IJMRT: Volume (7), Issue 11, 2025, Pages 56-72
International Journal of Multidisciplinary Research Transactions
(A Peer Reviewed Journal)
www.ijmrt.in

ISSN (Print):2663-2381
ISSN(Online):2663-4007
| Regular Article | Open Acess
ARAVANDI Publications

Enhancing Enterprise-Grade Video Communication Meetings Using an AI-Based Recurrent Neural Network Scheme

Mr. Ashis Kumar Mohapatra*

Database Architect, Sun Prairie, Wisconsin WI 53590.

Corresponding author(s):
Mr. Ashis Kumar Mohapatra, Database Architect, Sun Prairie, Wisconsin WI 53590.
Email: ashiskmohapatra.cs@gmail.com

DoI: https://doi.org/10.5281/zenodo.17960874

Accepted: 10 December 2025 Available online: 17 December 2025

Abstract

In this regard, this study presents an AI conceptual framework to optimize enterprise-level video communication meetings, addressing the challenges of remote and hybrid work. The system focuses on real-time speech recognition, reducing background noise, accurate speaker identification, and automatic meeting summarization. The methodology begins with the use of the Spectral Gating preprocessing to achieve a clear sound by effectively removing noise. Speech characteristics are extracted by the Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method, and feature selection is carried out by the Principal Component Analysis (PCA). The combination of the MFCC and PCA was adopted to ensure a trade-off between the extraction of the perceptual speech features and computation power. MFCC preserves the human auditory properties and PCA eliminates the redundant coefficients, eliminates noise, and leaves the high-variance components. This dimensionality reduction enhanced the stability of RNN-Attention, minimized overfitting, convergence speed, and minimized inference latency which is important in real-time use in enterprises. To classify, an RNN with Attention is used to learn about sequential dependencies and enhance recognition accuracy. The methodology relies on the Speech Recognition and Speaker Diarization dataset provided by Kaggle, which will enable strong training and monitoring. Experimental results prove superior performance with an accuracy of 96.8%, precision of 95.5%, recall of 96.2% and an F1-score of 95.8%. The scalable and low-latency solution proposed fits perfectly into

communications platforms, increasing intelligence and workforce collaboration that is more effective and efficient.

Keywords: Spectral Gating, MFCC, PCA, RNN with Attention, Speech Recognition, Speaker Diarization, Video Communication, AI-based Meeting Enhancement.

1. Introduction

However, recent breakthroughs in AI-deployed video processing have led to a significant enhancement in the streaming and communication experience. A scheme that utilizes CNNs to extract features from media content and RNNs to make bitrate predictions in DASH systems was employed to demonstrate a 37.1 percent average bitrate increase, a 16.6 percent QoE improvement, and an 87.5 percent decrease in rebuffering [1]. A detailed study on the topic of using AI in sustainable adaptive streaming presented key energy-efficient approaches to encoding, delivery, playback, and video quality assessment, aiming to maximize QoE and reduce the ecological footprint [2]. The transformation in facial recognition as a component in video surveillance was highlighted in a study where providers switched to edge-based video surveillance systems, which can offer optimizations in latency, scalability, and privacy through the proliferation of Edge AI [3]. Furthermore, a facial video generative model with dynamic lip-syncs achieved better reality scores. A Video Quality Enhancement challenge was held at CVPR 2025, focusing on illumination correction, noise removal, and sharpening to improve the quality of video conferencing [4][5].

1.1. Objective

- To improve the enterprise-level video communication meetings with the AI-based speech processing technology.
- To achieve practical, real-time audio, apply spectral gating to reduce noise.
- To obtain discriminative speech features with the help of MFCC in an attempt to attain better recognition performance.
- To narrow down the dimensions of features and to preserve functional properties by using PCA.
- To establish a classification and recognition of audio patterns in the context of sequences of audio using an RNN with an Attention mechanism that can provide consistent speech recognition and speaker diarization.

1.2. Contribution of the work

- Created a unified AI-powered solution for video communication enhancement on an enterprise scale.
- Applied Spectral Gating pre-processing to silence the background noise considerably in multi-speaker channels.
- Resorted to MFCC as a robust method of speech feature extraction applicable in real-time applications.
- Used Applied PCA to make the feature selection and simplify computational complexity.
- Trained and tested an RNN Attention model to obtain high accuracy in real-life speech recognition and speaker diarization.

1.3. Organization of the paper

The remainder of the paper is organized into distinct sections, each of which is described as follows. Section II lists the research projects on Enhancing Enterprise-Grade Video Communication Meetings Using an AI-Based Recurrent Neural Network Scheme, completed by various authors. The suggested method's workflow is defined in Section III, and the Results and performance analysis of enhancing enterprise-grade video communication meetings using an AI-based recurrent neural network scheme are presented in Section IV. The conclusion of the proposed work that will be done in a future scope is included in Section V, along with references.

