International Journal of Engineering and Techniques - Volume 12 Issue 2, March-April - 2026

RESEARCH ARTICLE OPEN ACCESS

Voice Command Recognition for Home Automation using Bi-Directional LSTM

S.Elakkiya[1], J.Dhivya Dharshini[2], R.Kavya[3], R.Karthick Rajan[4], P.Sivapriyan[5]

[1]*Assistant Professor, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India*
Email: elakkiya306@outlook.com

[2]*UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India*
Email: dhivyadharshinicse11@gmail.com

[3] *UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu 613006, India*
Email: kavyaravi0308@gmail.com

[4]*UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India*
Email: karthickrajan345@gmail.com

[5]*UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India*
Email: sivapriyan468@gmail.com

Abstract:

The rapid advancement of smart technologies has led to the increasing adoption of voice-based systems for home automation. Voice-controlled interfaces enable users to operate household appliances through spoken commands, providing a convenient and hands-free method of interaction. However, accurate recognition of speech commands remains challenging due to variations in speech patterns and background noise. This paper presents a voice command recognition system using a Bidirectional Long Short-Term Memory (Bi-LSTM) network. The system processes audio signals using Mel-Frequency Cepstral Coefficients (MFCC) and applies data augmentation techniques to improve robustness. The trained model classifies voice commands to control appliances such as TV, fan, and lights through a web-based interface. Experimental results show that the system achieves an accuracy of 94%, ensuring reliable performance and usability.

***Keywords*— Voice-Command Recognition(VCR), Bi-Directional LSTM(Bi-LSTM), Audio Preprocessing, Mel-Frequency Cepstral Coefficient(MFCC), Home Automation.**

I. INTRODUCTION:

The rapid advancement of smart technologies and intelligent systems has significantly increased the demand for automated home environments. Modern homes are increasingly equipped with smart devices that aim to enhance user comfort, convenience, and energy efficiency. Traditional methods of controlling household appliances rely on manual switches or mobile-based applications, which may not always provide a seamless or accessible user experience. In this context, voice-based control systems have emerged as an effective solution for enabling intuitive and hands-free interaction with home appliances.

Voice command recognition systems allow users to operate devices through natural speech, eliminating the need for physical interaction. However, accurate recognition of speech commands remains a challenging task due to variations in speech patterns, accents,

pronunciation differences, and the presence of background noise. These challenges can significantly affect the performance and reliability of voice-controlled systems, especially in real-world environments. Therefore, robust and efficient speech recognition techniques are required to ensure accurate command classification and system responsiveness.

Deep learning approaches have shown significant improvements in speech recognition tasks due to their ability to learn complex patterns from audio data. In particular, Bidirectional Long Short-Term Memory (Bi-LSTM) networks are well-suited for processing sequential data, as they capture contextual dependencies in both forward and backward directions. This paper proposes a voice command recognition system for home automation using a Bi-LSTM model. The system processes audio input, extracts Mel-Frequency Cepstral Coefficients (MFCC) features, and classifies voice commands to control appliances such as TV, fan, and lights through a web-based interface.

II. RELATED WORK:

Athul Chandran et al. [1] proposed an IoT-based smart home automation system that allows users to remotely monitor and control home appliances using internet-enabled devices such as smartphones and tablets. The system architecture includes sensors, microcontrollers, and cloud-based platforms to enable real-time communication and automation. Their approach focuses on improving energy efficiency by allowing users to schedule appliance operations and monitor energy consumption. Additionally, the system supports scalability, enabling the integration of multiple devices within a single platform. Despite its advantages, the system is highly dependent on stable internet connectivity. Issues such as network latency, packet loss, and cybersecurity threats can affect system reliability and performance.

Dong Yu et al. [2] introduced the use of Deep Neural Networks (DNN) in speech recognition systems, which significantly improved performance compared to traditional methods such as Hidden Markov Models (HMM). Their research highlights the ability of DNN models to learn complex features from large datasets, resulting in higher recognition accuracy and robustness. The study also emphasizes the importance of feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCC) in improving model performance. Furthermore, DNN models are capable of adapting to variations in speech patterns, accents, and pronunciations. However, the major limitation of this approach is the requirement for large-scale datasets and high computational resources, which increases training time and energy consumption.

