NLP & AI Speech Recognition: An Analytical Review

Ayush Thakur
Amity Institute of Information Technology,
Amity University
Noida, India
ayush.th2002@gmail.com

Laxmi Ahuja
Amity Institute of Information Technology,
Amity University
Noida, India
lahuja@amity.edu

Rashmi Vashisth
Amity Institute of Information Technology,
Amity University
Noida, India
rvashisth@amity.edu

Rajbala Simon
Amity Institute of Information Technology,
Amity University
Noida, India
rsimon@amity.edu

***Abstract*— The technical developments of Natural Language Processing (NLP) and Artificial Intelligence (AI) in the area of speech recognition are covered in this study. It explains how voice recognition systems operate, as well as the numerous types, models, and applications of speech recognition. It also discusses system characteristics, speech recognition algorithms, and the function of n-grams in natural language processing. We have also looked at how deep learning and neural networks fit into the development of voice recognition technologies. The study concludes by outlining the potential applications of artificial intelligence and natural language processing in voice recognition.**

Keywords*— *Speech Recognition, Deep Learning, Neural Networks, Speech Algorithms, Artificial Intelligence

I. INTRODUCTION

Speech is a common way to interact with a human. Automatic Speech Recognition, often known as ASR, is a technique that enables a computer to understand and identify human speech. Many different applications, including voice-controlled gadgets, voice-enabled search engines, and automated customer support systems, employ ASR technologies. ASR systems are typically trained using large datasets of human speech, which are then used to create models that can accurately recognize and interpret spoken language. ASR technology is continually advancing, with new techniques and algorithms being developed to improve accuracy and reduce errors [1]. The main responsibility of such a system is to accurately interpret user input and generate the desired output. However, unlike people, computers do not possess the intelligence to understand speech. Humans are capable of distinguishing the sound of interest among a variety of concurrently audible sounds, such as speech, music, and environmental noise. This ability is known as auditory scene analysis. Computers, however, lack this ability and must rely on techniques such as speech recognition or natural language processing to process spoken language. With the help of Neural Network, Deep Learning and NLTK it can identify and track the speaking individual in a group setting that would be of great value for many applications, such as human-computer interaction, automatic meeting transcriptions, and situational awareness. This work presents a system that uses a Convolutional Neural Network (CNN) to track the speaker in an audio recording. The CNN is trained on a large dataset of multichannel audio recordings that contain one or more speakers. The system is able to track the speaker even when the audio recording contains background noise and other speakers. The results of our experiments show that the proposed system is able to achieve an accuracy of 87.5% in identifying the speaker location.

In Speech Recognition, "guessing" is used to refer to the process of matching a spoken word or phrase to a specified keyword or phrase and selecting the one that is most likely to be correct [2]. This process is usually done using acoustic or language models, which are algorithms that compare the input speech to a database of previously known speech and determine the most likely match. If a match is found, the system will then use that match as its guess for what was said. This is done in statistical analysis to produce proper speech units that can produce accurate sentences and text from spoken words. Speech encoding, speech segmentation, and noise removal are various methods used to enhance the signal quality of a recorded speech signal. Speech encoding involves compressing the signal, so it takes up less space while still conveying the same amount of information. Speech segmentation involves separating out different parts of a speech signal, such as words, phrases, and even syllables. Noise removal reduces background noise and other unwanted sound that can interfere with the clarity of a speech signal. The stage of extracting features uses a computational technique to identify the distinguishing characteristics of the voice sample [3]. Different supervised or unsupervised speech classification algorithms are then used to convert the signal into a form that is easier to process. The model can then be used to recognize the signal and make decisions. Finally, post-processing is used to remove any artifacts or distortion that results from the encoding and the classification process. Features in speech technology are measurable qualities of human speech sounds

that allow us to distinguish between words or syllables, or between speakers of different ages, genders or dialects. Common speech technology features include fundamental frequency (pitch), voice quality, spectrogram analysis (vowel lengths and formants), Mel-frequency cepstral coefficients (MFCCs), articulatory features (lip movements, tongue gestures, facial and jaw motion) and prosodic features (intonation, phrasing, pauses) [4]. These features are crucial for accurately detecting, recognizing and analysing spoken language, enabling speech technologies to automate various tasks.

