19's presence was discovered by speech and voice using AI algorithms. The authors presented a relative analysis of the realization of the proposed system using main ML techniques grouped into Bayes, Functions, Lazy,

TABLE 1. Work classified on the basis of core contribution.

S. No	Core Contribution	Inference	References
1.	Smart Healthcare	i. Explores diverse clinical systems using ML for wearable and smartphone-based health monitoring systems. ii. Cognitive Healthcare system for Pathology Detection and monitoring. iii. Ubiquitous Healthcare framework.	[1], [2], [3]
2.	Human Device Interaction	i. Virtual voice-based assistant making use of Human Device Interaction. ii. Social human-robot interaction.	[4], [5]
3.	Speech technology and its applications in healthcare	Several healthcare case studies make use of speech processing for disease diagnosis.	[6], [7], [8]
4.	Speech Recognition	i. Survey of Speech Recognition using various DL frameworks. ii. ASR for a far-field scenario. iii. Speech recognition systems for suppressing noisy properties. iv. Investigates techniques for covering the issue of Vocabulary size.	[10], [11], [12], [13]
5.	C-RAN	A mobile network architecture that addresses a slew of issues that operators face as they try to satisfy the demands of growing end-users.	[14]

Meta, Rules, and Trees. Performance metrics include recall, Receiver Operating Characteristic (ROC), precision, specificity, F1 -score, and accuracy. The suggested model does have some drawbacks as the dataset is unbalanced and data collection is still ongoing. More data will provide a more comprehensive analysis which makes the model more robust and reliable. According to the findings, SVM had the highest accuracy of 97 percent in differentiating between a healthy and a disordered voice.

Furthermore in [8] authors have presented a comprehensive review of the several research frameworks from various speech-associated domains – ASR, TTS/STT also known as speech synthesis, speech biomarkers, and remote monitoring systems. Various applications and prospects of speech processing technology in the healthcare sector: fixing speech and hearing impairments, clinical documentation, and social robots for home and elderly care are also discussed. Challenges related to this field were also addressed: adversarial attacks that degrade the performance of the ASR systems, dearth of data, interoperability of the data generated by various segments within the healthcare system, and privacy concerns.

The authors of [9] surveyed the literature present at the time and presented an overview of how smart city technologies and techniques can be implemented to design a smart healthcare system and how they support various ML techniques. They studied techniques that are concerned with acquiring data from ambient sensors and mobile for health monitoring. The continuous collection of accumulating sensor data in day-to-day life interprets the behavior changes that are in sync with physical and cognitive health and are too subtle to be noticed without Information and Communication Technologies (ICT).

In [10] authors present a systematic review of DNN applied to speech recognition. A detailed description of the kinds of ML: Supervised, Unsupervised, Semi-Supervised, Reinforcement, and DL is also presented. DL is popular in the domain of speech recognition because of the increase in the number of hidden layers. The authors extracted information from 174 papers and presented this review. Several applications of speech recognition like ASR, Emotion cue-based speech recognition, Emotion recognition, Automatic health recognition, Automatic gender recognition, Age recognition, Accent recognition, and Language recognition. The authors concluded that MFCC is the most popular feature extraction technique followed by LDA, HLDA transform, and STFT.

In [11] the authors focus on ASR for far-field regions and describe its specific challenges. Authors measure Speech Recognition performance accuracy by Word Error Rate (WER). Speech enhancement can be done by Dereverberation and source separation. Dereverberation can be achieved by removing the late reverberation component, and source separation can be done by disintegrating speech into its components by beamforming and protruding the input signal into the 1-D subspace. The whole front-end system consists of Dereverberation, mask estimation, and beamforming. For a high ASR performance, we need to employ a more better and powerful speech enhancement. Multi condition Training Data, HMM-State alignments, Modification of the back end of ASR

TABLE 2. Related work.

Ref No.	Objective	Conclusion	Outcome	Advantages	Disadvantages
[1]	To explore a sophisticated healthcare system that highlights crucial areas of ML for wearable and smartphone-based health monitoring, assessment, and analysis.	i) Techniques for disease detection: Naive Bayes, Neural network, KNN, SVM, Decision tree, Random Forest, CNN, DLMNN, MDCNN, Logistic Regression, multilayer perceptron, etc. ii) Social Robots use Speech Interaction and Vision, Gesture Control, Touch Screen UI, Alarms, etc.	The advantages, shortcomings, and challenges of these smart frameworks are discussed.	i) ML Algorithms give promising solutions for disease detection. ii) Implementation of social robots for care was discussed.	i) Heterogeneous data integration from different sensors. ii) Concerns regarding security and privacy in wearable and IoT environments. iii) Loss of power results in loss of data.
[4]	To develop a virtual voice assistant that performs task automation using natural Human Device Interaction (HDI).	i) Affinity within an IoT system and a virtual voice assistant is guaranteed through Client-Server REST API architecture. ii) Speech to text and vice versa is converted to JSON format for communication with REST API.	AVS (Alexa Voice Service) performs ASR using NLU.	Faster, easier way of communication that minimizes human effort.	i) Security, Privacy, and accessibility are issues of concern. ii) Accessibility issues for people with disability or elderly people.
[6]	To realize a VPD system based on bimodal input that consists of EGG and voice signals fed into a pre-trained CNN.	i) Voice and EGG signals are inputs and sent to Edge Computing to capture spectrograms, which are then forwarded to cloud computing containing AI / ML / DL storage and servers. ii) The BiLSTM model connects and presents the features of the two approaches.	Results indicate that the suggested method achieves greater than 95% accuracy along with recall and precision.	i) Single input fares worse than bimodal input. ii) Decision is sent to the client and vice versa through 5G which is faster.	More data can be accumulated for training purposes.
[7]	To design a methodology proficient enough to catch Covid-19 disorder early using voice samples and to be used as a pre-screening test.	i) Techniques used are Bayes, SVM, SGD, KNN, LWL, Adaboost, Bagging, One-R, Decision Table, Decision Tree, and REPTree. ii) Performance metrics: F1-score, recall, precision, and ROC, accuracy, specificity.	Outcomes have shown that SVM achieved the best accuracy of 97% to detect Covid-19.	i) Early detection of Covid-19. ii) Helpful as a fast Covid-19 pre-screening test.	Data samples collected are fewer.
[8]	This paper covers various speech technology frameworks and their applications.	i) Focuses on vast applications of speech technology that revolutionized the healthcare sector. ii) Various Interoperability and cultural or language hurdles to name a few were discussed.	Overviews the conventional approaches and discusses open challenges.	Discusses applications like speech and hearing impairments, psychological disorders, and interactive interface for documentation, and care.	i) Less amount of data. ii) Interoperability issues. iii) Cultural and Language barriers.
[9]	Examines approaches, frameworks, and techniques that assess the data acquired from several ambient and mobile sensors for health assessment.	Techniques that are concerned with acquiring data from ambient sensors and mobile for health monitoring are studied.	Surveys present literature and introduces authentic research to present an outlook on how smart city technology and ML can be incorporated into healthcare.	i) Assists with the development of integrated solutions that can assist weak and non-tech adept patients. ii) Health assessment using ambient sensors.	i) Privacy and Security concerns. ii) Connection with other smart city services.
[10]	This paper presents an evaluative review of speech recognition using various DL architectures.	Background of Speech Recognition along with the related work and applications are discussed. DL architectures like CNN and RNN were discussed.	MFCC is the most widely used feature extraction. WER is the widely used parameter to govern system's efficiency.	Major Applications include Speaker Identification, speaker emotion recognition, speech enhancement, speech recognition, and speech transcription.	Cultural and Language barriers pose a crucial challenge for speech recognition and its extensive applications.
[11]	Extends the end-to-end method to	i) Discusses the algorithms for accurate speech recognition and	Emphasizes multichannel speech enhancement.	i) Effectiveness was enhanced by coupling	i) VAD and speaker diarization pose

TABLE 2. *(Continued.)* **Related work.**

	ASR to the far-field scenario.	extends it to the far-field scenario. ii) For better Speech Enhancement, Dereverberation is cascaded by denoising and followed by source separation. iii) Joint modeling of back-end ASR with Front-end speech enhancement.	Minimizing the size of neural networks makes the system work more robustly.	strong signal processing with DL. ii) Single neural network backpropagation performs both speech enhancement and recognition.	some challenges in far-field ASR. ii) Low latency online processing is challenging.
[12]	Proposes a DL solution with the use of a gating layer to suppress the effect of noisy properties.	i) The proposed system is a hybrid HMM-GNN model. ii) The gating Layer is optimal for managing with multimodal DL. iii) System's capability can be improved by summing gating layers at numerous levels.	Good performance can be attained by adding 25% more parameters.	i) Gating Layer learns to filter out noisy and inconsistent data. ii) Allows for realistic assimilation of AV-ASR into practical systems.	i) Only one gating layer is evaluated. Multiple gating layers can improve the capability of the system. ii) Models trained with only clean speech were considered.
[13]	To explore and investigate various techniques that cover the issue of vocabulary size and using Finnish and Estonian	i) For a broad glossary with a lump of words, n-gram models, are preferred over the normal word models. ii) In Finnish, the finest outcomes were attained from NNLM with a relatively short subword vocabulary with full softmax output. iii)In Estonian, the leading conclusions were obtained using 403ks subword vocabulary.	Experimental results have shown that Finnish and Estonian tasks were implemented (better than the previous work done) with a WER of 48.4% and 52.7% respectively.	Processes large vocabulary more efficiently.	i) The memory consumption grows as the number of threads increases. ii) Training takes more time.
[14]	To pool the Baseband Units from several base stations into a centralized BBU Pool.	i) C is regarded as Centralized, Clean, Cooperative Radio, Cloud, or Collaborative. ii) Consists of RRH and BBU, RRH is placed with an antenna and several BBUs are pooled together.	Pooling BBU reduces cost, energy, delays, and latency and improves the throughput, capacity, and statistical multiplexing gain.	Improves energy efficiency.	i) Imposes large overhead on the optical links between RRH and BBU Pool. ii) Requires high bandwidth.
[2]	To present a cognitive and smart healthcare system for monitoring and pathology detection.	i) Uses IoT and cloud technologies. ii) Uses smart sensors for data transmission and DL for intelligent decision-making.	The presented model achieves better accuracy than state-of-art models.	Accurate, timely, and cost-effective healthcare services.	i)Wearable sensors are not a good pick for communicable diseases. ii)Data integration in heterogeneous sensors is a challenge.
[3]	To present a Ubiquitous healthcare framework.	i) Uses Edge computing, Big Data, DL, HPC, and IoT. ii) 3 components: DLNTAP, DLNTAC, and FCA. iii) 4 layers: Mobile, Cloudlets, Network, and Cloud.	Results show that the latency obtained is 50% less than the traditional networks that are cloud-based.	i) Ubiquitous personalized and preventive healthcare. ii) Extensively addresses several network-related issues.	Security, Reliability, and Privacy need to be addressed.
[5]	To encourage patient participation and performance during therapy.	i) As a partner agent, SAR was included in the neurorehabilitation program. ii) Data acquisition, processing, on-line feedback, visualization, and data management comprise the system architecture.	Patients' thoracic and cervical postures improved by 18.44% and 32.23%, respectively, with SAR support.	After the robot's correction, patients were more focused on keeping a healthy posture.	Privacy and security concerns.

to the Front-end Speech Enhancement, and joint training are the realistic considerations for far-field ASR. Training of both ends jointly can be done by making use of only far-field speech and linked word transcripts which optimizes the entire system. Far-field ASR has several challenges like VAD (Voice Activity Detection), Speaker Diarization, online low latency processing, spontaneous speech conversations, signal extraction improvement using syntactic and semantic context information, and multimodality.