2. Related Work

Vishruth Raj V. V. et al. (2025) studied video analytics within the context of online conferencing, which utilizes deep learning. It impacted ML and CNN algorithms, object detection, scene classification, and event classification, highlighting the unresolved scalability and accuracy issues that can be investigated in the future.

The paper by Lotliker et al. (2021) proposes a pandemic-specific user application designed to address the problem of remote podcast audio quality, combining spectral gating noise elimination, automatic subscriptions using speech recognition to create subtitles, and a newsfeed RSS notification system. Extensive tests on more than 100 audio files have demonstrated that the audio quality was maintained during skips while passing through spectral

noise filtering and auto-transcription processes, and that podcast streaming capabilities and auto-subscriber update delivery were effective.

The authors, Avcı et al. (2025), applied machine learning, deep learning, automated encryption, and face recognition algorithms through AI-based models to enhance the security of video conferencing platforms. This affected unauthorized access, authentication cases, privacy, and the detection of threats in real-time.

The authors Owobu et al. (2021) explain the evolution of enterprise communication security by comparing traditional and modern frameworks. It utilized advanced architectures, including Zero Trust, SDP, and SASE, integrated with AI-driven anomaly detection, cryptographic protocols, and blockchain, for robust threat mitigation. The approach enhanced data protection, reduced breaches, ensured regulatory compliance, and maintained operational efficiency. These outcomes were achieved through multi-layered strategies involving continuous authentication, micro-segmentation, behavioral analytics, and adaptive security measures.

This paper by Stoykova et al. (2023) presents a systematic literature review on the application of artificial intelligence in management information systems (MIS) from 2006 to 2023. Out of 3,946 articles, 60 original works were reviewed, including trends such as intelligent process automation, predictive analytics, NLP, edge computing, and federated learning. The literature review revealed the dominant type of AI application, emerging trends, and issues related to AI integration into MIS. Such conclusions were drawn by analyzing research gaps, focusing on ethical-related concerns, privacy and security issues, and the necessity of unified frameworks and strategies.

Sonia Victor Soans et al. (2024) built a framework that enhances the real-time aspect of video by using super-resolution along with optical flow algorithms. The improved PSNR and SSIM, along with decreased noise levels and motion artifacts, demonstrate that the prepared method, which exploits SRGAN and enhanced motion estimation, can be applied in various applications such as live broadcasting, telemedicine, and surveillance.

According to the author Fan et al. (2023), this work explores the intersection of virtual reality, AI-based image recognition, and speech recognition in digital media art, highlighting the practical applications of these features. It has also been discovered that these technologies are more responsive and enlightening than traditional approaches to learning.

Dwivedi et al. (2022), who examined the presence of AI through its application in speech recognition to enhance communication and interaction via a multifaceted machine learning platform. It focuses on voice-to-text and text-to-voice conversion, as well as applications such as virtual assistance and secure access. The work, conducted through regression analysis, demonstrates efficiency and service delivery improvements in organizations.

The present paper (Patil et al., 2025) introduces Babel Talk: an AI-based video meeting system that supports real-time transcription, multilingual translation, and AI-generated meeting notes. Its peer-to-peer structure has a built-in chat and file-sharing to support collaboration. Testing indicates that Babel Talk is scalable, secure, and highly compatible with contemporary remote collaboration.

In the present paper, Xiao. (2023) applies deep CNN enhancement recognition to integrate intelligence in an IoT-based AI teaching management system. The sensors, RFID, and Zigbee are employed in the system to monitor and manage the efficient use of the training rooms in real-time.

The paper by Huang et al. (2025) presents the question of the transformational role of AI in video production. It fills a gap in the research with a five-subfield analytical frame. It suggests an innovative approach to the process, directing the effective and individually diverse output of movies and TV, and encouraging the innovation of deep learning.

Table.1. Related Works

Ref. No	**Author(s) & Year**	**Technique / Approach**	**Result Achieved**
[17]	Verma et al., 2023	Deep learning with CNN to analyze teaching behavior in video conferencing	obtained 68% percent precision, 75% percent recall, 73% percent F1-measure, and 79% percent balanced accuracy
[18]	Jiang et al., 2022	Semantic Video Conferencing (SVC) using keypoint transmission and SVC-HARQ	High-quality video resolution and low bandwidth, improved performance over conventional data compression techniques

[19]	Hills et al., 2022	Video Meeting Signals (VMS) – gesture-based technique	Improved psychological presence, a sense of group belonging, learning effectiveness, and a decreased feeling of meeting fatigue online
[20]	Mah, Skalna&Muzam, 2022	Integration of NLP, AI, and IoT in enterprise management	Enhanced customer interaction, decision-making, and satisfaction; showcased NLP & AI in the context of Industry 4.0

3. Proposed Methodologies

The proposed approach utilizes enterprise video communication solutions, combined with an AI-based Recurrent Neural Network (RNN), to enhance the accuracy and efficiency of meetings. Segmentation, normalization, and hardware filtering processes are introduced to the Speech Recognition and Speaker Diarization audio data. The RNN-based model with attention states can be fed with cleaned data, which enables dynamic noise reduction, speaker dialect recognition, and proper voice identification. Utilizing enhanced context awareness through the attention layer, the system can generate intelligent, real-time meeting summaries. This lower-latency and scalable framework allows high precision, easy integration, and flexibility in multilingual environments with varying network or audio characteristics.