Heorhii Ihor et al. [3] developed a deep learning-based system for real-time voice command recognition with a focus on reducing latency and improving system responsiveness. Their model utilizes optimized neural network architectures to achieve faster processing speeds without compromising accuracy. The system demonstrates effective performance in controlled environments with minimal background noise. However, in real-world scenarios, factors such as environmental noise, speaker variability, and microphone quality can significantly impact recognition accuracy. The study suggests the use of noise reduction algorithms and adaptive learning techniques to enhance system performance under varying conditions.

Jane Ortu et al. [4] proposed a speech recognition system based on Long Short-Term Memory (LSTM) networks, which are well-suited for handling sequential data. LSTM models can capture long-term dependencies in speech signals, making them highly effective for continuous speech recognition tasks. Their research demonstrates improved performance in recognizing complex speech patterns and maintaining contextual information. Additionally, LSTM networks can handle temporal variations in speech more effectively than traditional models. However, the training process of LSTM networks is computationally intensive and time-consuming, which may limit their applicability.

Khalil Rahman Amin [5] presented a real-time speech recognition system using advanced deep learning techniques. The primary objective of the study is to improve system responsiveness and reduce processing delays in real-time applications. The proposed system incorporates efficient feature extraction and classification techniques to achieve high accuracy. The results indicate that the system performs well in recognizing continuous speech and executing commands with minimal delay. However, the system may face challenges when dealing with multiple simultaneous users or complex voice commands, which can affect overall performance.

Leandro Filipe et al. [6] proposed a voice-controlled home automation system using machine learning algorithms. The system enables users to operate home appliances using simple voice commands, thereby enhancing usability and accessibility. This approach is particularly beneficial for elderly and disabled individuals, as it reduces the need for physical interaction with devices. The study also highlights the importance of integrating speech recognition systems with IoT devices to create a seamless automation environment. However, the accuracy of the system depends heavily on the quality and diversity of the training dataset. Inadequate or biased datasets may lead to reduced system performance and recognition errors.

Promod Kumar et al. [7] introduced a voice-activated home automation system using TinyML technology, which enables machine learning models to run on low-power embedded devices. This approach reduces dependency on cloud computing and enhances data

privacy by processing data locally on edge devices. The system is energy-efficient and suitable for real-time applications due to reduced latency. Additionally, TinyML-based systems are cost-effective and scalable for large-scale deployments. However, the limited computational capabilities of embedded devices restrict the complexity of models that can be implemented, which may affect performance in handling complex speech patterns.

Venkatesh et al. [8] developed a speech recognition system using Recurrent Neural Networks (RNN), which are effective in processing sequential data such as speech signals. The system demonstrates improved performance in recognizing continuous speech and capturing temporal dependencies. However, traditional RNN models suffer from issues such as vanishing gradient problems, which affect their ability to learn long-term dependencies. This limitation can reduce recognition accuracy in complex speech scenarios. The study suggests the use of advanced architectures such as LSTM and Gated Recurrent Units (GRU) to overcome these challenges. Furthermore, the research highlights that proper preprocessing techniques such as noise filtering and feature normalization can significantly enhance the performance of RNN-based systems. The model also benefits from large and diverse training datasets, which help in improving generalization across different speakers and accents. In addition, optimization techniques such as dropout and batch normalization can be applied to reduce overfitting and improve model stability.

III. PROBLEM DEFINITION

Despite advancements in speech recognition and smart home automation technologies, achieving accurate real-time voice command recognition remains a challenge. Variations in speech patterns, background noise, and differences in accents often reduce system performance and reliability. Additionally, dependence on internet connectivity can introduce latency and affect efficiency.