Furthermore, in [12] the authors have proposed a gating neural network (GNN) and have used an HMM model by creating a GNN-HMM hybrid model for Audiovisual Speech Recognition. All of the input nodes are connected to the gating layer unit. The gating layer's output is multiplied by the lower nodes' output. The gating layer has a sigmoid function as an activation function whose output varies from 0 to 1. Lower layer output results in 0, if the output of the gating layer is 0 and hence the output will not be propagated to the upper layers and vice-versa. Weight is calculated by backpropagation. Experimental results show that by adding 25% more parameters i.e., one gating layer, the proposed system can achieve good performance.

The authors of [13] explored several techniques that covered the issue of the size of vocabulary by bringing down the vocabulary size and its processing. Data sets were created for Finnish and Estonian conversations. A random sampling of a subset of each data set and weighing parameter updates are the approaches followed for neural network training. For both the tasks i.e., for both Finnish and Estonian the results have shown that the performance was great with a WER of 48.4% and 52.7% respectively. For both the tasks the best results were obtained from an interpolation of the NNLM and n-gram model.

In [14] the authors presented a comprehensive overview of a mobile network architecture known as C-RAN where C stands for Clean, Cloud, Centralized, Cooperative Radio, or Collaborative. In this architecture, the baseband units are pooled that reduces OPEX, latency, and heat generated. This technology has the capability to adapt to nonuniform traffic, minimizing the cost and energy savings, and also upgrades the networks and their maintenance. However, this technique causes considerable transport resources between BBU and RRH.

Authors of [2] have proposed a cognitive healthcare structure for classification and pathology detection that makes use of IoT and cloud technologies. Data acquisition is done through various IoT sensors. Short-range communication protocols like Bluetooth, RFID, etc. comprise the LAN. The Cloud layer consists of a cognitive engine, DL server, and a cloud manager. Raw time-domain EEG signals were fed as input to the CNN model. The model detects pathology and sends the same result to the cognitive system.

The authors of [3] presented a Ubiquitous healthcare structure that has three components and four layers. The components include DLNTAP, DLNTC, and FCA. DLNTAP uses DL, big data, and high-performance computing. DLNTC deals with the classification of application protocols of the traffic flows. FCA Component grouped the data to determine the data from various sources that come from same application protocols. The layers include Mobile, Cloudlets, Network, and cloud. The numerous network-related difficulties in next-generation healthcare systems are widely discussed. When compared to typical cloud-based connected healthcare systems, the suggested model achieved a 50% drop in latency.

In [5] the execution and analysis of a Human-Robot Interface is presented. It blends relevant cognitive and physiological aspects with the purpose of assessing user performance in gait rehabilitation using the Lokomat. Also, creation of SAR system was presented.

III. ALGORITHMS FOR SPEECH RECOGNITION

Speech recognition also known as ASR, computer speech recognition, or STT/TTS, is the ability to start a program for converting human speech to a written format. Speech Recognition can be achieved by various algorithms like Deep Learning Models (HMMs, GMMs, CNNs, DNNs, RNNs, Deep Belief Networks, Dynamic Bayesian Networks, Generative Models), SVM, K-Means clustering, Dynamic Time Warping, Decision Trees, EM algorithm, KPCA which are summarized in Table 3.

A. HIDDEN MARKOV MODEL (HMM)

HMMs are commonly used to capture the dynamic properties of visual features. HMMs are based on the discrete-time Markov process. The goal is to find the most likely sequence of words for some given acoustic input. Before HMM is applied to the input speech, the input speech is first processed for feature extraction to MFCC, LPCC, LPC, and other features. HMM is specified by a set of parameters = {P, A, B} and a set of states.

$$\begin{aligned} \mathrm{P} &= \text{Prior probabilities} \\ \mathrm{A} &= \text{Transition Probabilities} \\ \mathrm{B} &= \text{Emission Probabilities} \end{aligned}$$

Its operation is characterized by a hidden state sequence $\mathrm{J} = \{j_1, j_2, \ldots, j_N\}$ and an observation sequence $\mathrm{X} = \{x_1, x_2, x_3 \ldots\ldots x_N\}$ [27].

B. DEEP NEURAL NETWORK

Deep Learning structures like DNNs differ from artificial neural networks in the number of layers between the input and the output layer. The neurons are the fundamental element of a DNN which are entirely linked to adjacent layer's neurons to create a network. Input is passed into an activation function, $r = f(s; \theta)$ [8]. Widely used activation functions are Sigmoid Function, Hyperbolic Tangent, etc. DNNs have become more popular because it takes into account large data for training and thus enhances system's performance. DNNs are also known as feed-forward NN and uses backpropagation algorithm [8], [10]. Backpropagation is also called

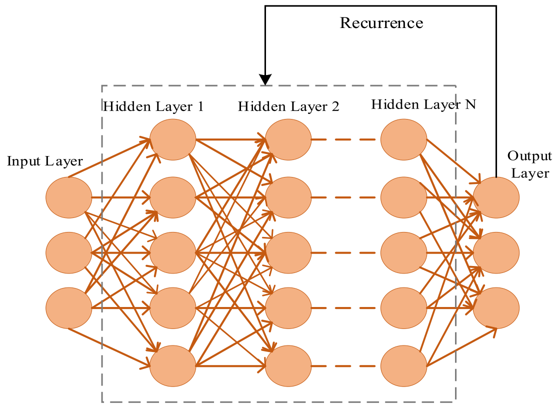

FIGURE 4. DNN with recurrence.

backpropagation of errors. Backpropagation is used for calculating the gradient with respect to the weight of the nodes.

C. CONVOLUTIONAL NEURAL NETWORK

CNN is a genre of ANN that consists of a Convolutional layer as the building block and is also known as ConvNet. This network comprises convolution and pooling layers stacked upon one another [10]. CNNs actually originated for image processing and then extended for natural language understanding and speech recognition. It is a version of multilayer perceptrons that has an input, hidden, and output layer with an activation function. A commonly used activation function for Convolutional Neural networks is ReLU. By applying a specific activation to the input, we obtain the weights [8].

D. RECURRENT NEURAL NETWORK

RNNs are DNNs with a long input data sequence. RNN as the name suggests uses recurrent connections within layers and has a strong representational memory [8]. It generates predictive results for sequential data [10]. RNN is more effective because it saves its state each time it runs into an input. The aim of RNN is to predict future sequences by making use of previous data sequences [10]. The hidden state 'w_t' sequence is calculated through the previous hidden state 'w_{t-1}'and produces an output vector sequence 'r_t' for an input sequence $s(t) = (s_1, s_2, \ldots, s_t)$. Gradient vanishing and failing to model long-term temporal events are the issues of simple Recurrent Neural Networks and thus multiple specialized RNN models were proposed like Long Short-Term Memory (LSTM).

E. GENERATIVE MODELS

Generative Adversarial Networks, Variational autoencoders, and autoregressive generative models are types of generative models. These are growing popular in the domain of speech technology because of their capability to learn and produce data distributions. Generative models consist of a generator and a discriminator neural network [8].

F. DEEP BELIEF NETWORK

It is another class of DNN that uses back-propagation for fine-tuning the entire network. This deep learning architecture is a graphical generative model that takes the advantage of unsupervised learning. It is made up of stacked restricted Boltzman machine layers that are trained one at a time [10].

G. GAUSSIAN MIXTURE MODEL (GMM)

GMM is a probabilistic model that, models univariate and multivariate datasets. GMM is quite popular in speech technology. The probability estimation is accurate and hence the classification produces the best results. It is a weighted representation of the weighted sum of Gaussian densities. GMM model consists of multiple Gaussians each identified by $m \in \{1,2,\ldots M\}$ where m is the number of clusters and is specified by parameters: $\{\mu, \Sigma, ï\}$ where μ is the mean, Σ is the covariance, ï is the mixing probability that defines the size of the Gaussian function. It is best used when a data point might belong to more than one cluster.

H. DYNAMIC BAYESIAN NETWORK (GMM)

The Bayesian network combines graph theory with probability and provides a convenient way of dealing with complexity and uncertainty. It is also known as the Acrylic Graphical Model and is an extension of the Bayesian Network. It is described as general and flexible as it is best used for complex temporal stochastic processes.

I. DYNAMIC TIME WARPING

Warping points are any two random time series that vary in time and speed. Dynamic Time Warping is applied to sequences like audio, video, graphics, or any time-varying sequence i.e., temporal in nature. It calculates a perfect match between two temporal sequences. Its applications include ASR, Speaker Recognition, online signature recognition, and also partial shape matching systems.

J. SUPPORT VECTOR MACHINE

SVM is a kind of supervised learning algorithm that categorizes data by finding a hyperplane that splits data of one class from those of all the other classes. It uses a kernel transform that transforms the non-linear data to higher-order dimensions where a hyperplane can be found. SVM can be classified into Linear and Non-Linear.

K. DECISION TREE

The decision tree resembles a flowchart in structure and its objective is to develop a model that estimates the output value derived from an input variable set. Each tree has a starting node also known as the root node. Every node is related to the input variables.