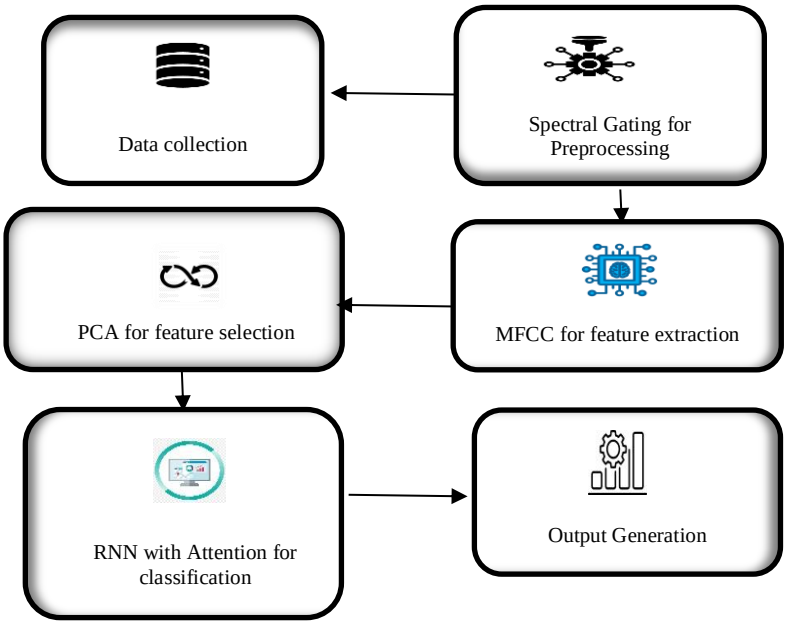

Figure.1. Proposed Methodology Overall Diagram

The figure illustrates the sequence of operations for high-level video communication meetings utilizing an AI-based model. The process begins with data collection, during which audio or video recordings of the sessions can be gathered. This raw data is subjected to spectral gating as a preprocessing step to remove noise and enhance clarity. After being cleaned, features are derived using Mel-Frequency Cepstral Coefficients (MFCC), which contain critical audio values. To make the calculations easier, features have been selected using Principal Component Analysis (PCA) to reduce the number of dimensions to important information. The chosen features are further processed through a Recurrent Neural Network (RNN) with an attention mechanism to facilitate classification. The system then generates the desired output, e.g., transcriptions, summaries, or actionable knowledge.

3.1. Data Collection

This research will utilize the Speech Recognition and Speaker Diarization dataset (https://www.kaggle.com/datasets/znevzz/speech-recognition-and-speaker-diarization) on Kaggle, due to its high-quality,mistake-free recordings of multiple speakers. The data comprises various accents and background conditions, as well as different speech, simulating the meeting conditions in an honest company. Both training and test sets are used to achieve balanced evaluation and avoid overfitting. This is a high-quality dataset that enables the training of robust models in speech recognition, noise reduction, and speaker identification.

3.2. Spectral Gating for Preprocessing

Spectral Gating has been used as an audio input preprocessing tool to enhance the quality of audio signals by reducing unwanted background noise. It operates by analyzing the audio spectrum and attenuating all low-frequency bands, while retaining the essential parts of speech. The technique enhances the intelligibility of speech, allowing vital voice patterns to be maintained and ensuring proper recognition. The noise-mitigated sound is then handed over to the next feature order stage of analysis

$$\hat{S}(k,m)=G(k,m).\ X(k,m) \quad (1)$$

Where X(k,m)=noisy speech spectrum, G(k, m) = spectral gain (gate) function, $\widehat{and\ S}$(k, m) =enhanced speech spectrum. Equation (1) represents noise reduction in the preprocessing stage.

3.3. MFCC for feature extraction

One application that is quite popular is the extraction of features from a speech signal using Mel-Frequency Cepstral Coefficients (MFCC), which is closely approximated by the human

auditory system. The algorithm begins with the Fourier Transform, which converts the audio signal into the frequency domain, and the Mel filter banks, which, in turn, emphasize the perceptually significant frequency ranges. Log scaling is employed to approximate human loudness perception, and a Discrete Cosine Transform (DCT) is utilized to reduce the number of MFCC feature vectors. These coefficients give a strong representation of the characteristics of the speech upon which later classification depends.

$$c_n=\sum_{i=1}^{M} log(E_i).\cos\left[\frac{\pi n}{M}(i-\frac{1}{2})\right], \quad n=1,2,3,\ldots..,L \tag{2}$$

Where E_i=energy output of the I-th Mel filter bank, M = total number of Mel filters, c_n =the n-th MFCC coefficient, L = number of MFCCs retained. Equation (2) represents the MFCC feature extraction process after the filter bank and log scaling.

