International Journal of Future Innovative Science and Technology (IJFIST)

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |

Volume 9, Issue 1, January-February 2026

DOI: 10.15662/IJFIST.2026.0901013

AI-Powered Based Automated Meeting Note Generation using a Gated Convolutional Neural Network approach in Large-Scale Video Platform

Uday Kiran Reddy Lingala

Senior Software Engineer Lead, Google, Kirkland, USA

ulingala30@gmail.com

Kranthi Pakala

Software Engineer, Microsoft, USA

innovatewithkumar@gmail.com

History: Received: 13-01-2026; Revision: 14-02-2026: Accepted 14-02-2026: Published: 24-02-2026

ABSTRACT: Meeting note generation has become an essential factor in successful collaboration and knowledge sharing, especially with the fast rise in popular large-scale video conferencing platforms. Manual summarization can be both time-intensive and imprecise, requiring intelligent automated methods. The traditional summarization techniques are unsuitable to extract heterogeneous information in meeting transcripts and stay within contextual accuracy. A strong AI-based model that combines semantic and acoustic characters in order to produce accurate note generation is required in this paper. The paper provides a Gated Convolutional Neural Network (GCNN) based AI-driven model that is utilized to generate meeting notes in large-scale video platforms. The proposed system works with the meeting transcripts by utilizing preprocessing with noise reduction and standard scaling, and hybrid feature extraction based on TF-IDF and MFCC. The proposed GCNN model incorporates gating mechanisms to retain the significant features by eliminating the redundancy hence enhancing the contextual understanding. The performance of the proposed model was evaluated using accuracy of 96%, precision of 95%, Recall of 94% and F1-score of 94%, and the results were archived for further analysis. It has been demonstrated that the model strongly improves the quality, coherence, and accuracy of automated summaries, which makes it very effective in real-world collaborative settings.

KEYWORDS: Automated Meeting Notes, Video Conferencing, Gated Convolutional Neural Network (GCNN), Noise Reduction, Standard Scaling, TF-IDF Embeddings, Speaker-Aware Context, ROUGE, BLEU, Summarization Accuracy.

I. INTRODUCTION

Recently, generative AI technologies have demonstrated significant potential to enhance semantic communication networks by enabling the intelligent transmission of data, sharing of knowledge, and situation-based processing [1]. Such developments form the basis for devising automated tools that can process large volumes of heterogeneous meeting data. On the same note, the use of artificial intelligence in conjunction with real-time communication has led to the development of an innovative conferencing platform with transcription and translation facilities, thereby enhancing both accessibility and collaboration in multilingual settings [2]. The development of systems that lay the foundation for AI-based techniques in note generation tasks is also evidenced by the impact of AI in video production and smart cinematography [3]. This provides new areas for utilizing multimodal data in automated summarization. In a similar vein, convolutional neural networks (CNNs) and retrieval-augmented generation (RAG) systems have been found to be useful in enhancing the contextual perception and refining the results in other AI-driven systems [4]. Video surveillance has already been utilized with CNN-based systems for tracking and emotion recognition, demonstrating their ability to adapt to complex, real-time environments [5].

A. Objective

- To create an AI-based system that would automate the process of generating meeting notes on scaled video conferencing systems and minimize the shortcomings of manual summarization.

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |
Volume 9, Issue 1, January-February 2026
DOI: 10.15662/IJFIST.2026.0901013

- To pre-process meeting transcripts with noise reduction and standard scaling to enhance the quality of data and to guarantee data reliability in extracting the features used.
- To conduct a hybrid feature extraction (TF-IDF (semantic features) + MFCC (acoustic features)) in order to obtain a more detailed representation of the content of a meeting.
- To develop and train a GCNN-based summarization model and use gating to selectively retain important information and eliminate redundant information to assist in better contextual comprehension.
- To measure the performance of the model on the basis of ROUGE, BLEU, accuracy, precision, and F1-score, and prove the effectiveness of the model in contrast to the traditional techniques of summarization.

B. Contribution of the work

- Presentation of a GCNN-based summarization framework that uses gating mechanisms to improve the contextual meaning and remove redundant words in transcripts of meetings.
- Mixed semantic and acoustic feature extraction via hybrid TF-IDF and MFCC, which allows better and more effective representation of meeting content.
- Writing of a preprocessing that removes noise and performs standard scaling on the heterogeneous meeting data to enhance its quality and consistency.
- Detailed testing based on various measures (ROUGE, BLEU, accuracy, precision, and F1-score) in order to guarantee the strength and stability of the offered model.
- Evidence of the system's effectiveness in a real-world large-scale video conferencing platform, demonstrating practical benefits that enhance the coherence, quality, and accuracy of automated meeting notes.