II. Literature Survey

A. Speech Recognition System Working

Speech recognition is the technology that enables machines to understand and respond to human speech. It involves capturing audio input, converting it into a digital format, and analysing it using algorithms and statistical models. The goal of speech recognition is to accurately transcribe spoken words into written text, or to understand and respond to spoken commands.

The basic process of speech recognition involves three main steps (Fig 1):

Acoustic analysis is the process of converting the audio input into a digital format, and then breaking it down into individual sound units called phones or phonemes. This step is critical for the system to be able to understand the spoken words. The most common method used to perform acoustic analysis is hidden Markov models (HMMs), which are trained on large datasets of human speech to recognize the patterns of sound.

Language modelling is the process of determining the most likely sequence of words that correspond to the sound units identified in the acoustic analysis step. This is done by analyzing the context of the spoken words, and comparing them to the patterns in a pre-built language model. The language model is a large database of words and phrases, along with their probability of occurring in a given context. The more data the model is trained on, the better it will be at understanding different accents, dialects, and speaking styles.

Decoding is the final step of the process, where the system takes the output from the language model and converts it into text or an action. The system will use the most likely sequence of words to transcribe the speech into text, or interpret the spoken command and carry out the appropriate action.

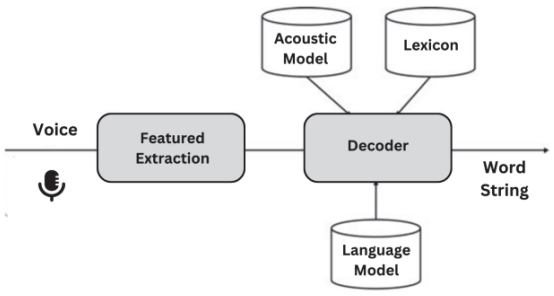

Fig. 1. Speech Recognition Model Working

One of the key factors that determine the accuracy of speech recognition is the quality of the audio input. Factors such as background noise, accent, and speaking rate can all affect the accuracy of speech recognition [5]. To overcome these challenges, many speech recognition systems use noise reduction techniques, such as filtering out background noise and echo, and adaptive learning algorithms that can adjust to the speaker's accent and speaking rate. Another important factor is the size and quality of the training dataset. The more data a speech recognition system is trained on, the better it will be at recognizing different accents, dialects, and speaking styles. High-quality datasets, which are typically annotated by human experts, are also crucial for achieving high accuracy.

B. Types of speech

There are many different types of speeches that a speech recognition bot can analyse, including:

- Spoken dialogue: This type of speech is usually in the form of a conversation between two or more people, in which the speech recognition bot will identify the words and sentences and transfer the information into text.
- Conversational spoken language: This type of speech is often used in everyday conversation, such as asking for directions or ordering food. The bot can recognize common phrases and intonation, which will help it understand the conversation better.
- Command/instruction speech: This type of speech is used to give instructions to the bot, such as "Open the window" or "Turn on the light". The bot can recognize the words and infer what action it should take.
- Audio recording speech: This type of speech is often found on podcasts or audio recordings, where the speech recognition bot will be able to identify the words and convert them into text.
- Accented speech: This type of speech can be difficult for the bot to understand, as certain dialects or accents may not be as recognizable [6]. However, the bot can adapt by using language algorithms to understand the accent better.

C. Speech Recognition Models

- Acoustic Models: Acoustic models are based on the concept that speech is composed of a series of distinct sounds. They analyze acoustic information such as the frequency, amplitude, and duration of sound waves and build a model that can recognize and distinguish words from one another. By utilizing acoustic information, acoustic models can improve the accuracy of speech recognition systems.
- Language Models: Language models are based on a set of rules that help the system to understand the context of the words used in phrases and sentences. This model uses probabilities to determine the most likely sequence of words and will consider the grammar and language structure to determine which strings of words are more likely to be used.