Therefore, there is a need for an efficient voice recognition system that can process speech signals in real time with high accuracy. The system should extract relevant features from audio signals and use deep learning techniques to ensure reliable performance in practical environments.

The system should handle variations in speech such as speed, tone, and pronunciation effectively. It must ensure data privacy and security while processing voice inputs. The model should be lightweight to support real-time execution on edge devices. Hence, a balance between accuracy, speed, and efficiency is required for practical implementation.

IV. PROPOSED METHODOLOGY

The proposed system is designed to implement a robust voice command recognition framework for home automation using a Bidirectional Long Short-Term Memory (Bi-LSTM) network. The system processes raw audio signals, extracts discriminative acoustic features, and performs sequence-based classification to identify user commands. By integrating signal processing techniques with deep learning models, the system ensures high accuracy and robustness under varying environmental conditions. The complete workflow consists of audio acquisition, preprocessing, feature extraction, sequence modeling, classification, and command execution.

The audio signals are processed and transformed into Mel-Frequency Cepstral Coefficients (MFCC), which effectively represent the spectral characteristics of speech. The extracted features are then fed into the Bi-LSTM model to capture temporal dependencies in both forward and backward directions. The final output is obtained using a Softmax classifier, and the predicted commands are used to control home appliances through a web-based interface.

a. Audio Data Acquisition

The initial stage of the proposed system focuses on the acquisition of speech signals from the user through a microphone interface. The input audio is captured at a sampling frequency of 16 kHz, which is widely adopted in speech processing applications to ensure an optimal balance between computational efficiency and signal fidelity. The analog voice signal is converted into a digital representation using an analog-to-digital conversion process and stored as a discrete-time waveform.

To maintain consistency across samples, the system restricts the duration of each audio input to a fixed time window. This ensures uniformity in subsequent processing stages and facilitates efficient batch training of the deep learning model. The acquired audio signals may contain variations due to environmental noise, speaker differences, and recording conditions; however, these variations are addressed in later preprocessing stages.

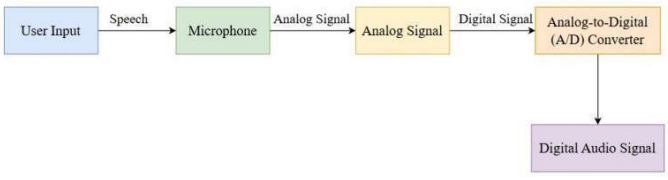

Fig1-Voice Data Acquisition

The captured digital audio serves as the primary input for the speech processing pipeline and forms the basis for feature extraction and sequence modeling in the proposed Bi-LSTM framework.

b. Audio Preprocessing

The audio preprocessing stage plays a crucial role in enhancing the quality and consistency of the captured speech signals before feature extraction. Raw audio signals obtained from the acquisition stage often contain noise, amplitude variations, and temporal inconsistencies that can negatively affect model performance. Therefore, several preprocessing operations are applied to standardize the input data.

Initially, the audio signal is resampled to a fixed sampling rate of 16 kHz to ensure uniformity across all input samples. Amplitude normalization is then performed to scale the signal within a consistent range, reducing the impact of varying recording volumes. To improve model robustness, data augmentation techniques such as additive noise injection and time shifting are applied. Noise augmentation simulates real-world environmental conditions, while time shifting introduces temporal variations, enabling the model to generalize better to unseen data.

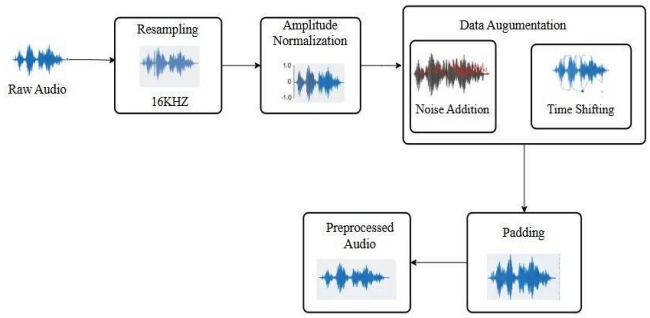

Fig2-Audio Preprocessing

Furthermore, each audio sample is adjusted to a fixed duration using padding and truncation techniques. If the signal length is shorter than the predefined duration, zero-padding is applied; otherwise, excess samples are truncated. This ensures that all input sequences have uniform dimensions, which is essential for batch processing in deep learning models.