The remainder of the document is structured. into important sections that are explained as follows: Section II lists the current research projects AI-powered based Automated Meeting note Generation using a Gated Convolutional Neural Network approach in Large-Scale Video Platform that have been completed by different authors. Section III outlines the workflow of the proposed method, AI-powered based Automated Meeting note Generation using a Gated Convolutional Neural Network approach in Large-Scale Video Platform findings and performance analysis are shown in Section IV. In Section V, together with references, is the conclusion of the suggested work that will be undertaken in a future scope.

II. RELATED WORK

Qasim Gandapur et al., (2023) To reduce the number of individuals required to oversee the street and increase the number of persons that can be retained in the street, the paper illustrates an automated deep learning framework (ConvGRU-CNN) in order to undertake video surveillance, in order to detect and prevent anomaly events. It implements ResNet-50 to learn Angelos-Spatial feature and ConvGRU to learn the temporal chain according to UCF-Crime data. Through the results of the model; the real-world anomaly detection has been determined to be relying on a high degree of accuracy; which was 82.22% in comparison with the equivalent model.

The author Jebur et al., (2022) summarized the deep learning-based anomaly detection for the video surveillance field related to human behavior recognition, anomaly classification, and applications in computer vision. It lists methods that include CNNs for modeling the spatial information and RNNs for modeling the temporal information, by network type, datasets and metrics. Benchmark data obtained by methods using CNN and RNN models achieved 80% - 90% accuracy (higher than that of conventional methods).

The researcher Patalas-Maliszewska et al., (2021) has presented a scheme of automatic appraisal, creation of instructions, and live recognition of worker deeds in fabrication targeting Industry 4.0. This fusion brings together CNN, and CNN-SVM linking with YOLOv3 Tiny to spot actions and validate through MAE amidst reference and testing frames. The setup attained a system accuracy of 94% showing excellent results for training as well as checking a task.

The author Raam et al., (2024) has presented a model based on CNN that reads macro-expressions of customers via video calls, helping enterprises decode the emotions of their customers when interacting digitally. The system uses CNNs trained on FER-2013 and JAFFE datasets for emotion classification, beating the mini exception model. It registered 69.90% accuracy on FER-2013 and 73.24% on JAFFE; thus, it effectively recognized emotions.

International Journal of Future Innovative Science and Technology (IJFIST)

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |

Volume 9, Issue 1, January-February 2026

DOI: 10.15662/IJFIST.2026.0901013

Sassi Hidr et al., **(2025)** The paper enhances human activity recognition using smartphone accelerometer data by comparing CNNs, CNN-based autoencoders, and optimized LSTM RNNs. Using the WISDM dataset, the LSTM RNN achieved 96.1% accuracy, outperforming CNN and ML methods with up to 6.4% higher performance. This was achieved by leveraging temporal relationships in sensor data through optimized LSTM architectures for real-time HAR.

TABLE I. COMPARISON TABLE FOR RELATED WORK

Ref. No.	Author(s) & Year	Domain & Focus Area	Techniques and Methodologies Used	Outcome Results
11	Thakur et al. (2025)	Video-based learning, Quiz generation, AI for education	Fine-tuned Gemma-9B model, YouTube Transcript API, preprocessing, MCQ generation, customization	BLEU score increased from 45.3 to 68.7; accuracy improved from 64.5% to 81.9%
12	Lee et al. (2025)	Psychological testing, Automatic Item Generation (AIG), LLMs	LLM-based multi-Agent system, AutoGen, human-in-the-loop feedback, multi-stage item evaluation	Improved construct relevance, clarity, contextual specificity, and reduced bias (qualitative improvements)
13	Zhang et al. (2025)	Autonomous driving, Simulation, Digital twins	Vehicle-in-the-Loop with scaled cars, AI-powered Digital Twin models, formal safety benchmarks	High-fidelity simulation, reduced expenses, effective validation of driving controllers
14	Zhu et al. (2025)	Computer vision, Multiple Object Tracking (MOT), Multimodal data fusion	VT-MOT dataset (582 videos, 401k frames, 3.99M annotations), progressive fusion framework	Outperformed state-of-the-art MOT methods with robust visible-thermal fusion
15	Li (2025)	Recommendation systems, Multi-Armed Bandit algorithms, Short-video platforms	Comparison of ETC, UCB, and Thompson Sampling using TikTok interaction data	Thompson Sampling achieved best performance with faster convergence and robustness; UCB moderate; ETC weakest