- Statistical Models: Statistical models are based on mathematical algorithms that are used to determine the probability of a certain string of words being used given the context of the conversation [7]. For example, the probability of a certain word or phrase being used in a certain context would be higher than the same words or phrases being used in a different context.
- Neural Network Models: Neural network models use deep learning algorithms to imitate the way humans process language. This model can understand the language context and fill in the blanks if any words are missed or not understood correctly. They are also able to identify patterns in the speech and learn continuously over time to improve the accuracy of their results.

D. Application of speech recognition system:

Speech recognition applications usually involve a speaker providing input by speaking into a microphone, which is then processed to derive meaning and intent. Some of the most common applications of speech recognition include:

- Virtual assistants: Virtual assistants, such as Siri, Alexa, and Google Assistant, use speech recognition to understand a user's query and to respond accordingly.
- Voice recognition software: Voice recognition software is commonly used in medical, legal and financial offices to accurately transcribe audio files into text documents.
- Automated call centres: Automated call centres use speech recognition to decipher customer requests and provide appropriate information.
- Automotive: Automotive companies are utilizing voice recognition as part of their infotainment systems to allow drivers to safely control the car's features.
- Security systems: Security systems use speech recognition to distinguish between authorized and unauthorized individuals.
- Handwriting and speech recognition apps: Some applications are specifically designed to recognize both handwriting and speech, allowing individuals to create and modify documents without typing.
- Disability benefits: Using closed captions, speech recognition software may translate spoken words into text, enabling hearing-impaired people to follow conversations. By using voice commands rather than typing, speech recognition can assist those with limited hand use in using computers.

E. Features of Speech Recognition system

- Natural Language Understanding: Natural language understanding (NLU) is a feature that enables machines to interpret elements of a language, such as grammar, meaning, context, and intent. NLU allows speech recognition systems to work with complete sentences and complex conversations rather than just isolated keywords [8]. Example: A NLU capable AI assistant can handle a wide range of natural language commands such as "Show me the weather for tomorrow."
- Ongoing Learning: Speech recognition systems are able to learn from their interactions with humans, improving the accuracy and reliability of speech recognition technology over time. Example: A speech recognition system can continuously gather data from its user interactions, enabling it to better understand its user's accents, verbal idiosyncrasies, and tone of voice.
- Context-Sensitive: Context-sensitive speech recognition enables machines to account for the context of a conversation, taking into account the current topic of conversation and the history of the conversation so far. This type of speech recognition can interpret user input within the context of the conversation and respond with more natural sounding responses. Example: An AI assistant equipped with context-sensitive speech recognition could interpret a phrase like "Turn off the lights" as a command to turn off the lights in the current room, rather than any lights in the entire house.
- Multilingual: Multilingual speech recognition enables machines to understand multiple languages. This feature is especially useful for applications in customer service, allowing them to accommodate more customers. Example: A multilingual AI assistant could engage with customers in their native language, understand their accents, and respond to a wide range of questions and commands.
- Adaptability: Speech recognition technologies are able to recognize different dialects, accents, and vocal patterns. This feature is especially useful for applications that involve interactions with customers from all over the world. Example: An AI assistant with adaptive speech recognition can understand and respond to customers from any part of the world without having to change the language settings of the application.

F. Speech Recognition algorithms

- Dynamic Time Warping: One of the most straightforward and widely used voice recognition techniques is dynamic temporal warping (DTW). It works by comparing the differences in time between audio samples and signals in order to identify patterns [9]. Technically, it uses an iterative method to search for an optimal alignment between the samples and signals. Applications like speaker recognition and automated speech recognition (ASR) can benefit from this technique.
- Hidden Markov Models (HMMs): HMMs are used to model individual words or phrases in speech. Each word or phrase is represented by its own HMM model [10]. The model consists of a set of "states" (phonemes) that are connected by "edges" (transition probabilities) [11]. By analyzing the acoustic features

of an audio sample, the algorithm can infer which words or phrases it corresponds to.