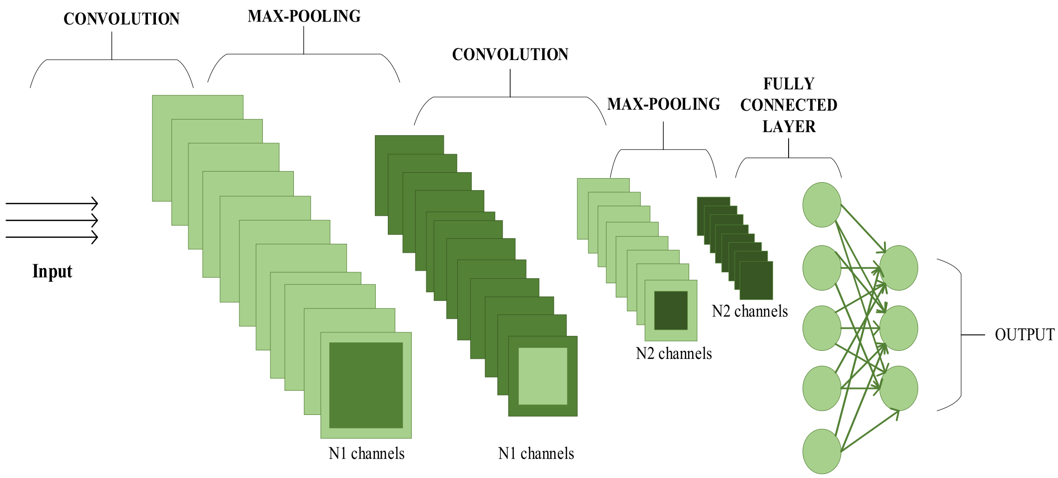

FIGURE 5. Convolutional neural network.

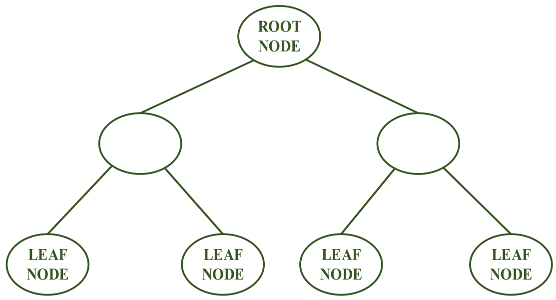

FIGURE 6. Decision tree with root and leaf nodes.

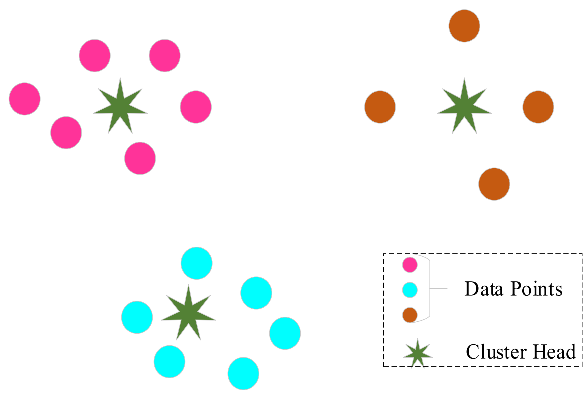

FIGURE 7. k-means clustering algorithm.

L. EXPECTATION MINIMIZATION (EM)

EM Algorithm is an iterative algorithm that can be used for latent variables. It is a type of unsupervised learning algorithm that calculates MLE which is calculated by maximizing the marginal likelihood of the observed data. This is achieved in two steps: Expectation Step (E) and Maximization Step (M).

M. KERNEL PRINCIPAL COMPONENT ANALYSIS (KPCA)

PCA implements linear transformation on the data such that the first few principal components capture the data that has the most variance or information present in the dataset. KPCA is a non-linear PCA that is developed using kernel techniques.

N. K-MEANS CLUSTERING ALGORITHM

Cluster analysis is a form of unsupervised learning technique that is further classified into Hard Clustering and Soft Clustering. K-Means clustering algorithm classifies data into mutually exclusive clusters. The distance of the data point from the cluster's center decides how well the said data point fits into a cluster. The nearest cluster center is determined for each data point, then each cluster center is replaced by an average of all the data points that are near to it. The process is continued alternatively until the local minimum is achieved through convergence. It is best used when the number of clusters is known and for fast clustering of large amounts of data sets.

IV. TECHNIQUES FOR RESOURCE ALLOCATION, ENERGY EFFICIENCY, POWER OPTIMIZATION AND MORE

Valkanis *et al.* [15] proposed an algorithm for resource and bandwidth allocation named DPPQ (Double Per Priority Queue). Services provided to the patients are updated by Tactile Internet. TI applications are widely used in healthcare applications. The proposed algorithm conserves the strict QoS demands. Two different queues are High Priority and Low Priority and transmission priority is determined by the CoS of the packet and the type of queue in which the packet is buffered. The proposed algorithm maximizes the efficiency of intra-ONU scheduling.

Saleh *et al.* [16] proposed an Emergency Efficiency architecture for Wireless Sensor Networks. A quaternary transceiver was used in the presented architecture of the sensor node. The quaternary interlink framework amends the data transfer from binary symbols to quaternary ones. It consists of amplitude and phase modulator and demodulator.

Mishra *et al.* [17] proposed an algorithm developed on the basis of an Evolutionary game also known as the Constant

TABLE 3. Algorithms for speech recognition.

S. No	Algorithm	Objective	Procedural Analysis
1.	Deep Learning Models i) Hidden Markov Model (HMMs)	To predict the most probable sequence of states, commonly used to analyze features or observations.	i) Its operation is characterized by a hidden state sequence. ii) Consists of two states: Hidden states and Visible states. iii) Determined by a set of states and parameters i.e., prior probabilities, transition probabilities, and emission probabilities.
	ii) Deep Neural Network (DNNs)	To improve the model's performance by increasing the number of layers.	i) Consists of fully integrated layers and representations. ii) A pipeline is created for non-linear translations that have the capability to learn transitional representations.
	iii) Convolutional Neural Networks (CNNs)	To capture high-level features from the input that were primarily built to recognize images but are now also long drawn out for speech recognition.	i) Prediction for Regression or Classification is achieved by fully connected layers. ii) Computationally demanding, requires GPU.
	iv) Recurrent Neural Networks (RNNs)	To generate predictive results for sequential data and well suited for modeling sequences like speech.	i) Stronger representational memory. ii) The hidden state is estimated from the previous one and produces an output vector sequence. iii) Recurrent connections within layers that have the ability to process already processed input.
	v) Generative Models	To learn data distribution of the training set to produce new data points with some variations.	i) Generator 'G' and a Discriminator 'D' are the two neural networks and play a min-max adversarial game. ii) Probabilistic models that model joint distribution.
	vi) Deep Belief Network	To produce a powerful generative model using a restricted Boltzman machine.	i) Gradient Descent Learning is used to enhance performance. ii) Consists of a stack of Restricted Boltzman Machine layers trained one at a time.
	vii) Gaussian-Mixture Model (GMM)	To maximize the likelihood value for the data or log-likelihood value of the data.	i)A powerful method to calculate the estimated density of an acoustic vector. ii)Flexible, easy to use, and accurate for multivariate data. iii) Uses a probabilistic approach to choose an input and wrongly chooses the more common option.
	viii) Dynamic Bayesian Network	To extend the Bayesian network to model temporal relationships.	i) Unsupervised layer-by-layer pre-training is applied to the entire network using back-propagation. ii) Temporal extension of Bayesian Network.
2.	Dynamic Time Warping	To find the correct alignment between two time series, and to measure the similarity between these time series.	i) Corresponding regions between two temporal sequences can be extracted easily for matching purposes. ii) Measures similarity between two sequences that differ in time and speed.
3.	Support Vector Machines	To classify data by finding hyperplanes.	i) Classifies data by finding a hyperplane that separates different classes of data. ii) Good Kernel Function is required. iii) Best used for high-dimensional, non-linearly separable data. iv) Uses kernel transform to transform non-linearly separable data.
4.	Decision Trees	To make the correct decision at the end of each node.	i) Trees spread their way out as more nodes are processed, they don't converge. Edges coming from the nodes have the total possible value of the node. ii)Overly complex models are created depending upon the type and amount of data. iii) Each tree starts with a Root Node and terminates on a leaf.
5.	Expectation-Minimization (EM) Algorithm	To find MLE (Maximum Likelihood Estimate) of parameters in a given statistical model.	i) Used to fit multivariate data into GMM models. ii) Iterated method for calculating the maximum likelihood estimates of parameters iii) Maximizing the marginal likelihood of the observed data gives MLE (Maximum Likelihood Estimate).
6.	Kernel Principal Component Analysis	To get PCA's representation in a higher-dimensional space that can create more diverse speech features.	i) Generated PCA Algorithm. ii) Reduces dimensionality with applications in face recognition, and image compression and is also now extended in speech recognition.
7.	K-Means Clustering Algorithm	To reduce the sum total of distances within points and their respective cluster centroids.	i) Best used in fast clustering of a large amount of data. ii) Data is segregated into k number of mutually exclusive clusters.

model hawk-dove game for allocation of resources on the basis of priority for medical emergencies with different time slots. Several Local Data Processing Units transmit data simultaneously which causes inconvenience at the time of some major health emergencies. Hence it becomes crucial to distinguish different LDPUs from each other. The proposed algorithm takes many factors into consideration like seriousness, urgency, etc. PATS presents higher priority and a higher number of time slots for situations that require medical emergencies.

Dai *et al.* [18] suggested a DRL-based DDPG algorithm to search for a solution to the Markov Decision Process and incorporate the action refinement in DRL. The proposed Algorithm incorporates action refinement in DRL for providing a solution for computation offloading and Resource Allocation. DRL is a division of AI in which an agent is acting an environment that must take certain steps to reach a certain state. DDPG is composed of three modules: Primary Network, Target Network, and Replay Memory. Primary Network consists of two DNNs (Primary Actor and Primary Critic) and maps present state x_t to an action state y_t. For Actor DNN training, the target network generates target values and Replay memory stores experience tuples.

Zhang *et al.* [19] present an efficient and reliable sleep scheduling strategy. The authors present two algorithms GAA and LAA. GAA stands for Global Approximation Algorithm, which is denoted by H $(M + \delta)$. The Polymatroid function is designed for constructing an MWmDS. LAA stands for Local Approximation Algorithm and minimizes the computational complexity as compared to GAA. An optimal node is selected from each one-hop region and selects numerous modes to MWmDS.