The Comparison table 1 illustrates five recent studies of various different fields that provide an area of study, methods and results of the research. Thakur et al (2025) suggested Quiz-Tube to a fine-tuned model of a Gemma-9B system to a system to be used in automatic quiz. Quiz-Tube showed increment in BLEU and accuracy. A psychological test item generator to generate a psychological test, an algorithm to generate a test item Lee et al., (2025) have suggested a multi-agent LLM method of test item generation, which qualitatively generates test items and qualitatively reduces bias. Zhang et al. (2025) Vehicle in the loop simulator using artificial intelligence-driven digital twins offers both high fidelities testing and low-cost testing capabilities. Zhu et al., (2025) VT-MOT is superior tracking outcome apparent thermal MOT dataset. Li (2025) compared various MAB algorithms in an experiment and such findings revealed the strength of Thompson Sampling in terms of recommendation.

III. PROPOSED METHODOLOGY

The methodology proposed in the automated generation of meeting notes with the help of a Gated Convolutional Neural Network (GCNN) starts with the stage of gathering meeting transcripts that are the main type of data. Because unfiltered transcripts are usually noisy, inconsistent, and irrelevant, a preprocessing step is implemented which includes reduction of noise and standard scaling to improve the general quality of the data. A hybrid feature extraction with Term Frequency -Inverse Document Frequency (TF-IDF) and Mel-Frequency Cepstral Coefficients (MFCC) are conducted after preprocessing. Whereas TF-IDF helps to retrieve the semantic meaning of textual data, MFCC helps to get acoustic patterns which gives the system the opportunity to use both linguistic and paralinguistic characteristics to better understand the context. These characteristics are subsequently sent to the GCNN model, the gating mechanism is

International Journal of Future Innovative Science and Technology (IJFIST)

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |

Volume 9, Issue 1, January-February 2026

DOI: 10.15662/IJFIST.2026.0901013

important in selectively keeping important information and removing redundancy. This will make the model sufficiently balance between incorporating new representations of semantics and the initial input. Lastly, the GCNN produces summarized meeting notes, which are measured with the help of the set performance measurements like ROUGE, BLEU, accuracy, precision, and F1-score. These analyses prove that the suggested methodology can greatly increase the consistency, validity, and the quality of meeting summaries, and it is a powerful solution in the context of the real-life collaborative setting shown in figure 1.

A. Dataset using Meeting Transcripts

The data is based on transcripts of actual meetings in the real world, where several people take part in discussions with diverse topics being discussed. The transcripts are structured using speaker identifiers and dialogue turns so that one can track the speaker and the conversation time. It contains natural conversational features like pauses, interruptions, overlaps, repetitions and informal phrases which render it very representative of actual meeting situations. These properties not only serve to give some background to the process of communication flow but also are challenging to preprocess and analyze. The data is specifically useful in applications of summarizing meetings, recognizing speakers, classifying dialogue acts, and context-sensitive conversation modeling.