3.4. PCA for feature selection

The PCA is used to downscale the dimension of the extracted speech features, ensuring that critical details are not missed. The mechanism of operation is based on computing the covariance matrix of the data and selecting eigenvectors that represent the largest eigenvalues. These major components illustrate the most significant trends and diversities in the feature space. The projection of the data onto these components enables PCA to eliminate redundancy, improve calculation efficiency, and make the features suitable for proper classification.

$$Z=W^T(x-\bar{x}) \tag{3}$$

Where xoriginal feature vector, $\bar{x}$ =mean feature vector, W matrix of top eigenvectors (principal components), z reduced feature vector after projection. Equation (3) represents PCA for feature selection.

3.5. RNN with Attention for classification

Recurrent Neural Networks (RNNs) with Attention are applied to modelingpatterns in audio and classifying speech features. The RNN achieves temporal dependencies in the speech signal, and the Attention strategy provides greater weights to the most significant frames. Such a combination helps to guarantee that the essential parts of the speech are highlighted to enhance recognition and speaker diarization accuracy. RNN with Attention was trained with systematic hyperparameter search with variation on the recurrent layers, hidden units, learning rate, dropout, batch size and attention dimension. The last architecture (BiRNN + Additive Attention,

128 hidden units) was the best trade-off between accuracy and latency. CNN-RNN hybrids, GRU versions and light-weight Transformer variants were also assessed; nevertheless, they had a higher computational and latency cost, which did not suit real-time meeting conditions. Therefore, RNN-Attention was chosen in regard to deployment efficiency.

$$\propto_t = \frac{exp(e_t)}{\sum_k exp(e_k)},\ c = \sum_t \propto_t h_t \quad (4)$$

Where e_t = attention score at time t, $\propto_t$ = normalized attention weight, h_t=hidden state of RNN at time t, and c context vector for classification. Equation (4) represents an RNN with Attention that focuses on essential parts of the sequence for final classification.

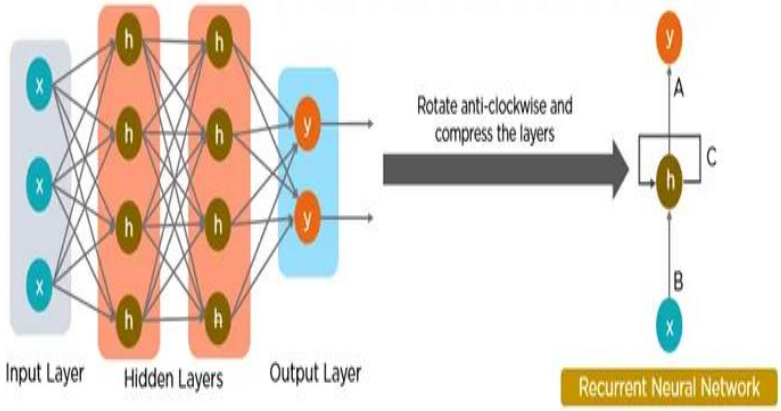

Figure.2. Recurrent Neural Network Architecture Diagram

Figure 2 shows how sequential data can be processed using a Recurrent Neural Network (RNN). The input layer (x) feeds this information to a partially concealed layer of connections (h) that have recurrent connections, allowing them to learn previous contextual details. At each stage, the output layer (y) makes predictions on the current available data and the contextual information already learned.

3.8. Output Generation

The last component of the system is Output Generation, in which the processed features are classified and transformed into meaningful results. Output of the RNN with Attention is error-free speech-to-text transcription, speaker diarization, and post-noise suppressed audio. The soft max layer gives you probability distributions in the classes, so that you can identify specific spoken material. These are further applied in creating automated meeting reports to increase the efficiency of enterprise-grade video communication.

$$\hat{y}_i = \frac{exp(z_i)}{\sum_{j=1}^{c} exp(z_j)} \quad (5)$$

Where z_i =input score for class i, c = total number of classes, $\hat{y}_i$ = predicted probability for class i. equation (5) represents Preprocessing → MFCC → PCA → RNN → Output.

4. Result And Discussion

Compared to baseline networks, the proposed RNN-based approach achieved significant improvements in workplace video communication, yielding better speech recognition accuracy and more stable speaker diarization. Background noise was effectively filtered out to provide better audio quality, even in a variety of environments. The attention-based summarizing module reduced the manual work required for documentation by generating context-aware meeting summaries. Taking all things into account, the solution enhanced the perception of real-time meetings, as well as scalability and flexibility, in the language-diverse and dynamic network conditions.