- Naive Bayes Classifier: A machine-learning system called Naive Bayes classifier employs feature extraction to identify patterns in speech. It uses a set of labelled audio samples to create a statistical model of each word or phrase being spoken. From there, it uses Bayesian probability to classify input audio signals according to the models it has. This algorithm is great for robust speech recognition since it can generalize from a limited set of labelled data [12].
- Neural Networks: Neural networks use a combination of artificial intelligence (AI) and deep learning algorithms to recognize features in speech. It's one of the most powerful and accurate algorithms available. The models can also be fine-tuned over time to get even better results. Neural networks are used for many applications including image recognition, natural language processing, and of course, speech recognition [13].
- Support Vector Machines (SVMs): SVMs are another machine-learning algorithm that is used for speech recognition. These algorithms use mathematical equations (vectors) to identify patterns in data. The algorithm can detect patterns in speech patterns and use them to classify input audio signals [14]. This algorithm is often used for speaker recognition due to its ability to recognize different speakers.

G. Natural Language Processing (NLP)

In the subject of artificial intelligence, or AI, natural language processing focuses on giving computers the ability to understand spoken and written language in a similar way to how humans do. Computational linguistics, or rule-based modelling of human language, combines statistical, machine learning, and deep learning models. Through the use of these technologies, computers can now completely "understand" what is being communicated or written, including the intents and mood of the speaker or writer, and translate human language into text or audio data. NLP powers a variety of computer systems, including those that translate text across languages, respond to spoken requests, and sum up voluminous volumes of text quickly—even in real-time [15]. In the form of virtual assistants, speech-to-text dictation tools, customer service bots, and some other consumer conveniences, NLP has undoubtedly been utilized by us. However, as a method of optimizing business operations, increasing worker productivity, and optimizing mission-critical business processes, the usage of NLP in corporate solutions is growing. It employs a number of NLP techniques to break down human text and voice data in order to help the computer interpret text and speech data. Some of these tasks contain the following:

- Part of speech tagging, sometimes referred to as grammatical tagging, is the process of determining a word's part of speech depending on its use and context [16].
- Co-reference resolution: Co-reference resolution is the process of determining whether two words refer to the same thing.
- Word sense identification is the process of selecting the meaning of a word from among all of its potential meanings by applying semantic analysis to determine which word is most meaningful in the given context.

Application of the NLP:

- An illustration of how frequently NLP technology is utilized is Google Translate. For machine translation to be genuinely useful, it must be capable of doing more than simply replacing words from one language with those from another. To generate content that has the same meaning and the desired impact in the target language, the translation must accurately capture the meaning and tone of the original language.
- Virtual agents and chatbots: Virtual assistant like Apple's Siri and Amazon's Alexa utilize speech recognition and natural language generation to detect patterns in voice requests and respond with the appropriate action or helpful comments.
- NLP has developed into an essential commercial tool for sentiment analysis, exposing hidden insights and data from social media networks. Sentiment analysis may be used to extract attitudes and sentiments in reaction to goods, campaigns, and events by analyzing the language used in social media postings, comments, reviews, and more [17]. This data may be used by businesses to develop new goods, start new marketing campaigns, and more.

NLP approaches are used to summaries enormous quantities of digital text and present highlights and overviews for indexes, research papers, or busy readers who lack the time to read the entire text [18]. The finest text summarization software produces summaries of pertinent context and conclusions using natural language generation (NLG) [19] and semantic reasoning.

H. Role of N-Gram in Natural Language Processing

In a text, n-grams are uninterrupted word, symbol, or token sequences. They might theoretically be referred to as the neighboring groupings of things in a document. They are pertinent for doing Natural Language Processing (NLP) operations on text data [20]. The simplest and most apparent use of n-gram modelling in natural language processing is word prediction, which is frequently used in text messaging and email. The n-gram model is frequently the sole ML technology we need to do this task because the whole use case relies on anticipating what will happen next based on what has already happened [21]. This is not the case with many other applications of n-grams, which depend on a variety of technologies combined to create a single coherent engine.

For example, n-gram models are used by auto-correct systems, such as the ones in word processors that fix grammar and spelling. It makes sense to use an n-gram model to determine whether to employ the words "there," "their," or "they're" in a phrase based on the context in which they appear. However, applying grammatical rules necessitates considering sentences as a whole, so what comes next could matter just as much as what came before.