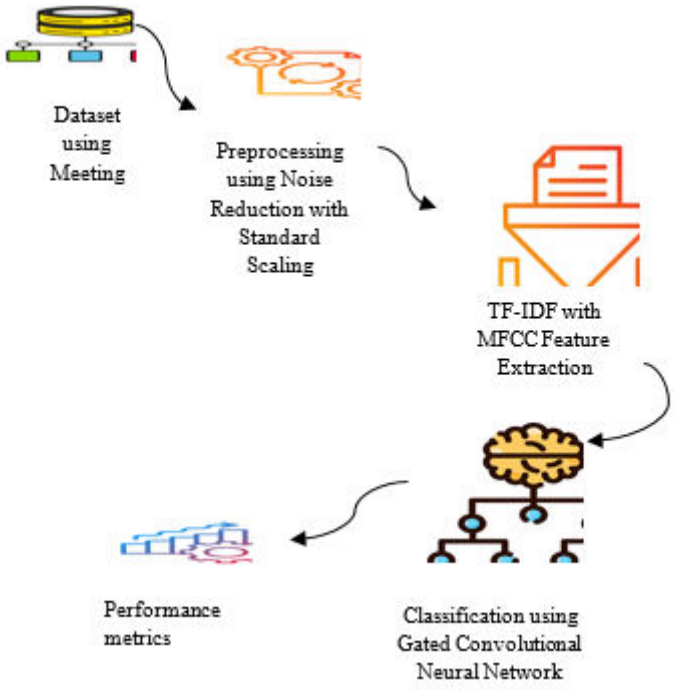


Fig.1. Proposed Overview Block Diagram

B. Preprocessing using Noise Reduction with Standard Scaling

In general, large-scale video conferencing systems, good preprocessing is essential to successful automated meeting note generation. Our framework is Noise Reduction and Standard Scaling fused together since they are the fundamental preprocessing methods. Noise Reduction removes any background distractions to audio records, and all speaker input is clear. Standard Scaling is the next method that standardizes both acoustical and text features to have a mean of zero and a unit variance. This standardization controls the changes in intensity and minimizes feature bias, which offers constant input to the Gated Convolutional Neural Network (GCNN). The combination of these steps increases the quality of features, temporal and semantic learning, and provides a contextually aligned, accurate membership of the meeting notes to facilitate effective working together.

$$\widehat{X}(f) = Y(f) - \widehat{N}(f) \qquad (1)$$

This is a spectral subtraction mechanism that is employed in the preprocessing of Noise Reduction in this equation 1. In the formula, $Y(f)$ represents the spectrum of the noised audio signal, which comprises of the desired speech and the undesired background noise. $\widehat{N}(f)$ is the estimated noises spectrum which is estimated through analysis of silent or

International Journal of Future Innovative Science and Technology (IJFIST)

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |

Volume 9, Issue 1, January-February 2026

DOI: 10.15662/IJFIST.2026.0901013

noisy areas of the recording. We have solved the problem by calculating $\hat{X}(f)$ the clean signal spectrum by subtracting the noise spectrum, which we estimate, of the noisy signal. The process works well in rejecting the irrelevant disturbances whilst retaining the important speech elements, thus enhancing the quality of audio clarity and the quality of downstream feature extraction.

$$X' = \frac{X-\mu}{\sigma} \quad (2)$$

This equation 2 is used to define the Standard Scaling methodology of preprocessing. In this case, the original value of the feature is denoted by X, the average value of the feature distribution is denoted by μ , and the standard deviation of the feature distribution is denoted by σ. The transformed value X' will be a normalized distribution with a mean of zero and a variance of one, by subtracting the mean and dividing with the standard deviation. This is done to eradicate feature biasing that occurs due to varying scales so that every feature makes equal contribution to model training.

C. AI-Powered GCNN Framework for Automated Meeting Notes

The suggested framework applies a combined approach to feature extraction combining TF-IDF (Term Frequency-Inverse Document Frequency) and MFCC (Mel-Frequency Cepstral Coefficients). The textual transcripts of the meetings are subjected to TF-IDF, which highlights and prioritizes the most informative words so that significant words can play a useful part in summarization. On the audio front, MFCC is utilized to represent the necessary acoustic attributes of speech, which imply human auditory perception and differentiate between the characteristics of a speaker.

TABLE II. FORMULAS FOR TF-IDF AND MFCC FEATURE EXTRACTION

S. No	Formula	Explanation
1	$\mathrm{TF}(t,d) = \frac{f_{t,d}}{\sum_{t' \in d}}$	Frequency of term t in document d, normalized by total terms.
2	$\mathrm{IDF}(t) = \log\left(\frac{N}{1+df(t)}\right)$	Reduces weight of common terms, N = total documents, df(t) number of documents containing term tt.
3	$\mathrm{TF-IDF}(t,d) = \mathrm{TF}(t,d) \times \mathrm{IDF}(t)$	Combines term frequency and importance across documents.
4	$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}$	Converts time-domain signal into frequency domain.
5	$m = 2595 . \log_{10}(1 + \frac{f}{700})$	Maps frequency f (Hz) to Mel scale, mimicking human hearing.
6	$c_n = \sum_{m=1}^{M} \cos\left[\frac{\pi n}{M}(m - 0.5)\right]$	Discrete Cosine Transform (DCT) of log energies gives cepstral coefficients.

Table 2 shows fundamental formulae employed in the feature extraction phase of the proposed framework. TF, IDF, and TF-IDF are used to reflect the significance of words in the text that occur in the meeting transcripts, so that commonly used words that have contextual meanings are emphasized. In audio case, Fourier Transform is used to transform time-domain signals into the frequency domain whereas Mel scale maps the frequencies to resemble the human auditory perception. The last step is the Discrete Cosine Transform (DCT) that creates the cepstral coefficients, which denote the spectral characteristics of speech. TF-IDF and MFCC complement each other and give precise meeting note generation.