Luo *et al.* [20] presented a DBN-based path for power reduction. The DBN presented by the authors has three stages: data preparation, Training, and Running. Data preparation is done by randomly generating a set of channel gains, and applying a genetic algorithm to determine the training sample's output. Training of the data is done through both supervised and Unsupervised Learning methods. In the running stage, the well-trained networks are used and overworked to predict and foretell the way out to the said power minimization issue.

Liu *et al.* [21] proposed a DRL framework that predicts the overall needs and requirements of the network. They focused on the MD-IMA design for TDM. MD-IMA is designed by the collaboration of both LSTM and DRL. Resource management is done through clustering, allotment of subchannels, and power allocation among users [21].

Sodhro *et al.* [22] presented an algorithm namely MMMM that is an ML-driven method for Mobility Management. The proposed algorithm is designed for industrial NIB communication networks that are energy efficient in nature. The entire proposed system is based on security needs. The proposed algorithm works with less overhead and low arithmetic complications. The entire unified process is ranging from key generation to information preservation keeps on going until the desired level is attained.

Sodhro *et al.* [23] proposed an algorithm namely AETPC (Adaptive Energy-Efficient Transmission Power Control). It adjusts temporal variations in the static and dynamic postures in the wireless channels. In comparison to the traditional TPC and PTPC techniques, the proposed algorithm saves 11.25% energy. The authors proposed a joint duty cycle and power transmission power control adaptation model consisting of a Base station, Access Point, and sensor nodes for energy saving in the BSNs.

V. SECURITY ISSUES AND COUNTERMEASURES

Users are concerned about privacy and security, which is among the most talked-about concerns encountered by wireless networks. Both industry and academia are increasingly interested in developing privacy-assured ML algorithms in order to reap the benefits of ML while respecting user privacy [30].

Authors of [4] suggested a secure Wi-Fi network developed by a respectable manufacturer and configuring it according to the application. Other simple ways include changing the passwords and updating software periodically.

DeSVig seeks to detect adversarial attacks in IAISs. A control plane, data plane, and OpenExample protocol are all part of the proposed system model [28]. The data plane contains n Deep Learning Models, each of which serves a large number of consumers. A mobile computing agent, a CGAN, and a discriminator make up the control plane's controller. Among several DL models, the control plane efficiently detects threats and executes vigilance quickly. The proposed system is put to the test using two datasets: MNIST and a self-created industrial dataset. It is more reliable, efficient, and scalable than other solutions.

SSL makes use of both labeled and unlabeled data and is a powerful tool to discover hidden information. The authors of [29] suggested a DeNeB's SSL method requires fewer data poisoning expenses and results in greater backdoor efficacy. Trigger patterns and adversarial concern data are fed to the unlabeled training data which in turn gives them a higher success ratio. To resist newly designed attacks for a secure SSL, the authors have proposed a novel DePuD, i.e., Detection and Purification Defense strategy to rectify the discovered threat and encourage SSL algorithms. The suggested DePuD comprises a detector that locates the poisoned data and filters it out before forwarding it for further procedure.

Authors of [30] explore the issue of privacy-preserving collaborative DL by considering unreliable participants and proposed a novel solution called SecProbe. It effectively secures and safeguards each participant's data privacy while learning an accurate model. In the proposed model they have considered a global model and an additional validation dataset on the server. There are N participants present whose aim is to grasp and acquire a common model. Participants only exchange

TABLE 4. Techniques for implementation of healthcare networks.

S. No	Metric	Algorithm Proposed	Objective	Inference	Ref
1.	Resource and Bandwidth Allocation	DPPQ (DWBA)	To conserve rigorous and strict QoS demands of TI applications.	i) Provides an amalgamation of techniques and approaches that provide energy-efficient, equitable management, and resource allotment. ii) There is an implementation of two queues viz High Priority and Low Priority, double per CoS queues improve ONU scheduling efficiency.	[15]
2.	Energy Efficiency	NN-SRAM	To alter data transmission in a Wireless Sensor Network in an energy-efficient manner.	i) Proposes a Quaternary interlink framework that amends and improves data transmission. ii) The proposed architecture accumulates data, stores it, and remains sustainable until the power is supplied and is made up of internal latches. iii) Drops the amount of power to 76.99% more than the conventional one.	[16]
3.	Time Slot Allocation	Priority Based Allocation for Time Slots (PATS)	To present a relative measure that prioritizes the nodes according to their importance and influence.	i) In the case of medical emergency scenarios, several LDPUs transmit their data simultaneously. Thus, it becomes crucial to distinguish the LDPU transferring crucial and sensitive data from the regular data. ii) Proposed Algorithm is based on the evolutionary game also known as the Constant Model hawk-dove game. LDPUs select propositions developed on the basis of fitness.	[17]
4.	Resource Allocation and Computation Offloading in Beyond 5G	DRL based DDPG	To present a DRL-based algorithm for Resource allocation and computation offloading.	i)The algorithm uses an actor-critic architecture, in which the actor generates actions and the critic guides the actor to perform and generate better actions. ii) Comprises three components i.e., Primary and Target Network, and Replay Memory. iii) Energy utilization is minimized considering meticulous constraints into account.	[18]
5.	Energy Efficiency	• GAA • LAA	To propose an efficient and reliable scheduling algorithm.	i) A Dominating Set (DS) with m-folds is constructed where m denotes the number of links present from DS to the nodes and vice-versa. ii) A polymatroid function is designed by an MWmDS through GAA. iii) GAA is also represented as H (M+ δ). iv) LAA minimizes computational complexities and is denoted by 1+ln (mδ).	[19]
6.	Power minimization	DBN based approach	To minimize energy dissipation and prolong the battery life.	i) Power minimization issue is addressed by developing a joint optimal solution. ii) The proposed framework is made up of three parts: Data Preparation, Training, and Running.	[20]
7.	Situation Aware Resource Allocation	DRL based approach	To estimate the overall network necessities and provisions for the long haul.	i) Allotment of the Resources for MD-IMA is powered by DRL estimation. ii) DRL and the LSTM are integrated together to work collectively.	[21]

TABLE 4. *(Continued.)* Techniques for implementation of healthcare networks.

				iii) Power allocation in the MD-IMA system is presented using DDPG Algorithm.	
8.	Energy Efficient mechanism for 6G enabled	MMMM	To develop an ML-aware energy allotment algorithm.	i) Key generation and information preservation continue until the desired authentication level is obtained. ii) Resources having high energy efficiency are allocated through a dynamic wireless link.	[22]
9.	Power Control	AETPC	To adjust temporal disparities during static and dynamic body postures in the wireless channel.	i) In static and dynamic postures, the proposed algorithm has a steady value of RSSI levels. ii) With realistic PLR the proposed algorithm saves 11.5% more energy.	[23]

the parameters and the server deals with exchanging, storing, and communicating with the participants. SecProbe uses both exponential and functional mechanisms to protect data privacy and data quality.

In [31] authors have presented a trustworthy privacy-preserving framework for ML in IIoT systems. Because ML models are built on sensitive data, they have a tendency to leak private information, limiting their potential in Industry 4.0. PriModChain, the suggested paradigm, mandates privacy and trustworthiness by combining differential privacy, federated ML, Ethereum blockchain, and smart contracts. FedML was used to federate and share ML models globally, while DP ensured that the ML models were kept private. The addition of smart contracts and the EthBC to the framework adds traceability, transparency, and immutability. IPFS combines safe P2P content transport with immutability, low latency, and quick decentralized archiving.

Conventional cryptographic algorithms have been used to solve the security and privacy challenges in IoT networks. The authors of [32] have presented a review of IoT security solutions based on ML and DL.

IAI is applied to various problems persisting in industry 4.0. Performance of the system can be calculated by exploiting the shared parameter adversaries. In [33] the authors have proposed privacy-enhanced federated learning for IAI such that even if numerous entities collaborate, it is possible to prevent private data from being disclosed. A key generating center, a cloud service, and several participants make up the proposed system. PEFL is noninteractive in each aggregation. Furthermore, PEFL protects the privacy of training data throughout and after the training process, even when an opponent conspires with multiple groups.

Data providers face significant challenges in sharing their data through wireless networks due to security and privacy concerns. In [34] the authors modified the data-sharing problem by applying Federated learning to build data models and sharing them instead of raw data. They proposed a collaborative architecture that is blockchain empowered, for sharing data among several parties and minimizing the data leakage risk. Furthermore, for data protection, they integrated differential privacy into federated learning.

Also, the authors of [35] have taken into an account a case in which several data owners want to apply an ML algorithm for a joined dataset and to achieve significant learning results without having to share the local datasets due to privacy concerns. They designed systems for such scenarios making use of the SGD and its variants. The systems are named Server-aided Network Topology, and Fully-connected Network Topology. The designed system can handle any activation function, and instead of the gradients calculated by SGD, weights are being transferred.

VI. PROPOSED ARCHITECTURE

Healthcare is one of the most crucial, growing, and advancing fields as far as the application of AI/ML is concerned. Machine Learning has tremendous applications in the healthcare sector. Since the pandemic, the health care system has faced serious challenges, ranging from critical health problems to basic health problems. The rapid growth in the health problems and decrease in the number of caregivers and healthcare facilities makes it even more important to design a smart healthcare system that provides basic healthcare facilities to the people in need.

This paper is proposing a novel smart healthcare system that is designed for basic health problems using speech recognition. The accuracy of such speech recognition systems depends on the type and the amount of data that is being fed to it to train the machine. Healthcare and Speech recognition are two such fields that have a lot of applications in the field of AI. Hence this paper is proposing a concept to combine these two popular prime sectors to present a framework that will provide health services to the people in need and take the load off of the caregivers and health service providers.

TABLE 5. Privacy and security countermeasures.