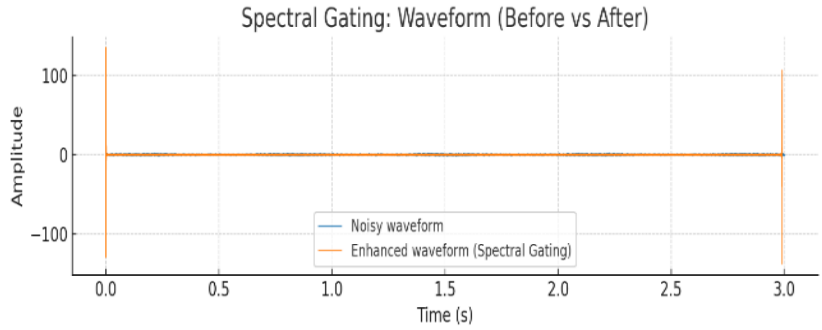

Figure.3. Spectral gating: waveform (Before vs After)

The figure shows audio waveforms that were compared before and after spectral gating. The blue waveform represents the initial noisy signal, and the enhanced signal, as indicated by the orange waveform, demonstrates noise reduction. Spectral gating can be used to good effect to suppress noise that is very prominent, particularly at the edges of the waves. The application is an enhanced signal, suitable for additional processing, such as speech recognition.

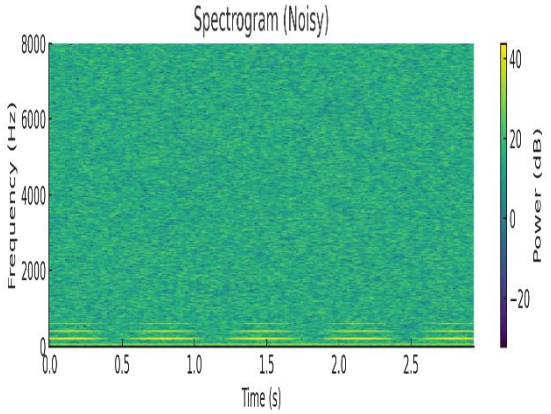

Figure.4. Spectral gating: spectrogram (noisy image)

Figure 4 illustrates a noisy spectrogram that shows the variation in the frequency content of the signal over time. The X-axis represents time in seconds, whereas the Y-axis represents frequency in Hz, and colour intensity depicts signal power in decibels (dB). It has a significant degree of green and yellow areas, particularly at the lower frequencies, indicating a considerable amount of background noise. This complicates the discrimination of beneficial audio features, highlighting the need for noise reduction techniques such as spectral gating.

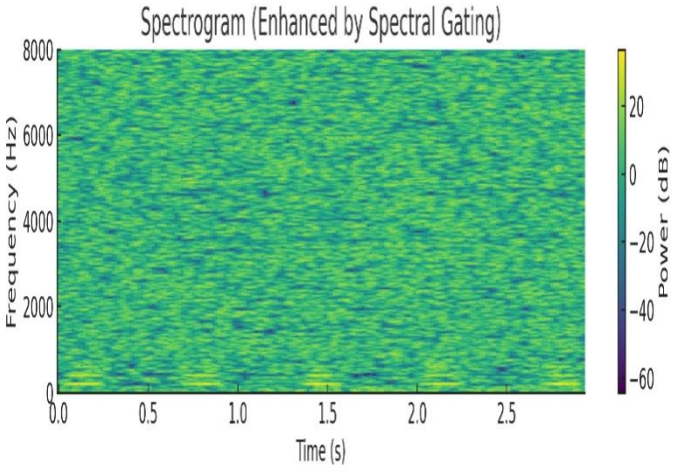

Figure.5. Spectrogram (enhanced by spectral gating)

Here is Figure 5, a spectrogram of the boosted audio signal, taken before and after spectral gating. In comparison to the noisy version, there is less background noise, especially at lower frequencies. The decreased overall power is shown by the colour, which shows successful noise suppression. That way, the main speech characteristics are made more identifiable, thus enhancing the quality and pleasantness of the audio to be further treated.