Implementation of n-grams:

- Dataset analysis
- Feature extraction
- Training and testing separation
- Basic pre-processing
- N-gram generating code
- Unigram construction
- Bigram construction
- Trigram construction

N-grams frequently play a supportive role in what is known as "search space reduction." The n-gram will essentially look at the earlier transcribed words to limit the choices for the following words [22]. Instead of searching through a broad vocabulary, the computer "gets a concept of where to seek" and can do its duty more quickly and accurately.

III. Deep Learning and Neural Network in Speech Recognition

Neural networks are typically referred to as neurons or circuits. Currently, artificial neural networks, which comprise artificial neurons or nodes, are referred to as "neural networks". It is a network of basic building blocks known as "artificial neurons" that takes in input, modifies state as a result of that input, and produces an output [23]. The two most important components of contemporary statistically-based speech recognition systems are language modelling and acoustic modelling. The speech recognition methods are shown in the ensuing section. In the Deep Neural Network generative model, it is thought that audible characteristics are produced by a hidden Markov process that alternates between states [24].

The observation probability of the traditional HMM used in automatic speech recognition was approximated using a Gaussian Mixture Model (GMM). Even though the GMM offered a plethora of benefits, they were statistically ineffective for modelling data that are on or close to the non-linear diversity in the space [25]. For instance, modelling extremely close-to-surface sphere points virtually never needs any parameters when done with the right model class, but it does need a lot of diagonal Gaussians or a lot of full-covariance Gaussians. As a result, different models may do better than GMM at using the information that is embedded throughout a significant number of frames. On the other hand, ANN (Artificial Neural Network) may develop far better models of data and manage the data residing on or near a non-linear model more efficiently. Deep neural networks can be trained via back-propagating derivatives of something like a cost function that computes the distance in between neural network's output and the intended output. This is known as error backpropagation [26].

Error backpropagation is an algorithm used in artificial neural networks (ANNs) to calculate a gradient that is used to update the weights of the network in order to minimize error [27]. It is commonly used in supervised learning systems, where the network is provided with input and expected output, and it tries to adjust the weights to produce the expected output. The algorithm consists of propagating the error back through the network, starting from the end of the network (where the output error is known), and calculating the gradient of the error with respect to each weight. This information is then used to update the weights in order to reduce the overall error. Error backpropagation uses the chain rule to calculate how much each weight parameter impacts the cost, and then adjusts these parameters to reduce the cost (Fig 2). The process involves propagating the error from the output layers of the neural network to the input layers, and determines how to adjust the weights so that the error gets reduced. This is done using an optimization algorithm like gradient descent, which determines which direction for the weights that minimizes the change in the cost function.

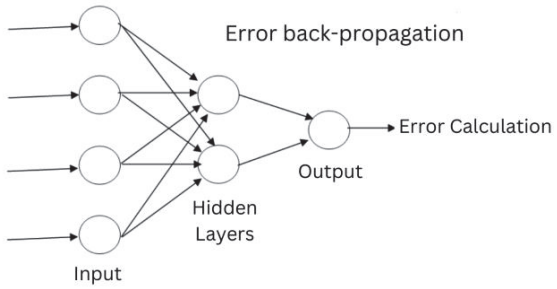

Fig. 2. Error backpropagation

In simple terms we can say, Deep learning and neural networks are two key components of AI speech recognition systems. Artificial Intelligence (AI) speech recognition systems use deep learning algorithms to analyze audio and recognize speech patterns. This process allows AI systems to learn from auditory patterns, often from human training data.

Once the AI system has analyzed the data, a neural network is used to interpret the data and generate an output, such as understanding and responding to a user's commands or queries. Neural networks are composed of linked layers of neurons that process and store information in the form of weights and signals [28]. Each neuron acts as a mini-computer, processing input and making connections to other neurons based on its experience. The combination of deep learning and neural networks allow AI speech recognition systems to accurately interpret human speech and accurately respond. The deep learning algorithms allow the system to learn from data and improve over time, while the neural networks make the connections necessary to interpret and respond to the data.