S.No	Algorithm Proposed	Objective	Inference	Reference No.
1.	DeSVig	To identify Adversarial attacks in IAISs.	i) Designed to overcome the challenges of ultralow latency caused by industries and generative adversarial networks. ii) Robust, scalable, and highly efficient than existing state-of-art.	[28]
2.	i) DeNeB ii) DePuD	DeNeB: To identify insidious backdoor threats. DePuD: To identify the location of the poisoned data.	i) Specially designed for SSL. ii) DeNeB uses poisoned unlabeled data to create a strong neural backdoor. iii) DePuD identifies the location of the poisoned data so as to not forward it for further procedure.	[29]
3.	SecProbe	To protect the data privacy while learning the model effectively.	i) Uses both exponential and functional mechanisms that protect both data quality and privacy.	[30]
4.	PriModChain	To present a trustworthy privacy-preserving strategy for ML in IIoT systems.	i) Combines differential privacy, FedML, and EthBC. ii) Designed for Industrial IoT systems.	[31]
5.	PEFL	To solve the issue of privacy concerns while sharing parameters for IAIs.	i) Designed for IAIs. ii) It is non-interactive in each aggregation. iii) Preserves privacy of training data during and after the training.	[33]
6.	Blockchain empowered collaborative architecture	To share data over various distributed entities to minimize data leakage.	i) Minimizes data leakage. ii)Integrates differential privacy into federated learning which further protects data privacy.	[34]
7.	i)SNT ii)FNT	To attain the best possible output for a number of data owners to share data for a combined dataset.	i) Instead of gradients, they share weights. ii) Uses any activation function.	[35]

The proposed smart healthcare framework consists of a Front-End system performing human device interaction and executing speech recognition, C-RAN network architecture, and different computing techniques based on the priority of the case. What makes this work original and novel is the proposed architecture that uses a Speech Interactive system in a C-RAN architecture and does its computing on the basis of priority (allotted on the basis of data rate). The outputs of several front-end interactive systems are pooled together in a centralized BBU pool incorporating C-RAN architecture, which is then transferred for computing (edge/fog/cloud based on the priority of the case).

Each element and its work concerning the suggested system is described below.

A. FRONT END SYSTEM

The front end of the proposed system consists of a human device interaction (HDI) system. The proposed HDI system performs speech recognition and has a voice assistant. Speech recognition is a methodology that takes speech as an input and transcribes it to text [10], [11], [12], [13]. Speech is a naturally occurring time sequence. To transcribe speech waves to the text, we train it with speech data. Data that is fed for training needs to vary in age, gender, accent, and environmental noise. In the proposed system there are virtual voice assistants for communication between the users and machine. It translates the user's voice request to text and then into JSON format [4]. JSON is an interchangeable data format, is language-independent, and is used for sending data from the server to the client, and vice versa. In this scenario, it sends a request from the user to the server through REST API. REST is a client-server architecture that ensures communion within the system and voice virtual assistant [4]. REST API is used to send the JSON User's request to the server and the server's response back to the user [4]. Apart from JSON, there is another way to communicate data from

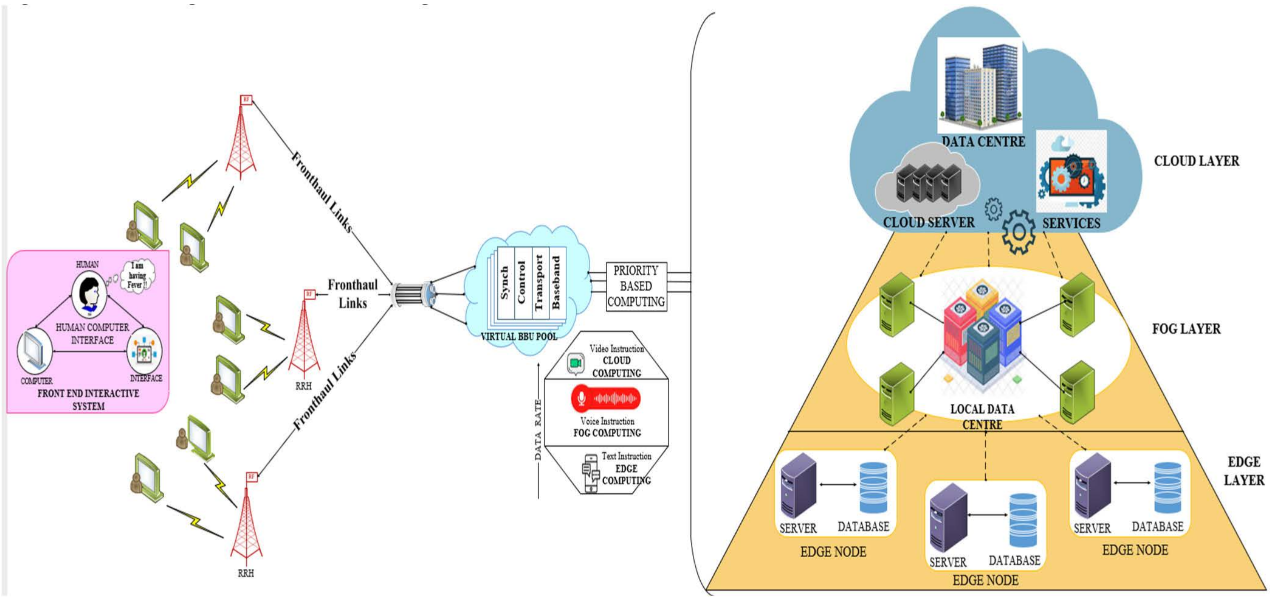

FIGURE 8. Proposed architecture for intelligent and interactive healthcare system.

client to server, and that is through the XML format. However, JSON is widely used as it is much faster than XML as it was specifically designed for data interchange. XML is not just used for data interchange; hence, it is much slower. JSON encoding is brusque and its parsers are less complicated and hence take less processing time. The most widely used protocol for these requests and responses is HTTP. The HTTP methods are POST, GET, PUT, and DELETE, corresponding to create, read, update, and delete (CRUD) [4].

In this proposed model HMM is applied in the processing of I^{2}HS. Hidden Markov Model is popular for training temporal sequences and has various applications in Speech Recognition, Pattern Recognition, and Activity Recognition (for security surveillance purposes).

B. C-RAN

C-RAN is a network architecture where C can be transcribed as Cloud, Cooperative Radio, Centralized processing, Clean or Collaborative. It is specially designed for addressing the issues faced by the operators because of the increase in the number of users and their needs. The objective of C-RAN is to amalgamate the BBU Units from many base stations into a centralized BBU pool [14]. In this architectural style network area is split into cells and resources are shared between the base stations. This multiplexes the gain and shifts the load to fast wireline transmission of quadrature and in-phase data and improves the capacity of the network.

C-RAN architecture benefits both macro and small cell [14] networks by adapting to non-uniform traffic and scalability, decreasing delays, increasing throughput, increasing statistical multiplexing gain, saving energy and cost, and easing maintenance and network upgrades. By creating a reconfigurable mapping between RRH (Remote Radio Head) and BBU (Base Band Unit), statistical multiplexing gain can be maximized. Multiple BBUs are pooled together that are automatically reconfigured in the event of a loss. Optical fiber or microwave links are used to connect the backend to the virtual BBU pool [14].

C. PRIORITY BASED COMPUTING

In the proposed system computing is priority-based and priority is given on the basis of data rate. As the number of internet users grows, the need for a computation offloading approach has come to the light and that is why edge, fog, and cloud computing are used. Text instructions have a lower data rate so they should perform edge computing. Voice instructions have a relatively higher data rate than text instructions thus they should perform Fog computing. Video instructions have the highest data rate and hence they perform Cloud Computing. All these techniques of computing i.e., Edge, Fog, and Cloud are associated with distributed computing and focus on the physical stationing of storage and compute resources about the data that is being generated. Albeit making use of the cloud yields immense indefinite space and computational power, it does have a ramification i.e., time lag in transferring data.

1) EDGE COMPUTING

Edge computing moves some storage and computes resources closer to the end-users so that it can be performed in high-priority cases like text instructions that have low data rates. Edge computing forms edge nodes, each node consisting of an edge server with some local database. The original purpose for using edge computing was to minimize the bandwidth costs for devices, to communicate via long distances. However, with the increase in the number of edge devices, the volume of data generated increases as well.

Edge computing leads from the cloud because it substantially reduces latency issues, improves operation efficiency, and reduces bandwidth costs.

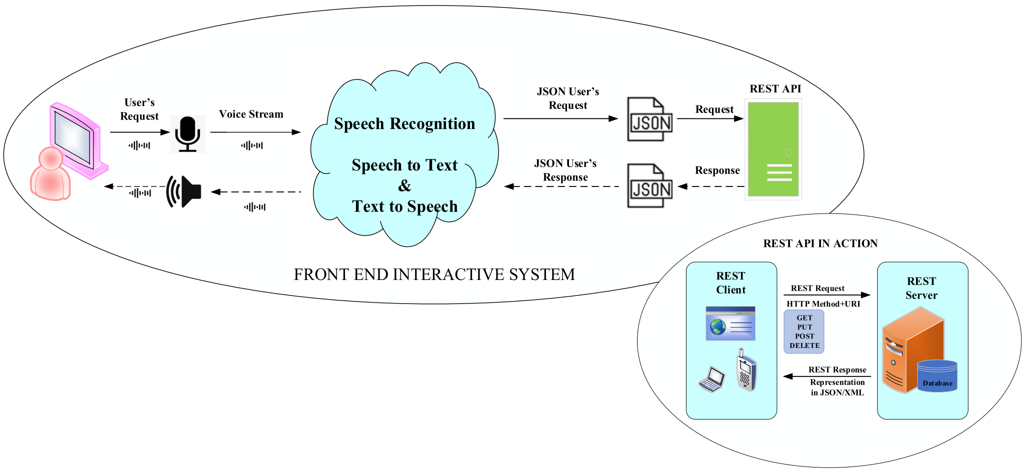

FIGURE 9. Front end interactive system.

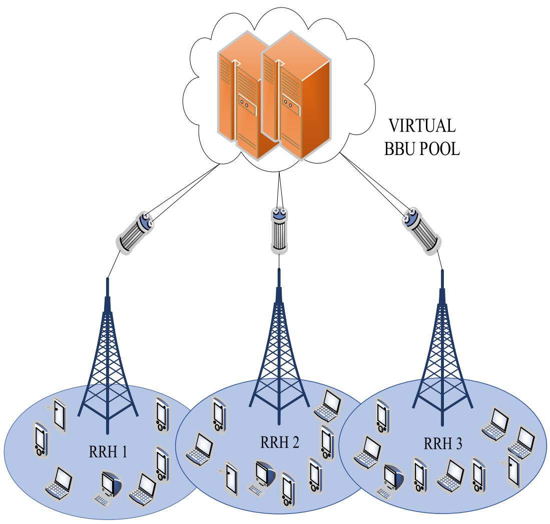

FIGURE 10. C-RAN.