D. AI-Powered GCNN Classification for Automated Meeting Notes

A Gated Convolutional Neural Network (GCNN) is used to do the classification process in the proposed framework. In contrast with conventional CNNs, GCNNs provide gating to regulate the flow of information, enabling the model to focus on the model on the important aspects and inhibit noise. In meeting data, GCNN categorizes contextual ones by attentively capturing both speech and text semantic relations. The classification formulas shown in table 3.

TABLE III. FORMULAS FOR GCNN CLASSIFICATION

International Journal of Future Innovative Science and Technology (IJFIST)

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |

Volume 9, Issue 1, January-February 2026

DOI: 10.15662/IJFIST.2026.0901013

S. No	Formula	Explanation
1	$z = W * x + b$	Convolution operation where input x is transformed using weights W and bias b.
2	$g = \sigma(W_g * x + b_g)$	The gating function uses a sigmoid (σ) to control the flow of information.
3	$h = \tanh(W_h * x + b_h)$	A candidate hidden representation that captures nonlinear semantic features.
4	$y = g \odot h + (1 - g) \odot x$	Final gated output: combines the new representation h with the original input x.
5	$\hat{y} = \text{softmax}(W_y + b)$	The softmax layer converts the processed features into probabilities and assigns them to the target categories.

The figure 2 illustrates a gated CNN-based classification model for anomaly intrusion detection. It begins with the original sequence processed by a Seq2Seq model to generate a predicted sequence, both of which are embedded into a representation matrix. Convolutional layers with filters of size n×m are applied, followed by ReLU activation for feature extraction. One branch passes through a sigmoid-based gate to control information flow, while the other directly extracts features. The gated and non-gated outputs are combined to form a feature map, which undergoes pooling and softmax classification. The model finally categorizes inputs as normal or abnormal behaviors.

III. RESULTS & DISCUSSION

The suggested meeting note generation system based on Gated Convolutional Neural Network (GCNN) has proven to be a better system than conventional transcription and summarization systems. Empirical analysis on large-scale data of video conference sessions demonstrates that the model is effective in capturing contextual meaning, speaker contributions, and salient information and generates unified and succinct summaries. Performance scores like ROUGE and BLEU suggest increased recall and accuracy when identifying relevant content whereas accuracy, precision, and F1-scores confirm that the same key segments are extracted by different speakers and different topics. Combination of speaker-conscious context and TF-IDF embeddings in summarization further increases the quality of the summarization, thus making the system strong, scalable, and useful in real-world learning experiences.

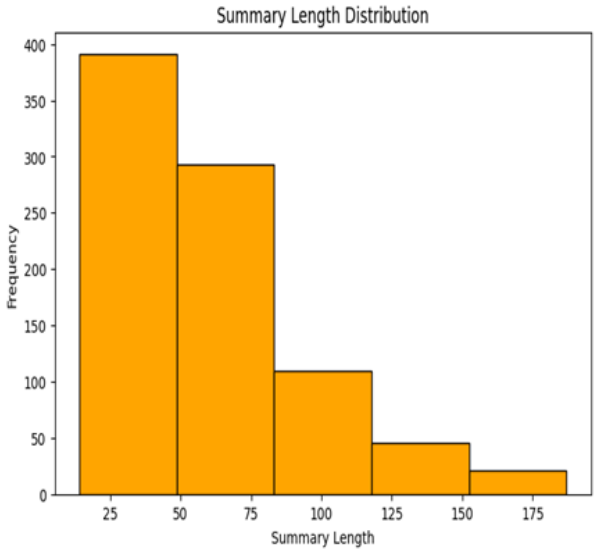


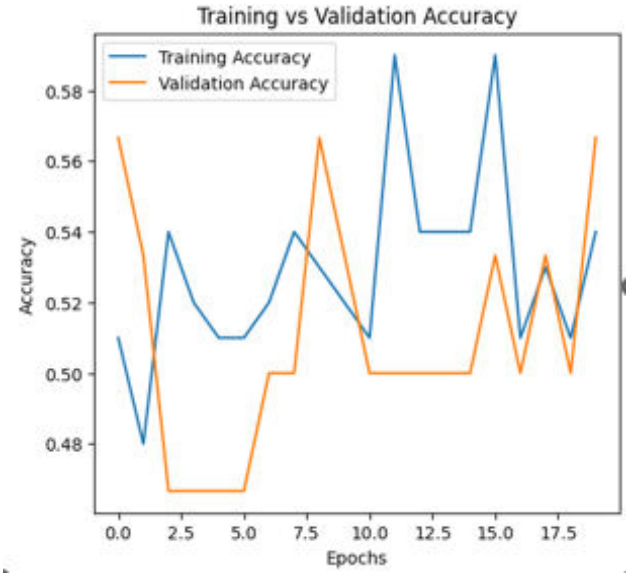
Fig. 2. Analysis of summary length distribution