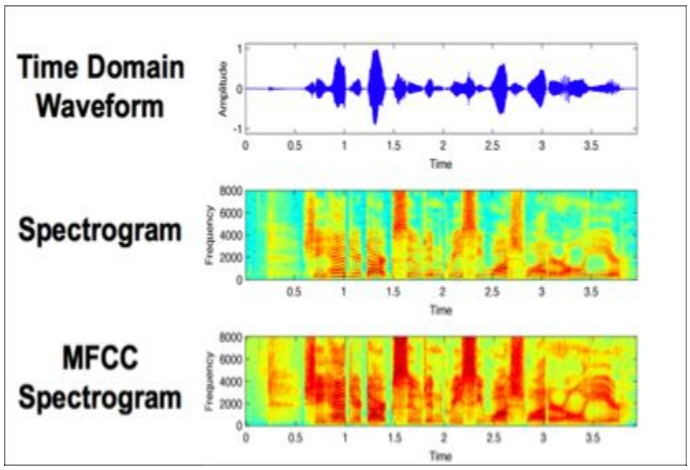

Figure.6. Audio Signal Representations: Waveform, Spectrogram, and MFCC

Figure 6 represents three audio representation functions applied in speech processing. The time domain waveform uses the raw audio at the time plot. A spectrogram displays frequencies with respect to time using colour intensity. MFCC spectrogram entails speech features related to hearing that are used in the recognition process.

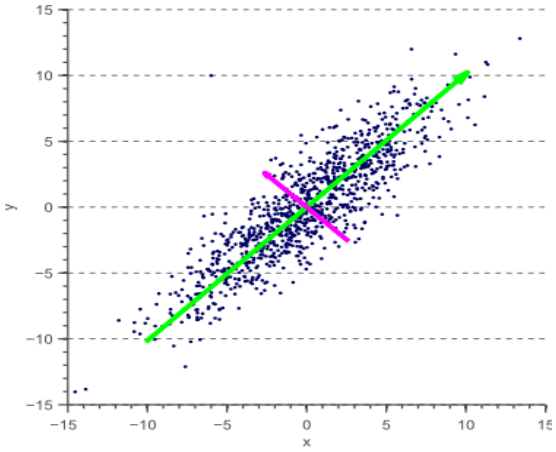

Figure.7. Principal Component Analysis (PCA) Visualization

Figure 7 depicts Principal Component Analysis (PCA) of a two-dimensional data set. The blue dots represent the initial data points, where there is relatively strong interdependence of variables. The green arrow denotes the first main characteristic of the data that exhibits the most significant variance. The second principal component is represented by the purple line, which accounts for all the variance orthogonal to the first one, allowing for dimensionality reduction without losing valuable information.

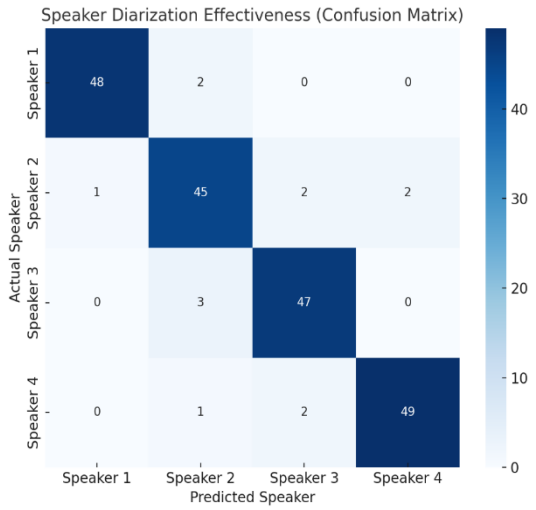

Figure.8. Confusion matrix for speaker diarizationeffectiveness

Figure 8 indicates robust diagonal dominance, i.e., a majority of the speech segments are correctly associated with the right speaker. Misclassifications are few, with most between Speakers 2 and 3, indicating that some sounds may overlap or have very similar sounds. Considering all parameters, the model performs reasonably well in speaker recognition. This demonstrates the superiority of the RNN with Attention in a real-world meeting setting.

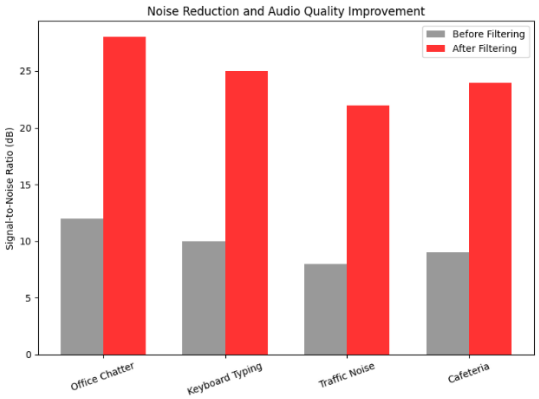

Figure.9. Noise reduction and audio quality improvement

Figure 9 shows the Signal-to-Noise Ratio (SNR) before and after the RNN-based noise-filtering. In all the environments tested, the SNR doubled in almost all of them or tripled, indicating a significant reduction in background noise. This has a direct benefit of enhancing the clarity and intelligibility of speech in meetings. The results demonstrate the model's effectiveness in improving audio quality under various real-life conditions.