These preprocessing steps collectively enhance the signal quality, reduce variability, and improve the reliability and accuracy of the subsequent feature extraction and classification stages.

c. Feature Extraction using MFCC

Feature extraction is a critical stage in speech recognition systems, as it transforms raw audio signals into a compact and informative representation suitable for machine learning models. In the proposed system, Mel-Frequency Cepstral Coefficients (MFCC) are employed as the primary acoustic features due to their effectiveness in capturing perceptually relevant characteristics of speech signals.

Initially, the preprocessed audio signal is segmented into short overlapping frames to preserve temporal information. Each frame is then transformed from the time domain to the frequency domain using spectral analysis. The resulting spectrum is passed through a set of Mel-scaled filter banks, which mimic the human auditory system by emphasizing frequencies that are more perceptible to the human ear. The logarithm of the filter bank energies is computed to compress the dynamic range of the signal.

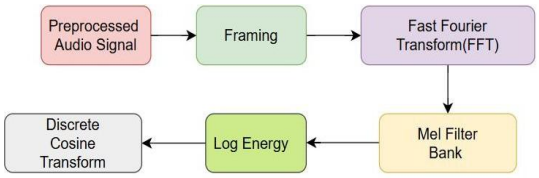

Fig3-Feature Extraction using MFCC

Subsequently, the Discrete Cosine Transform (DCT) is applied to the log Mel spectrum to obtain a set of decorrelated coefficients known as MFCCs. These coefficients represent the short-term power spectrum of the speech signal in a compact form. In this work, a fixed number of MFCC features are extracted for each frame and arranged as a time-series feature matrix.

To ensure consistency and improve model convergence, the extracted MFCC features are normalized across the dataset. The resulting feature sequences effectively capture both spectral and temporal properties of speech and are used as input to the Bidirectional LSTM model for voice command classification.

d. Bi-LSTM based Voice Command Recognition

The Bidirectional Long Short-Term Memory (Bi-LSTM) network is employed in this work to effectively model temporal dependencies present in speech signals for accurate voice command recognition. Unlike conventional Recurrent Neural Networks (RNNs), LSTM networks are capable of learning long-term dependencies using memory cells and gating mechanisms, thereby overcoming issues such as vanishing and exploding gradients. In addition, the

bidirectional architecture processes the input sequence in both forward and backward directions, enabling the model to capture contextual information from past and future time steps simultaneously.

In the proposed system, the extracted MFCC feature vectors are structured as sequential time-series inputs and fed into a multi-layer Bi-LSTM network. The first Bi-LSTM layer is configured with a higher number of hidden units to capture complex temporal patterns, followed by a second Bi-LSTM layer that refines the learned representations. Batch normalization is applied after each Bi-LSTM layer to stabilize training and accelerate convergence, while dropout regularization is incorporated to reduce overfitting and improve generalization performance.