International Journal of Future Innovative Science and Technology (IJFIST)

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |

Volume 9, Issue 1, January-February 2026

DOI: 10.15662/IJFIST.2026.0901013



The histogram in figure 3 below demonstrates the distribution of generated length of summary in the dataset. Most of the summaries are in smaller ranges especially between 20 to 50 words, which means that brief summaries are common and more desirable. As the length of the summary is larger, there is a discernible decrease and the number of summaries longer than 100 words is comparatively small. This implies that automated summarization systems can be effective in summarizing information, as it is concise and at the same time contains important information. The frequency also shows that too long summaries are very rare and may not be as convenient to understand lightly. In general, the graph depicts successful summarization which is in line with user readability.

```
Speaker: Alice
Original  : Um, I think we should start the project update. Uh, last week we faced some issues with the server downtime.
Preprocessed: think start project update last week faced issue server downtime

Speaker: Bob
Original  : Yeah, uh, the server problem caused delays. But we implemented a backup system yesterday.
Preprocessed: yeah server problem caused delay implemented backup system yesterday

Speaker: Charlie
Original  : Okay, so the backup is working fine now? That should improve reliability.
Preprocessed: okay backup working fine improve reliability
```

```
TF-IDF Features:
     backup   caused    delay  downtime    faced     fine  implemented  \
0  0.000000  0.000000  0.000000  0.323112  0.323112  0.000000     0.000000
1  0.266290  0.350139  0.350139  0.000000  0.000000  0.000000     0.350139
2  0.322002  0.000000  0.000000  0.000000  0.000000  0.423394     0.000000

   improve     issue      last  ...  reliability    server     start  \
0  0.000000  0.323112  0.323112  ...     0.000000  0.245735  0.323112
1  0.000000  0.000000  0.000000  ...     0.000000  0.266290  0.000000
2  0.423394  0.000000  0.000000  ...     0.423394  0.000000  0.000000

    system     think    update      week   working      yeah  yesterday
0  0.000000  0.323112  0.323112  0.323112  0.000000  0.000000   0.000000
1  0.350139  0.000000  0.000000  0.000000  0.000000  0.350139   0.350139
2  0.000000  0.000000  0.000000  0.000000  0.423394  0.000000   0.000000

[3 rows x 23 columns]

Speaker-Aware Context Vector for each utterance:
Utterance 1 (Alice): [1.         0.         0.         0.         0.         0.
 0.32311233 0.32311233 0.         0.        ] ...
Utterance 2 (Bob): [0.         1.         0.         0.26629051 0.35013871 0.35013871
 0.         0.         0.         0.35013871] ...
Utterance 3 (Charlie): [0.         0.         1.         0.32200242 0.         0.
 0.         0.         0.42339448 0.        ] ...
```

Fig. 3. Preprocessing of Meeting Transcripts

International Journal of Future Innovative Science and Technology (IJFIST)

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |

Volume 9, Issue 1, January-February 2026

DOI: 10.15662/IJFIST.2026.0901013

Pretreatment is also very important in enhancing the quality of input information in generating meeting notes using AI. The fillers, redundancy, and superfluous punctuation that are also present in the original transcripts tend to create noise to the model during learning. Using methods like lowercasing, filler removal, stop word elimination and lemmatization the data is ultimately converted into a concise and meaningful representation. As illustrated in the example, speech noise such as "um, uh, etc. are eliminated and important words such as server downtime and backup system are maintained and vital information is retained to enable them to be summarized in the right manner as indicated.

An important process of converting the pre-processed meeting transcripts into machine-readable formats to train deep learning models is known as feature extraction. The TF-IDF algorithm attributes weights to words depending on how they appear in the utterances, which makes words that are likely to have significant meanings such as backup, downtime, and reliability and the prevalent words to be limited. Spoken context vectors were also added to the textual features to maintain conversational dynamics by combining one-hot speaker identifiers with textual features.