2) FOG COMPUTING

Also known as fogging or fog networking. It is a decentralized computing technique situated between cloud and edge devices that generates data. It is performed in cases that are at a higher priority than text instructions like voice instructions. The goal of this technique is to build better operational connectivity and perform local data analysis and filtering. It has a local data Centre with state-based servers and is used for real-time analytics. Fog computing bridges the cloud and the end edge devices for various services, networking, and storage.

3) CLOUD COMPUTING

Cloud is not just used to store data but it also makes use of the internet to run many of the software applications and networks from a remote cloud server. Cloud computing is efficient, flexible, provides services, and is easily accessible.

It's about developing a hybrid environment to house the capabilities of the above-mentioned computing techniques in a way that will optimize the interests of each.

Incorporating fog and edge computing techniques plays a crucial role in developing this intelligent healthcare system by substantially narrowing the computing load from cloud servers, which assures services with quick response time.

VII. INTELLIGENT AND INTERACTIVE HEALTHCARE SYSTEM USING HIDDEN MARKOV MODEL (HMM)

A. DESCRIPTION OF HMM

HMM is used best in modeling high-complexity systems. HMM has a finite number of states which are ruled by a set of transition probabilities. HMM has found huge applications in Bioinformatics, genomics, and also for security like anomaly detection and intrusion detection systems by present-day researchers [26]. With a defined probability distribution, and for a definite state, observation is generated which is generated by an observer and not the state.

1) The number of states in the model is denoted by $\Phi = \{\Phi_1, \Phi_2, ..., \Phi_{Ns}\}$, where $\Phi_k = 1, 2, ..., N$ denote an individual state. The corresponding state at any time instant 't' is given by δ_t.
2) Number of distinct observation symbols is depicted by η, and then are updated according to the output of the system. The set of symbols is given by $\Psi = \{\Psi_1, \Psi_2, ..., \Psi_\eta\}$, where $\Psi_k = 1, 2, ..., \eta$ represent a distinct symbol.
3) State Transition Matrix is given by $R = [r_{xy}]$, where $r_{xy} = P(\delta_{t+1} = \Phi_y | \delta_t = \Phi_x)$, $1 \leq x \leq N_S$, $1 \leq y \leq N_S$; $t = 1, 2, \ldots$
4) State transition is achieved when $r_{xy} \geq 0$, $\forall$ x,y and $\sum_{n=1}^{Ns} r_{xy} = 1$, $1 \leq x \leq N_S$.
5) Probability matrix of observation symbol is denoted by $B_{obs} = [b_n(z)]$, where $b_n(z) = P(\Psi_Z | \Phi_n)$, $1 \leq n \leq N_S$, $1 \leq z \leq \eta$ and $\sum_{z=1}^{N_S} b_n(z) = 1$, $1 \leq n \leq N_S$.

$\varphi = |\varphi_i|$ represents the initial state probability vector.

6) Observation sequence is represented as $\varpi = \{\varpi_1, \varpi_2, \ldots, \varpi_r\}$ where each ϖ_t is an observation symbol among the set Ψ and r denotes the observations in the sequence.

The above description makes it clear that parameters N_S, η, probability distributions (State Transition Matrix, Observation symbol probability matrix and φ) are required for HMM estimation. The entire set of parameters is given by $\zeta = \{R, B_{obs}$ and $\varphi\}$ [27].

B. HMM MODEL FOR I²HS PROCESSING

The operation of HMM is characterized by an observation sequence and a hidden sequence. The hidden state sequence is represented as $\delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$, where $\delta \in \Phi$. The observation sequence is represented as $\varpi = \{\varpi_1, \varpi_2, \ldots, \varpi_r\}$ [26].

For a sequence of states $\delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$, according to first-order Markov assumption probability of observation at time n only depends on observation at time n-1 and is given by:

$$P(\delta_n|\delta_{n-1}, \delta_{n-2}, \ldots \delta_1) = P(\delta_n|\delta_{n-1}) \tag{7}$$

The output sequence for which the above equation is satisfied is known as the First-order Markov chain.

For second-order assumption, the probability of the output sequence is given by:

$$P(\delta_n|\delta_{n-1}, \delta_{n-2}, \ldots \delta_1) = P(\delta_n|\delta_{n-1}, \delta_{n-2}) \tag{8}$$

Similarly for third-order assumption, the probability of the output sequence will be given as:

$$P(\delta_n|\delta_{n-1}, \delta_{n-2}, \ldots \delta_1) = P(\delta_n|\delta_{n-1}, \delta_{n-2}, \delta_{n-3}) \tag{9}$$

The joint probability of past and current observations for certain sequence $\delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$ using the Markov assumption is given by

$$P(\delta_1, \delta_2, \ldots, \delta_n) = \prod_{i=1}^{n} P(\delta_i|\delta_{i-1}) \tag{10}$$

I. **State Representation:** In accordance with the approach followed in the proposed strategy, the four states considered are the Front-End System (FES), Pre-Processing and Feature Extraction, Training, and Prediction. These are denoted as

$$\Phi = (\text{FES, PFE, T, and P})$$

An entirely linked HMM with transition probabilities of the proposed HMM given in Table 6 is shown in Figure 11 where each and every stage can be obtained by a single hop and every client will be trained and maintained by HMM.

II. **Conditional Probability of the Hidden state and the Observation Sequence:** The probability of a certain state $\delta_i \in \Phi$ can only be based on the observation ϖ_i.

TABLE 6. Proposed HMM for intelligent and interactive healthcare system.

Demand	Response			
	Front End System (FES)	**Pre-processing and Feature Extraction (PFE)**	**Training (T)**	**Prediction (P)**
Front End System (FES)	0.015	0.41	0.55	0.025
Pre-processing and Feature Extraction (PFE)	0.60	0.015	0.30	0.085
Training (T)	0.085	0.3	0.015	0.60
Prediction (P)	0.09	0.3	0.60	0.01

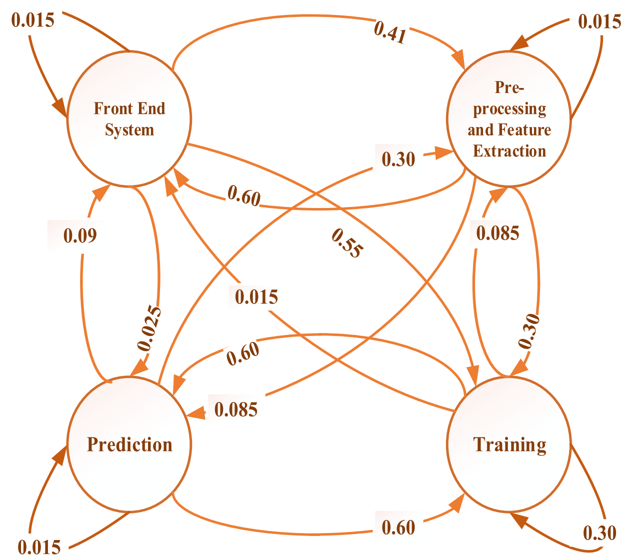

FIGURE 11. HMM for intelligent and interactive healthcare system with probabilities according to table 4.

The conditional probability $P(\delta_i|\varpi_i)$ according to the Baye's Rule is given by:

$$P(\delta_i|\varpi_i) = \frac{P(\varpi_i|\delta_i)P(\delta_i)}{P(\varpi_i)}. \tag{11}$$

For state sequence $\delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$ and observation sequence $\varpi = \{\varpi_1, \varpi_2, \ldots, \varpi_n\}$is denoted as

$$P(\delta_1, \delta_2, \ldots, \delta_n|\varpi_1, \varpi_2, \ldots, \varpi_n) = \frac{P[(\varpi_1, \varpi_2, \ldots, \varpi_n)|(\delta_1, \delta_2, \ldots \delta_n)]P(\delta_1, \delta_2, \ldots \delta_n)}{P(\varpi_1, \varpi_2, \ldots \varpi_n)} \tag{12}$$

The probability P (ϖ_1, ϖ_2, ..., $\varpi_n|\delta_1$, δ_2, ..., δ_n) can be calculated as $\prod_{i=1}^{n} P(\varpi_i|\delta_i)\forall$ i the δ_i, ϖ_i are independent of all $\delta_i$and ϖ_i also, i$\neq$ j.
The likelihood proportional to probability and we denote it as Θ.

$$
\begin{aligned}
&\mathrm{P}(\delta_1, \delta_2, \ldots., \delta_n|\varpi_1, \varpi_2, \ldots., \varpi_n)\\
&\quad \times \alpha\Theta(\delta_1, \delta_2, \ldots., \delta_n|\varpi_1, \varpi_2, \ldots.\varpi_n)\\
&= P(\varpi_1, \varpi_2, \ldots., \varpi_n|\delta_1, \delta_2, \ldots., \delta_n)\\
&\quad . P(\delta_1, \delta_2, \ldots., \delta_n) \qquad (13)
\end{aligned}
$$

With first-order Markov assumption, it is given by:

$$
\begin{aligned}
&\Theta(\delta_1, \delta_2, \delta_3, \ldots., \delta_n|\varpi_1, \varpi_2, \ldots., \varpi_n)\\
&= \prod_{i=1}^{n} P(\varpi_i|\delta_i). \prod_{i=1}^{n} P(\delta_i|\delta_{i-1}) \qquad (14)
\end{aligned}
$$

In our case, equations 7,8 and 9 can be re-written as

$$
\begin{aligned}
&\mathrm{P}(\delta_{FES}|\delta_{PFE}, \delta_{T,}\delta_P, \delta_{FES})\\
&\quad = P(\delta_{FES}|\delta_P); \quad \text{first-order} \qquad (15)\\
&\mathrm{P}(\delta_{FES}|\delta_{PFE}, \delta_{T,}\delta_P, \delta_{FES})\\
&\quad = P(\delta_{FES}|\delta_P, \delta_T); \quad \text{second-order} \qquad (16)\\
&\mathrm{P}(\delta_{FES}|\delta_{PFE}, \delta_{T,}\delta_P, \delta_{FES})\\
&\quad = P(\delta_{FES}|\delta_P, \delta_T, \delta_{PFE}); \quad \text{third-order} \qquad (17)
\end{aligned}
$$

Equation (17) is a third-order Markov assumption equation in which the probability of the Front-End System depends on the previous states i.e., Pre-Processing and Feature Extraction, Training, and Prediction.