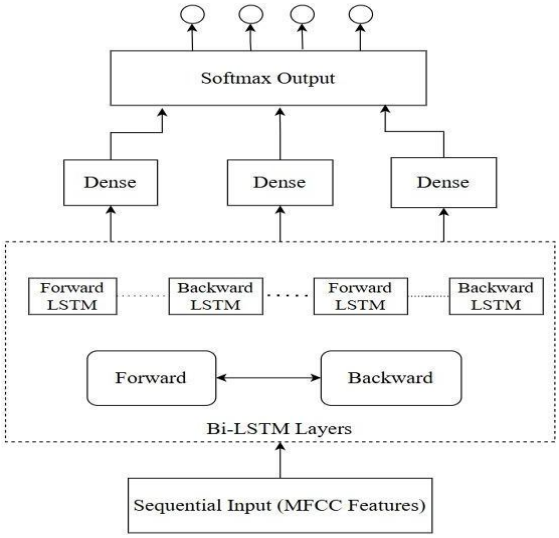

Fig4-Bi-LSTM based Voice Command Recognition

The output from the Bi-LSTM layers is passed through fully connected dense layers with ReLU activation to learn higher-level feature representations. Finally, a Softmax classification layer is used to predict the probability distribution over predefined voice command classes. The model is trained using the AdamW optimizer, which combines adaptive learning rate optimization with weight decay for better generalization. The overall architecture enables robust recognition of spoken commands under varying conditions, making it suitable for real-time home automation applications.

e. Command Processing Home Automation Control

The command processing and home automation control module is responsible for interpreting the predicted output from the Bi-LSTM model and executing the corresponding actions on connected devices. Once the Softmax layer generates the probability distribution over predefined command classes, the command with the highest probability is selected as the final predicted label. This predicted command is then forwarded to the command processing unit for further validation and execution.

The command processor performs semantic interpretation and validation of the recognized command to ensure correctness and consistency. It maps the predicted label to a predefined set of control actions such as turning appliances ON or OFF. In addition, the module incorporates basic error handling mechanisms to manage invalid or ambiguous commands. If the confidence score of the prediction is below a predefined threshold, the system may reject the command or request user feedback to improve reliability.

Following successful validation, the processed command is transmitted to the home automation controller, which acts as an interface between the software model and physical devices. The controller generates appropriate control signals and communicates with smart appliances through web-based interfaces or IoT protocols. The system supports operations such as switching lights, fans, or other devices based on user voice commands.

f. Output Generation

The output generation of the Voice Recognition for Home Automation system begins with capturing the user's voice through a microphone, followed by preprocessing steps such as noise reduction, normalization, and segmentation to improve audio quality. The system then extracts important speech features using MFCC, which are fed into the Bidirectional LSTM model for deep sequence analysis.

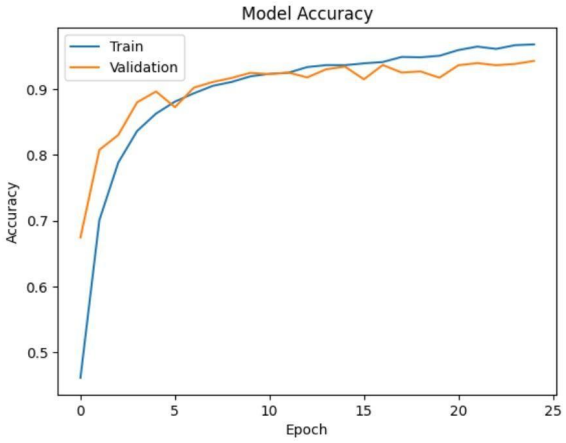

Fig5-System Accuracy Analysis

This model processes the input in both forward and backward directions, enabling better understanding of context, tone, and variations in speech patterns. Based on this analysis, the system accurately classifies the spoken command into predefined actions like switching lights or fans ON/OFF, controlling appliances. The recognized command is then converted into a machine-readable signal and sent to the home automation controller via a web-based interface, where the output is also visually displayed to the user for confirmation. Additionally, the system provides real-time response, supports continuous listening, minimizes recognition errors, adapts to different voice inputs, and ensures reliable performance even in slightly noisy environments, making the overall output generation process efficient, fast, and user-friendly for smart home control.

V. CONCLUSION

The project titled "Voice Command Recognition for Home Automation using Bi-Directional LSTM" was successfully developed to create a voice-based smart home automation system that allows users to control household appliances through simple voice commands. In this system, the user's voice input is captured through a microphone and processed through stages such as audio preprocessing and feature extraction to obtain meaningful speech features. These features are then given to a Bi-Directional Long Short-Term Memory (Bi-LSTM) model, which analyzes the speech sequence in both forward and backward directions to better understand the context of the command.