IV. Conclusion

The field of Speech Recognition is constantly evolving and improving. As technology improves, speech recognition algorithms are getting better at interpreting spoken words and commands accurately. Today, companies utilize the power of speech recognition to automate the processing of customer calls, automate the navigation of complex software, automate health care processes and provide improved accuracy in court systems and financial institutions. Advances in artificial intelligence are helping machines to understand the context of conversations and process natural language with greater efficiency and accuracy. Moving forward, we can expect to see even more applications of speech recognition that help us to interact with computers and technology in an even more natural and effective manner. A comparison of various techniques and

their accuracy shows that the employment of HMM and ANN models is a considerably more popular technique for continuous voice recognition. As we are advancing and developing our speech recognition algorithms and models in order to incorporate the contextual aspects of a conversation, the accuracy and applications of the technology will only continue to grow.

V. LIMITAIONS

a) Limited Accuracy in Understanding Natural Language: Natural language processing (NLP) is an artificial intelligence area that allows computers to comprehend the meaning of human language and analyse it to obtain valuable information. Unfortunately, NLP algorithms are currently restricted in their capacity to process natural language effectively, which means they may miss out on minor details, subtle subtleties, complicated metaphors and idioms, inferred meanings, and other difficult to analyze characteristics of natural language [29]. When attempting to comprehend and use natural language, this might lead to misunderstandings and confusion.

b) Limited Context Understanding: NLP algorithms are occasionally unable to comprehend the context of a statement, resulting in erroneous results. Context is critical for processing natural language since the meaning of a statement or phrase can vary greatly depending on the context [30]. For example, the statement "I'm hungry" can refer to either the speaker's bodily hunger or the speaker's desire for food. It is impossible to establish which interpretation is intended without context. NLP systems may be unable to analyze sentences effectively in a context-dependent way, resulting in misinterpretations and erroneous outcomes.

c) Limited Scalability: NLP techniques are frequently computationally demanding and hence difficult to scale, making them unsuitable for large-scale applications. To work correctly, NLP algorithms must handle vast volumes of data and demand substantial computational resources. As a result, they are a poor fit for applications that demand the processing of massive volumes of data in a short period of time. Furthermore, because they often process data slowly, they may be incapable of handling real-time applications.

d) Limited Domain Knowledge: NLP algorithms frequently lack domain-specific expertise, which means they may be unable to interpret sentences correctly in a domain-specific context. For example, they may be unable to comprehend the meaning of words and concepts used in a certain profession. Moreover, they may be unable to detect the distinct syntax and structure of a given dialect or language. As a result, individuals may be unable to correctly evaluate sentences including domain-specific terminology or phrases. Moreover, they are incapable of evaluating text for sentiment or tone, making them less useful for sentiment analysis applications.

e) Limited Accuracy in Speech Recognition: Although artificial intelligence (AI) voice recognition technology has advanced significantly, its accuracy remains restricted. AI voice recognition systems may struggle to distinguish between words with similar sounds and may fail to understand words uttered in a loud environment. This might result in inaccurate or missing findings, as well as a lack of comprehension on the side of the AI.

f) Limited Ability to Support Human-Like Conversations: AI voice recognition algorithms may struggle to interpret extended, complicated discussions, as well as human speech characteristics like sarcasm, comedy, and irony. AI voice recognition systems are still far from being capable of supporting human-machine interactions; generally, they can only handle simple orders and enquiries. This hinders their capacity to have long-term, meaningful discussions with humans.

VI. FUTURE SCOPE

AI has revolutionized speech recognition technology and with its rapid development, its potential use cases are growing. In the future, AI-powered speech recognition technology could further increase its accuracy and allow for real-time and comprehensive dialogue with voice-activated personal assistants, such as Alexa and Google Home. This technology could be used in transportation and navigation, such as providing voice-controlled navigation information when driving or flying. Furthermore, AI could be used to improve the accuracy of automated customer service systems, allowing them to understand conversations and be able to process personal requests. Lastly, AI can be used to detect, analyse and actuate on emotion in user requests, allowing personal assistants to better understand user needs. With its potential to improve communication, differentiate sentiment and better handle the increasing amount of information in our lives, the future of speech recognition technology could be limitless.

REFERENCES

[1] Peacocke, R. D., & Graf, D. H. (1995). An introduction to speech and speaker recognition. In Readings in Human–Computer Interaction (pp. 546-553). Morgan Kaufmann.