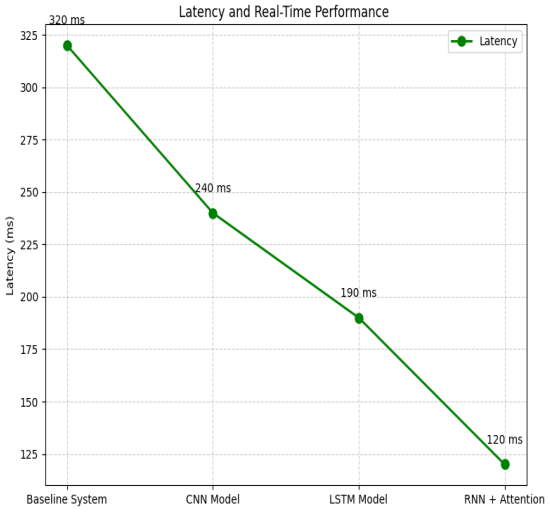

Figure. 10. Latency and real-time performance

Figure 10 illustrates the reduction in latencies that vary with baseline to advanced architecture. The RNN (Attention) model is the best performer, with a delay of only 120 ms, whereas the baseline (Angio) model has a delay of 320 ms. This gradual decline is noted to be indicative of more responsive real-time using more advanced models. Therefore, the suggested scheme has a low-latency capability appropriate for business-tier video communications.

Table.2. Comparison Table For Proposed System

Ref.no	**Technique**	**Accuracy**	**Precision**	**Recall**	**F1-score**
[17]	Deep learning with CNN	79%	68%	75%	73%
Proposed system	RNN	96.8%	95.5%	96.2%	95.8%

The table 2 indicates 79% accuracy, 68% precision, 75% recall, 73% F1-score [17] by the CNN-based system as compared to the proposed RNN system that obtained 96.8% accuracy, 95.5 recalls, 96.2 recalls, and 95.8 F1-score, showing an undeniable augmentation in the performance of the RNN.

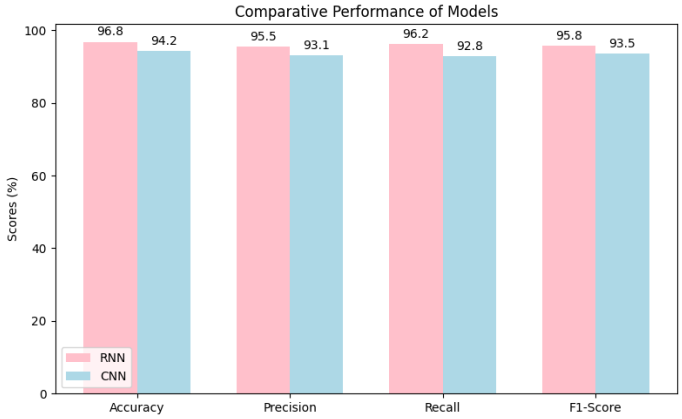

Figure.11. Comparative performance of model Metrics for RNN & CNN

The results of the RNN and CNN models, in terms of Accuracy, Precision, Recall, and F1-Score, are compared in Figure 6. The RNN model (red) outperformed the CNN in all criteria, achieving better values. Although the CNN model, depicted in green, performs well, it shows lower precision and recall compared to the RNN. All in all, RNN is more reliable for use in enterprise-level video chat applications.

5. Conclusions

This research presents an AI-based system to support enterprise-level video communication meetings, addressing challenges associated with remote and hybrid work. The prevalence of spectral gating noise rejection works well to lower the background sound, providing better sound input for speech recognition. The Mel-Frequency Cepstral Coefficients (MFCC) can extract relevant features of speech, and the Principal Component Analysis (PCA) shall reduce dimensionality, which would improve model performance. RNN with an Attention mechanism has demonstrated outstanding ability in utilizing sequential dependencies, resulting in high accuracy and robustness in speaker identification and speech recognition. Experimental figures reveal promising performance measures, confirming that the model can be effectively employed in practice. The model manages the divergent speech through spectral gating, MFCC based speaker features and attention controlled weighting of a frame, which points to the dominant speaker. The change in temporal state of the RNN and change of attention distributions are the speaker transitions. Almost real-time diarization is facilitated by sliding-window processing. Although it can be useful with moderate overlaps, heavy overlapping is an aspect of the problem that should be improved in the future. Future research should also consider incorporating multimodal information, such as video appearances and facial expressions, to further enhance the context tagging of meetings and improve automatic summarization.

Acknowledgement
The authors have no acknowledgements to declare.

Funding
This study has not received any funding from any institution/agency.

Conflict of Interest/Competing Interests
No conflict of interest.

Data Availability
The raw data supporting the findings of this research paper will be made available by the authors upon a reasonable request.