III. **Training:** In the processing of the proposed system, the training stage plays a very important role. The proposed system is trained with the input speech data through the four stages i.e., Front-End System, Pre-Processing and Feature Extraction, Training, and Prediction using HMM with transition probabilities as presented in table 6. The probabilities in every row of table 6 sum up to 1.

Reinforcement Learning is another type of ML that consists of an agent acting in an environment that needs to know what steps to take to acquire a certain stage. The agent performs an action that is based on the rewards or penalties that the agent earns in various different states.

During a specific stage, the agent performs an action that leads the environment in acquiring a new state and awarding a reward. Q-Learning is a model-free approach that is based on function value. Q-value is determined by state-action pairs. Performance Analysis can also be done through Q-learning algorithm of I^{2}HS.

VIII. FUTURE RESEARCH CHALLENGES

Despite the encouraging and bright future of speech technology and its vast applications in healthcare, it faces several hurdles which are demonstrated here.

Real-time patient monitoring faces a challenge when dealing with incomplete data. The proposed system is designed for remote areas where the loss of electric power is quite common. Loss of electricity results in the loss of data acquisition before archiving it to a central location.

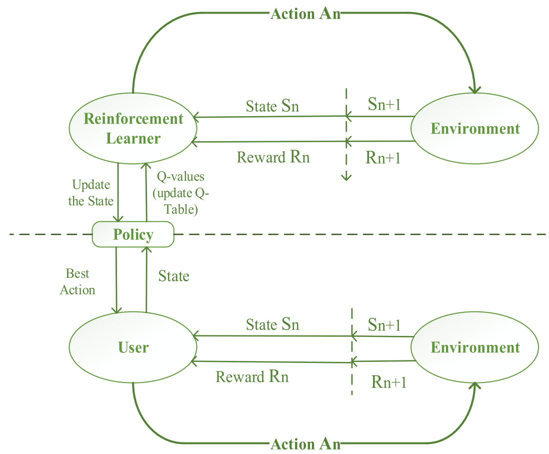

FIGURE 12. Q-learning reinforcement learning.

Scarcity of data is another challenge that we come across and to fit ML models into a small amount of data is an issue. Speech varies from tone, accent, and different languages, and thus training the data with respect to different features and developing healthcare systems for rural and local areas becomes difficult.

Global healthcare solutions are available but local healthcare solutions are not available. Cultural and Language barriers pose some major challenges in the advancement of speech technology. The rapid growth of interactive systems is halted because of cultural barriers.

Not many datasets are available for speech processing. And hence we may have to use some third-party databases for training.

Speech technology and processing in the healthcare sector can be best utilized when the data produced from various medical devices and several Electronic Health Records (EHR) is interoperable.

Developing a prediction algorithm for better optimization with greater accuracy in providing healthcare services is a challenge.

IX. CONCLUSION

In this paper, we have articulated the flourishing necessity for an Intelligent and Interactive Healthcare System in today's time. We have analyzed the previous work done so far in developing speech interactive systems and have also mentioned the research gap and the contribution of our paper. This paper explains why speech interaction method is taken as the core of our proposed framework. Various algorithms available and applied in speech recognition are also discussed. Also, several energy and power efficient techniques along with few resource allocation techniques are discussed. We have proposed a novel architecture for implementing an Intelligent and Interactive Healthcare System (I^{2}HS) for remote areas.

TABLE 7. Ongoing projects.

S. No	Project Name	Aim of Research	Area of Research	HTTP location
1.	ESSENCE (Emphatic platform to personally monitor, Stimulate, enrich, and aSsist elders aNd Children in their Environment)	To develop a model for domiciliary care that centers on remote monitoring, stimulation, and communication between users, and professionals.	Digital Tools and AI analytics, Medical and Health sciences	https://cordis.europa.eu/project/id/101016112
2.	India Covid-19 Emergency Response and Health Systems Preparedness Project	To detect, prohibit and address the threat of Covid-19.	Healthcare Services	https://projects.worldbank.org/en/projects-operations/project-detail/P173836
3.	PyXy.AI	To provide a multi-parametric system for cardio-pulmonary chronic patients early Covid-19 detection.	Healthcare Services and AI	https://cordis.europa.eu/project/id/101016046
4.	ICovid	To rapidly detect and diagnose covid through chest CT analysis.	Computer and Information Sciences	https://cordis.europa.eu/project/id/101016131
5.	VASCOVID	To design a transportable, compact platform for evaluation and analysis of microvascular health in Covid-19 patients at ICU.	Spectroscopy, Medical and Health Sciences.	https://cordis.europa.eu/project/id/101016087
6.	PORSAV (Protecting OR Staff from Aerosolized Virus)	To accumulate and assess data regarding aerosols from the operating room using Schlieren imaging for better surgical safety in the pandemic.	Healthcare services	https://cordis.europa.eu/project/id/101015941
7.	CleanAir	To protect health practitioners from Covid-19 virus Infection by creating an air decontamination system.	Medical and Health Sciences	https://cordis.europa.eu/project/id/101016174
8.	ICU4Covid (ICU for Covid)	To develop an Intensive Care Unit for Covid-19 Patients with an efficient and effective diagnosis while minimizing the risk of infection within the care providers.	Robotics, Critical Care Medicine	https://cordis.europa.eu/project/id/101016000
9.	Covid-X (Covid exponential Program)	To provide eHealth solutions during Covid-19 pandemic using data sandbox with AI capabilities to provide easy access, and providing data services consisting of visual analytics, decision-based systems, and a few more features.	Data Protection, Artificial Intelligence	https://cordis.europa.eu/project/id/101016065
10.	An Electronic Health Record-Based Screening Tool to Support Safe Discharges of Covid-19 Patients in the energy department.	To develop a screening tool for Covid-19 patients for earlier hospitalization of the most at risk from the disease.	Predictive Modelling, Natural Language Processing	https://digital.ahrq.gov/ahrq-funded-projects/ehr-based-screening-tool-support-safe-discharges-covid-19-patients-emergency-department
11.	COVIRNA	To develop a predictive system making use of AI and digital tools on the basis of Covid-19 outcomes of cardiovascular biomarkers.	Artificial Intelligence, Medicinal Chemistry	https://cordis.europa.eu/project/id/101016072
12.	ENVISION	To develop and design a predictive, intelligent system for real-time monitoring, treatment, and prediction.	AI-driven data analytics	https://cordis.europa.eu/project/id/101015930
13.	METRICS (Metrological Evaluation and Testing of Robots in International CompetitionS)	To structure the four most precedence scenarios: Inspection and Maintenance, Agri-food, agile production, and healthcare in a sustainable manner.	Interactive Systems, Healthcare, Robotics, and Artificial Intelligence	https://cordis.europa.eu/project/id/871252

The architecture proposed makes use of C-RAN network architecture, and computing techniques like edge/fog/cloud for faster communication, reducing time delay, better storage, and computing services. A mathematical model for the proposed architecture using HMM is also presented. We have also discussed various ongoing projects.

TABLE 8. Abbreviation list.

S. No	Abbreviation	Meaning
1.	I^{2}HS	Intelligent and Interactive Healthcare System
2.	LSTM	Long Short-Term Memory
3.	RNN	Recurrent Neural Network
4.	ASR	Automatic Speech Recognition
5.	VAD	Voice Activity Detection
6.	BBU	Baseband Unit
7.	RRH	Remote Radio Head
8.	C-RAN	Cloud Radio Access Network
9.	KPCA	Kernel Principal Component Analysis
10.	MLE	Maximum Likelihood Estimation
11.	GAN	Generative Adversarial Networks
12.	VAE	Variational Autoencoders
13.	OPEX	Operating Expense
14.	LDA	Linear Discriminant Analysis
15.	STFT	Short-Time Fourier Transform
16.	WER	Word Error Rate
17.	MFCC	Mel Frequency Cepstral Coefficient
18.	TI	Tactile Internet
19.	CoS	Cost of Service
20.	ONU	Optical Network Unit
21.	LDPU	Local Data Processing Unit
22.	GAA	Global Approximation Algorithm
23.	LAA	Local Approximation Algorithm
24.	PATS	Priority Based Time Allocation Slots
25.	AETPC	Adaptive Energy-Efficient Transmission Power Control
26.	MMMM	ML-driven Mobility Management Method
27.	HDI	Human Device Interaction
28.	HMI	Human-Machine Interaction
29.	JSON	Java Script Object Notation
30.	XML	Extensible Markup Language
31.	REST	Representational State Transfer
32.	DRL	Deep Reinforcement Learning
33.	EHR	Electronic Health Records
34.	HMM	Hidden Markov Model
35.	IoMT	Internet of Medical Things
36.	CGAN	Conditional Generative Adversarial Network
37.	SSL	Semi-Supervised Learning
38.	IIoT	Industrial Internet of Things
39.	FedML	Federated Machine Learning
40.	IAISs	Industrial Artificial Intelligence Systems
41.	DeNeB	Deep Neural Backdoor
42.	DeSVig	Decentralized Swift Vigilance
43.	DLNTAP	Deep Learning Network Traffic Analysis and Prediction
44.	DLNTC	Deep Learning Network Traffic Classification
45.	FCA	Flow Clustering and Analysis
46.	SNT	Server-aided Network Topology
47.	FNT	Fully-connected Network Topology
48.	SAR	Socially Assistive Robot
49.	RSSI	Received Signal Strength Indicator
50.	PLR	Packet Loss Ratio

APPENDIX

Table 7 contains a list of current ongoing projects in the healthcare domain around the world. In conjunction with this, a list of the abbreviations used in the paper along with their full form meaning is presented in Table 8.

REFERENCES

[1] M. Nasr, M. M. Islam, S. Shehata, F. Karray, and Y. Quintana, "Smart healthcare in the age of AI: Recent advances, challenges, and future prospects," *IEEE Access*, vol. 9, pp. 145248–145270, 2021.

[2] S. U. Amin, M. S. Hossain, G. Muhammad, M. Alhussein, and M. A. Rahman, "Cognitive smart healthcare for pathology detection and monitoring," *IEEE Access*, vol. 7, pp. 10745–10753, 2019.

[3] T. Muhammed, R. Mehmood, A. Albeshri, and I. Katib, "UbeHealth: A personalized ubiquitous cloud and edge-enabled networked healthcare system for smart cities," *IEEE Access*, vol. 6, pp. 32258–32285, 2018.