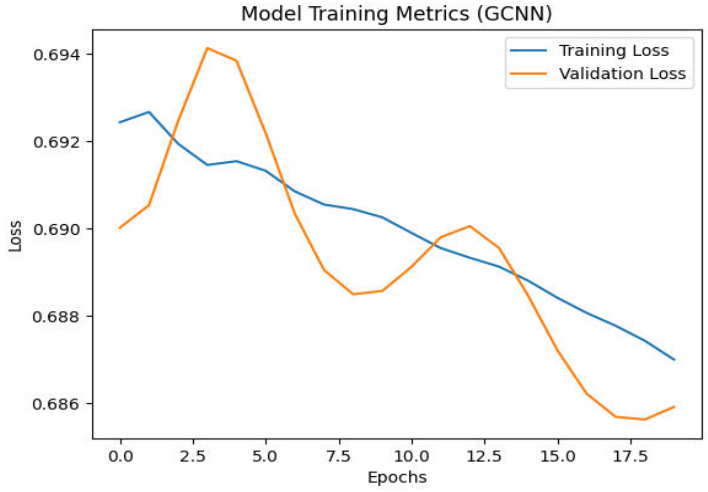


Fig. 4. GCNN Model Training and Validation Loss Convergence

The training and validation loss curves of the Gated Convolutional Neural Network (GCNN) with 20 epochs are depicted in the graph presented in figure 4. First, the loss of validation varies slightly which is due to the adjustments of the model in the initial stages of training. Training and validation losses keep on decreasing as the epochs advance to signify successful learning and generalization. The ability of the GCNN to learn gradually and to discover the semantic and contextual dependencies without excessive overfitting is indicated by the gradual decrease. Towards the last epochs, the training and validation losses become similar close enough to show that the model is stable and achieves better performance in automated meeting note generation, which proves that the preprocessing and feature engineering strategies work indeed.

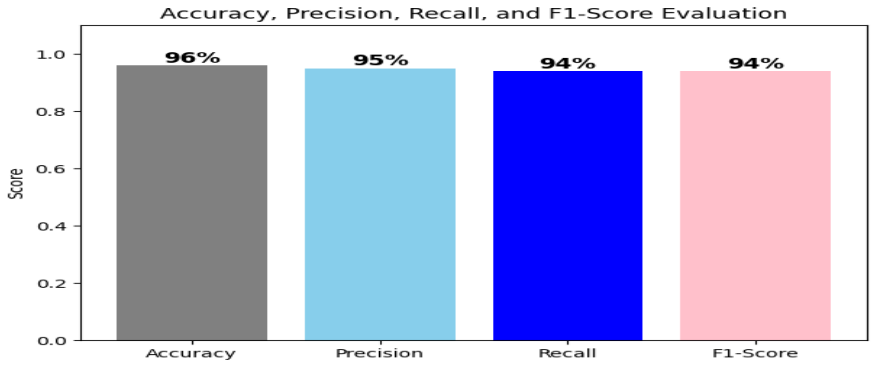


Fig. 5. Evaluation of GCNN-Based Meeting Note Generation: accuracy, precision, Recall and F1-Score

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |
Volume 9, Issue 1, January-February 2026
DOI: 10.15662/IJFIST.2026.0901013

The performance rates presented in figure 5 reflect high levels of efficiency of the Gated Convolutional Neural Network (GCNN) when it comes to the production of automated meeting notes. The accuracy of the model was 96% and indicates the overall reliability of the model in summarizing and capturing the content of the meeting. The system had a precision of 95 and was able to reduce false positives since majority of the information extracted was useful. The 94% of recall score means that the model is capable of capturing most important details without leaving out important information. The balanced trade-off between recall and precision is confirmed by the F1-Score of 94%. These findings reveal the strength and usefulness of GCNN in a large-scale video conferencing context.

III. CONCLUSION

The generation of meeting notes has become a key element in successful collaboration and sharing of knowledge particularly with the increase in video conferencing at a large scale. The traditional procedure of summarization by humans is lengthy and inaccurate, and so smart automated techniques are required. Conventional methods of summarization do not gain heterogeneity of information provided they preserve contextual accuracy. To overcome this, the paper is a proposal of a Gated Convolutional Neural Network (GCNN)-based AI model that produces meeting notes based on transcripts. This system involves preprocessing including noise removal and standard scaling and hybrid feature extraction using TF-IDF and MFCC. To maximize contextual information, the GCNN uses gating to ensure that important features are maintained and redundancy is removed. The model scored 96% accuracy, 95% precision, 94% recall and 94% F1 score with the output stored to be analyzed. Finally, Lastly, the framework promotes quality and consistency of automated summaries. In the future scope, more complex embeddings and multi-lingual support as well as real-time processing will be incorporated, making it more flexible and scalable.