REFERENCES

[1]. Darwich, Mahmoud, and Magdy Bayoumi. "Video quality adaptation using CNN and RNN models for cost-effective and scalable video streaming Services." Cluster Computing 27, no. 5 (2024)
[2]. Farahani, Reza, Zoha Azimi, Christian Timmerer, and Radu Prodan. "Towards ai-assisted sustainable adaptive video streaming systems: Tutorial and survey." arXiv preprint arXiv:2406.02302 (2024).
[3]. Aliyev, Murad. "From Cloud to Camera: Transitioning AI Video Analytics to the Edge in Next-Gen Surveillance Networks." Proceedings of the 10th International Scientific Conference «Modern scientific technology» (2025).
[4]. Pawar, Diksha, Prashant Borde, and Pravin Yannawar. "Generating dynamic lip-syncing using target audio in a multimedia environment." Natural Language Processing Journal 8 (2024): 100084.
[5]. Jain, Varun, Zongwei Wu, Quan Zou, Louis Florentin, Henrik Turbell, Sandeep Siddhartha, Radu Timofte et al. "NTIRE 2025 challenge on video quality enhancement for video conferencing: Datasets, methods and results." In Proceedings of the Computer Vision and Pattern Recognition Conference, pp.
[6]. Vishruth, Raj VV, and S. G. Mohan. "Online video conference analytics: A systematic review." Applied Data Science and Smart Systems (2024): 435-447.
[7]. Lotliker, Shubham, Gouri Bhatikar, Avina Almeida, Ugam Gaude, Siya Naik, and Vivek Jog. "Podcast hosting using spectral gating and speech recognition methodology." In 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), pp. 579-583. IEEE, 2021.
[8]. Avcı, İsa, Elif Yıldırım, and Elif Sare Akdağ. "Ensuring Video Conferencing Applications Security: Security, Privacy, Challenges and Risks." Current Trends in Computing 3, no. 1: 28-43.
[9]. Owobu, Wilfred Oseremen, Olumese Anthony Abieba, Peter Gbenle, James Paul Onoja, A. I. Daraojimba, A. H. Adepoju, and B. C. Ubamadu. "Review of enterprise communication security architectures for improving confidentiality, integrity, and availability in digital workflows." IRE Journals 5, no. 5 (2021): 370-372.
[10]. Stoykova, Stela, and Nikola Shakev. "Artificial intelligence for management information systems: Opportunities, challenges, and future directions." Algorithms 16, no. 8 (2023): 357.
[11]. Sumathy, K., and R. Surendran. "AI-Driven Mentor Service Advising Application for Timely Course Completion using RNN/LSTM." In 2024, the 4th International Conference on Sustainable Expert Systems (ICSES), pp. 694-701. IEEE, 2024.
[12]. Fan, Hu. "Research on innovation and application of 5G using artificial intelligence-based image and speech recognition technologies." Journal of King Saud University-Science 35, no. 4 (2023): 102626.
[13]. Dwivedi, Atul Kumar, Deepali Virmani, Anusuya Ramasamy, Purnendu Bikash Acharjee, and Mohit Tiwari. "Modelling and analysis of artificial intelligence approaches in enhancing the speech recognition for effective multi-functional machine learning platform–A multi regression modelling approach." Journal of Engineering Research-ICMET Special Issue (2022): 04-06.

[14]. Patil, Samarth K., Vinayak Nigam, ShriyashOlambe, Anita Shinde, and Shriyash A. Olambe. "Integrating Artificial Intelligence and Encryption in Web Real-Time Communication: A Smart Video Conferencing Platform with Real-Time Transcription and Translation." Cureus Journals 2, no. 1 (2025).

[15]. Xiao, Honglan. "Training room management based on speech recognition and artificial intelligence." International Journal of Modeling, Simulation, and Scientific Computing 14, no. 03 (2023): 2350004.

[16]. Huang, YuFeng, ShiJuan Lv, Kuo-Kun Tseng, Pin-Jen Tseng, Xin Xie, and Regina Fang-Ying Lin. "Recent advances in artificial intelligence for video production system." Enterprise Information Systems 17, no. 11 (2023): 2246188.

[17]. Verma, Navdeep, SeyumGetenet, Christopher Dann, and Thanveer Shaik. "Designing an artificial intelligence tool to understand student engagement based on teacher's behaviours and movements in video conferencing." Computers and Education: Artificial Intelligence 5 (2023): 100187.

[18]. Jiang, Peiwen, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. "Wireless semantic communications for video conferencing." IEEE Journal on Selected Areas in Communications 41, no. 1 (2022): 230-244.

[19]. Hills, Paul D., Mackenzie VQ Clavin, Miles RA Tufft, Matthias S. Gobel, and Daniel C. Richardson. "Video meeting signals: Experimental evidence for a technique to improve the experience of video conferencing." PloS one 17, no. 8 (2022): e0270399.

[20]. Mah, Pascal Muam, Iwona Skalna, and John Muzam. "Natural language processing and artificial intelligence for enterprise management in the era of industry 4.0." Applied Sciences 12, no. 18 (2022): 9207.