[2] Spanias, A. S., & Wu, F. H. (1991, June). Speech coding and speech recognition technologies: a review. In 1991., IEEE International Sympoisum on Circuits and Systems (pp. 572-577). IEEE.

[3] Chandra, E., & Sunitha, C. (2009, March). A review on Speech and Speaker Authentication System using Voice Signal feature selection and extraction. In 2009 IEEE International Advance Computing Conference (pp. 1341-1346). IEEE.

[4] Lee, L. S., & Pan, Y. C. (2009, November). Voice-based Information Retrieval—how far are we from the text-based information retrieval?. In 2009 IEEE Workshop on Automatic Speech Recognition & Understanding (pp. 26-43). IEEE.

[5] Meng, J., Zhang, J., & Zhao, H. (2012, August). Overview of the speech recognition technology. In 2012 fourth international conference on computational and information sciences (pp. 199-202). IEEE.

[6] Jiang, D., Tan, C., Peng, J., Chen, C., Wu, X., Zhao, W., ... & Deng, L. (2021). A gdpr-compliant ecosystem for speech recognition with transfer, federated, and evolutionary learning. ACM Transactions on Intelligent Systems and Technology (TIST), 12(3), 1-19.

[7] Reddy, D. R. (1976). Speech recognition by machine: A review. Proceedings of the IEEE, 64(4), 501-531.

[8] Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A review on speech recognition technique. International Journal of Computer Applications, 10(3), 16-24.

[9] Müller, M. (2007). Dynamic time warping. Information retrieval for music and motion, 69-84.

[10] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286.

[11] Ganapathiraju, A., Hamaker, J., & Picone, J. (2000, October). Hybrid SVM/HMM architectures for speech recognition. In INTERSPEECH (pp. 504-507).

[12] Saritas, M. M., & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification. International journal of intelligent systems and applications in engineering, 7(2), 88-91.

[13] Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford university press.

[14] Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567.

[15] Vogel, S., Ney, H., & Tillmann, C. (1996). HMM-based word alignment in statistical translation. In COLING 1996 Volume 2: The 16th International Conference on Computational Linguistics.

[16] Hermansky, H., Ellis, D. P., & Sharma, S. (2000, June). Tandem connectionist feature extraction for conventional HMM systems. In 2000 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 00CH37100) (Vol. 3, pp. 1635-1638). IEEE.

[17] Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. Fundamentals of artificial intelligence, 603-649.

[18] Dhuria, S., Taneja, H., & Taneja, K. (2016, March). NLP and ontology based clustering—An integrated approach for optimal information extraction from social web. In 2016 3rd international conference on computing for sustainable global development (indiacom) (pp. 1765-1770). IEEE.

[19] Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18(5), 544-551.

[20] Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. Science, 349(6245), 261-266.

[21] Grosz, B. J., Sparck-Jones, K., & Webber, B. L. (Eds.). (1986). Readings in natural language processing. Morgan Kaufmann Publishers Inc..

[22] Wolpert, D. H. (1992). Stacked generalization. Neural networks, 5(2), 241-259.

[23] Singh, A., Thakur, N., & Sharma, A. (2016, March). A review of supervised machine learning algorithms. In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 1310-1315). Ieee.

[24] Learning, D. (2020). Deep learning. High-dimensional fuzzy clustering.

[25] Singh, N., Agrawal, A., & Khan, R. A. Gaussian Mixture Model: A Better Modeling Technique for Speaker Recognition.

[26] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[27] Bohte, S. M., Kok, J. N., & La Poutre, H. (2002). Error-backpropagation in temporally encoded networks of spiking neurons. Neurocomputing, 48(1-4), 17-37.

[28] Zue, V. W. (1985). The use of speech knowledge in automatic speech recognition. Proceedings of the IEEE, 73(11), 1602-1615.

[29] Barnawi, A., Tsaramirsis, G., Buhari, S. M., & Aserey, N. A. S. (2016, March). Natural language to ontology chart. In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 1333-1337). IEEE.

[30] Hasan, I., Rizvi, S., Jain, S., & Huria, S. (2021, March). The AI enabled Chatbot Framework for Intelligent Citizen-Government Interaction for Delivery of Services. In 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 601-606). IEEE.