The proposed model achieved an accuracy of 94%, showing that the system can recognize voice commands with high reliability. Once the command is identified, the system sends the appropriate signal to control home appliances such as lights and fans. This system improves convenience by enabling hands-free control and reducing the need for manual switches or mobile applications.

The proposed system can be applied in smart homes, assistive living environments for elderly and disabled individuals, and smart workplaces where voice-based automation can improve efficiency. In the future, the system can be enhanced by supporting multiple languages, larger datasets, and integration with IoT platforms for better automation and remote control. Overall, the project demonstrates that Bi-Directional LSTM is an effective approach for accurate voice command recognition in home automation systems.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to their project guide and faculty members of the Department of Computer Science and Engineering for their continuous support, valuable guidance, and encouragement throughout this research work.

The authors also extend their thanks to their external guide, Mr. P.Boopathy Pandi, from Monzha Research Lab, for providing valuable insights, technical guidance and necessary resources that contributed to the successful completion of this work.

Additionally, the authors acknowledge the use of publicly available datasets and tools that supported the development and evaluation of the proposed system.

REFERENCES

[1] A. Chandran, A. Anu, and V. Raj, "IoT Based Smart Home Automation System," *IEEE*, vol. 7, no. 3, pp. 45–50, 2019.

[2] D. Yu, G. Hinton, and L. Deng, "Application of DNN for Voice Recognition System," *IEEE Transactions on Audio, Speech and Language Processing,* vol. 26, no. 5, pp. 1020–1030, 2018.

[3] I. Ihor, H. Heorhii, and O. Oleksii, "Application of Deep Neural Network for Real-Time Voice Command Recognition," *IEEE Access,* vol. 10, pp. 55678–55687, 2022.

[4] J. Oruh and S. Viriri, "Speech Recognition using Long Short-Term Memory," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 33, no. 4, pp. 1205–1215, 2022.

[5] K. Amin and R. Khalil, "Real-Time Speech Recognition using Deep Learning Techniques," *IEEE Access,* vol. 7, pp. 135245–135255, 2019.

[6] L. Filipe and R. Silva Peres, "Voice Controlled Home Automation using Machine Learning," *IEEE Internet of Things Journal,* vol. 8, no. 6, pp. 4521–4530, 2021.

[7] P. Kumar, S. Bhudhani, and T. Malche, "Voice Activated Home Automation using TinyML," *Springer Journal of Ambient Intelligence and Humanized Computing,* vol. 14, no. 2, pp. 987–995, 2025.

[8] V. Venkateswarlu, V. Kumar, and V. Jayasri, "Speech Recognition using Recurrent Neural Network," *International Journal of Scientific and Engineering Research,* vol. 8, no. 10, pp. 1120–1125, 2017.

[9] A. Graves, N. Jaitly, and A. Mohamed, "Speech Recognition with Deep Recurrent Neural Networks," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* pp. 6645–6649, 2013.

[10] S. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 28, no. 4, pp. 357–366, 1980.

[11] H. Sak, A. Senior, and F. Beaufays, "Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* pp. 338–342, 2014.
[12] T. N. Sainath and B. Li, "Deep Convolutional Neural Networks for Large-Scale Speech Tasks," *IEEE Signal Processing Magazine,* vol. 29, no. 6, pp. 29–39, 2012.
[13] F. Eyben, M. Wöllmer, and B. Schuller, "OpenSMILE: The Munich Versatile and Fast Open-Source Audio Feature Extractor," *Proceedings of the ACM Multimedia Conference*, pp. 1459–1462, 2010.
[14] A. Graves and J. Schmidhuber, "Framewise Phoneme Classification with Bidirectional LSTM Networks," *Proceedings of the International Joint Conference on Neural Networks (IJCNN),* pp. 2047–2052, 2005.
[15] J. L. Gauvain and C. H. Lee, "Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains," *IEEE Transactions on Speech and Audio Processing,* vol. 2, no. 2, pp. 291–298, 1994.