[4] C. Jimenez, E. Saavedra, G. Del Campo, and A. Santamaria, "Alexa-based voice assistant for smart home applications," *IEEE Potentials*, vol. 40, no. 4, pp. 31–38, Jul. 2021.

[5] N. Cespedes, M. Munera, C. Gomez, and C. A. Cifuentes, "Social human–robot interaction for gait rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1299–1307, Jun. 2020.

[6] G. Muhammad and M. Alhussein, "Convergence of artificial intelligence and Internet of Things in smart healthcare: A case study of voice pathology detection," *IEEE Access*, vol. 9, pp. 89198–89209, 2021.

[7] L. Verde, G. De Pietro, A. Ghoneim, M. Alrashoud, K. N. Al-Mutib, and G. Sannino, "Exploring the use of artificial intelligence techniques to detect the presence of coronavirus covid-19 through speech and voice analysis," *IEEE Access*, vol. 9, pp. 65750–65757, 2021.

[8] S. Latif, J. Qadir, A. Qayyum, M. Usama, and S. Younis, "Speech technology for healthcare: Opportunities, challenges, and state of the art," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 342–356, 2021.

[9] D. J. Cook, G. Duncan, G. Sprint, and R. L. Fritz, "Using smart city technology to make healthcare smarter," *Proc. IEEE*, vol. 106, no. 4, pp. 708–722, Apr. 2018.

[10] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech recognition using deep neural networks: A systematic review," *IEEE Access*, vol. 7, pp. 19143–19165, 2019.

[11] R. Haeb-Umbach, J. Heymann, L. Drude, S. Watanabe, M. Delcroix, and T. Nakatani, "Far-field automatic speech recognition," *Proc. IEEE*, vol. 109, no. 2, pp. 124–148, Feb. 2021.

[12] F. Tao and C. Busso, "Gating neural network for large vocabulary audiovisual speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 7, pp. 1290–1302, Jul. 2018.

[13] S. Enarvi, P. Smit, S. Virpioja, and M. Kurimo, "Automatic speech recognition with very large conversational Finnish and Estonian vocabularies," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 11, pp. 2085–2097, Nov. 2017.

[14] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks—A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 1st Quart., 2015.

[15] A. Valkanis, P. Nicopolitidis, G. Papadimitriou, D. Kallergis, C. Douligeris, and P. D. Bamidis, "Efficient resource allocation in tactile-capable Ethernet passive optical healthcare LANs," *IEEE Access*, vol. 8, pp. 52981–52995, 2020.

[16] N. Saleh, A. Kassem, and A. M. Haidar, "Energy-efficient architecture for wireless sensor networks in healthcare applications," *IEEE Access*, vol. 6, pp. 6478–6486, 2018.

[17] S. Misra and S. Sarkar, "Priority-based time-slot allocation in wireless body area networks during medical emergency situations: An evolutionary game-theoretic perspective," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 2, pp. 541–548, Mar. 2015.

[18] Y. Dai, K. Zhang, S. Maharjan, and Y. Zhang, "Edge intelligence for energy-efficient computation offloading and resource allocation in 5G beyond," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12175–12186, Oct. 2020.

[19] R. Zhang, A. Nayak, S. Zhang, and J. Yu, "Energy-efficient sleep scheduling in WBANs: From the perspective of minimum dominating set," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6237–6246, Aug. 2019.

[20] J. Luo, J. Tang, D. K. C. So, G. Chen, K. Cumanan, and J. A. Chambers, "A deep learning-based approach to power minimization in multicarrier NOMA with SWIPT," *IEEE Trans. Signal Process.*, vol. 7, pp. 17450–17460, 2019.

[21] Y. Liu, X. Wang, J. Mei, G. Boudreau, H. Abou-Zeid, and A. B. Sediq, "Situation-aware resource allocation for multi-dimensional intelligent multiple access: A proactive deep learning framework," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 116–130, Jan. 2021.

[22] A. H. Sodhro, N. Zahid, L. Wang, S. Pirbhulal, Y. Ouzrout, A. S. Seklouli, A. V. L. Neto, A. R. L. D. Macedo, and V. H. C. D. Albuquerque, "Toward ML-based energy-efficient mechanism for 6G enabled industrial network in box systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 7185–7192, Oct. 2021.
[23] A. H. Sodhro, S. Pirbhulal, G. H. Sodhro, A. Gurtov, M. Muzammal, and Z. Luo, "A joint transmission power control and duty-cycle approach for smart healthcare system," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8479–8486, Oct. 2019.
[24] K. S. Rao and A. K. Vuppala, *Speech Processing in Mobile Environments* (SpringerBriefs in Electrical and Computer Engineering). Cham, Switzerland: Springer, 2014, pp. 103–106.
[25] A. Abrol and R. K. Jha, "Power optimization in 5G networks: A step towards GrEEn communication," *IEEE Access*, vol. 4, pp. 1355–1374, 2016.
[26] A. Gupta, R. K. Jha, P. Gandotra, and S. Jain, "Bandwidth spoofing and intrusion detection system for multistage 5G wireless communication network," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 618–632, Jan. 2018.
[27] Barbara Resch (Modified by Erhard and Car Line Rank). *Hidden Markov Models-A Tutorial for the Course Computational Intelligence*. Signal Process. Speech Commun. Lab., Graz, Austria, [Online]. Available: https://www.academia.edu/8630285/Hmm_Tutorial_Barbara_Exercises
[28] G. Li, K. Ota, M. Dong, J. Wu, and J. Li, "DeSVig: Decentralized swift vigilance against adversarial attacks in industrial artificial intelligence systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3267–3277, May 2020.
[29] Z. Yan, J. Wu, G. Li, S. Li, and M. Guizani, "Deep neural backdoor in semi-supervised learning: Threats and countermeasures," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 4827–4842, 2021.
[30] L. Zhao, Q. Wang, Q. Zou, Y. Zhang, and Y. Chen, "Privacy-preserving collaborative deep learning with unreliable participants," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1486–1500, Sep. 2019.
[31] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "A trustworthy privacy preserving framework for machine learning in industrial IoT systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6092–6102, Sep. 2020.
[32] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine learning in IoT security: Current solutions and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1686–1721, 2020.
[33] M. Hao, H. Li, X. Luo, G. Xu, H. Yang, and S. Liu, "Efficient and privacy-enhanced federated learning for industrial artificial intelligence," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6532–6542, Oct. 2020.
[34] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 4177–4186, Jun. 2020.
[35] L. T. Phong and T. T. Phuong, "Privacy-preserving deep learning via weight transmission," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 11, pp. 3003–3015, Nov. 2019.
[36] *ESSENCE: Empathic Platform to Personally Monitor, Stimulate, Enrich, and ASsist Elders aNd Children in Their Environment Project*. Accessed: Oct. 31, 2022. [Online]. Available: https://cordis.europa.eu/project/id/101016112
[37] *India Covid-19 Emergency Response and Health Systems Preparedness Project*. [Online]. Available: https://projects.worldbank.org/en/projectsoperations/projectdetail/P173836
[38] *PyXy.AI Project*. Accessed: Dec. 1, 2020. [Online]. Available: https://cordis.europa.eu/project/id/101016046
[39] *ICovid.Project*. Accessed: Aug. 31, 2022. [Online]. Available: https://cordis.europa.eu/project/id/101016131
[40] *VASCOVID Project*. Accessed: Nov. 30, 2022. [Online]. Available: https://cordis.europa.eu/project/id/101016087
[41] PORSAV. *Protecting OR Staff from Aerosolized Virus Project*. Accessed: Jul. 31, 2022. [Online]. Available: https://cordis.europa.eu/project/id/101015941
[42] *CleanAir Project*. Accessed: Oct. 31, 2022. [Online]. Available: https://cordis.europa.eu/project//id/101016174
[43] ICU4Covid. *ICU for COVID Project*. Accessed: Dec. 31, 2022. [Online]. Available: https://cordis.europa.eu/project/id/101016000
[44] COVID-X. *COVID Exponential Program Project*. Accessed: Oct. 31, 2022. [Online]. Available: https://cordis.europa.eu/project/id//101016065
[45] *An Electronic Health Record-Based Screening Tool to Support Safe Discharges of COVID-19 Patients in the Energy Department*. Accessed: Sep. 30, 2021. [Online]. Available: https://digital.ahrq.gov/ahrq-funded-projects/ehr-based-screening-tool-support-safe-discharges-covid-19-patients-emergency-department
[46] *COVIRNA Project*. Accessed: Nov. 1, 2020. [Online]. Available: https://cordis.europa.eu/project/id/101016072
[47] *ENVISION Project*. Accessed: Dec. 1, 2020. [Online]. Available: https://cordis.europa.eu/project/id/101015930
[48] METRICS. *Metrological Evaluation and Testing of Robots in International CompetitionS Project*. Accessed: Jan. 1, 2020. [Online]. Available: https://cordis.europa.eu/project/id/871252

RASHIKA RAINA received the B.E. degree in electronics engineering from the Vivekanand Education Society's Institute of Technology, University of Mumbai, India, in 2020. She is currently pursuing the M.Tech. degree in electronics and communication engineering with Shri Mata Vaishno Devi University, Katra, Jammu and Kashmir, India. She is involved in the research work in interactive speech recognition systems. She is working in Python and MATLAB for AI/ML models. Her research interest includes intelligent and interactive systems in the field of AI/ML.

RAKESH KUMAR JHA (Senior Member, IEEE) worked as an Associate Professor at SMVD University, Jammu and Kashmir, India. He is currently an Associate Professor with the Department of Electronics and Communication Engineering, Indian Institute of Information Technology Design and Manufacturing, Jabalpur, India. He is among the top 2% researchers of the world. He has published more than 61 science citation index journals articles, including many IEEE TRANSACTIONS, IEEE journal, and more than 25 international conference papers. His research interests include wireless communications, optical fiber communication, computer networks, and security issues. He is a Senior Member of GISFI, SIAM, International Association of Engineers (IAENG), and Advance Computing and Communication Society (ACCS). He is also a member of ACM and CSI, with many patents and more than 4101 citations on his credit. He has received the Young Scientist Author Award by the ITU in December 2010. He has received an APAN Fellowship in 2011, 2012, 2017, and 2018, and a Student Travel Grant from COMSNET 2012. His one concept related to the router of wireless communication was accepted by the ITU, in 2010.