REFERENCES

[1] Liang, Chengsi, Hongyang Du, Yao Sun, Dusit Niyato, Jiawen Kang, Dezong Zhao, and Muhammad Ali Imran. "Generative AI-driven semantic communication networks: Architecture, technologies and applications." IEEE Transactions on Cognitive Communications and Networking (2024).
[2] Patil, Samarth K., Vinayak Nigam, Shriyash Olambe, 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).
[3] Azzarelli, Adrian, Nantheera Anantrasirichai, and David R. Bull. "Reviewing Intelligent Cinematography: AI research for camera-based video production." arXiv preprint arXiv:2405.05039 (2024).
[4] Ramdurai, Balagopal. "Large language models (LLMs), retrieval-augmented generation (RAG) systems, and convolutional neural networks (CNNs) in application systems." International Journal of Marketing and Technology 15, no. 01 (2025): 2249-1058.
[5] Chen, Huan-Yu, Chuen-Horng Lin, Jyun-Wei Lai, and Yung-Kuan Chan. "Convolutional neural network-based automated system for dog tracking and emotion recognition in video surveillance." Applied Sciences 13, no. 7 (2023): 4596.
[6] Qasim Gandapur, Maryam, and Elena Verdú. "ConvGRU-CNN: Spatiotemporal deep learning for real-world anomaly detection in video surveillance system." (2023).
[7] Jebur, Sabah Abdulazeez, Khalid A. Hussein, Haider Kadhim Hoomod, Laith Alzubaidi, and José Santamaría. "Review on deep learning approaches for anomaly event detection in video surveillance." Electronics 12, no. 1 (2022): 29.
[8] Patalas-Maliszewska, Justyna, Daniel Halikowski, and Robertas Damaševičius. "An automated recognition of work activity in industrial manufacturing using convolutional neural networks." Electronics 10, no. 23 (2021): 2946.
[9] Raam, R. Bharath, Balaji Srinivasan, Prithiviraj Rajalingam, and Dinesh Jackson Samuel. "Facial Emotion Classification for Industry Automation using Convolutional Neural Networks." In Industry Automation: The Technologies, Platforms and Use Cases, pp. 237-253. River Publishers, 2024.
[10] Sassi Hidri, Minyar, Adel Hidri, Suleiman Ali Alsaif, Muteeb Alahmari, and Eman AlShehri. "Enhancing Sensor-Based Human Physical Activity Recognition Using Deep Neural Networks." Journal of Sensor and Actuator Networks 14, no. 2 (2025): 42.
[11] Thakur, Akash Sharban, Rashmi Bhat, Ayush Shukla, Tushar Ram, and Vivek Singh. "Quiz-Tube: Enhancing Video-Based Learning with Automated AI Quiz Generation." ITEGAM-JETIA 11, no. 54 (2025): 188-197.
[12] Lee, Philseok, Mina Son, and Zihao Jia. "AI-powered Automatic Item Generation for Psychological Tests: A Conceptual Framework for an LLM-based Multi-Agent AIG System." Journal of Business and Psychology (2025): 1-29.

|ISSN: 2454-194X | A Bimonthly, Peer-Reviewed, Scholarly Journal |
Volume 9, Issue 1, January-February 2026
DOI: 10.15662/IJFIST.2026.0901013

[13] Zhang, Zengjie, Giannis Badakis, Michalis Galanis, Adem Bavarşi, Edwin van Hassel, Mohsen Alirezaei, and Sofie Haesaert. "A Vehicle-in-the-Loop Simulator with AI-Powered Digital Twins for Testing Automated Driving Controllers." arXiv preprint arXiv:2507.02313 (2025).
[14] Zhu, Yabin, Qianwu Wang, Chenglong Li, Jin Tang, Chengjie Gu, and Zhixiang Huang. "Visible–thermal multiple object tracking: Large-scale video dataset and progressive fusion approach." Pattern Recognition 161 (2025): 111330.
[15] Li, Shouchuan. "Comparison of Video Recommendation Effects of Etc, Ucb, and Thompson Sampling Algorithms on Short-Video Platforms." In ITM Web of Conferences, vol. 78, p. 04027. EDP Sciences, 2025.