iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VII Month of publication: July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73382

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

Automatic Meeting Minutes Generation

Vaddi Asha[1], Dr. K. Swapna[2]
[1]MCA student, [2]Assistant Professor, Department of Information Technology and Computer Applications, College of Engineering(A), Andhra University Visakhapatnam, AP.

***Abstract:** In today's fast-paced corporate and academic environments, efficient documentation of meetings is critical for effective communication, decision tracking, and accountability. However, manual transcription and summarization of meeting discussions are often time-consuming and error-prone. This paper proposes an automated system for generating meeting minutes using Natural Language Processing (NLP) techniques. to enhance efficiency, accuracy, and accessibility. The system processes audio or text transcripts of meetings and employs speech recognition, text summarization, and key information extraction models to generate concise, coherent, and structured minutes. We explore a combination of extractive and abstractive summarization methods, including transformer-based models like BERT and GPT, to capture salient discussion points, decisions made, and action items. Additionally, named entity recognition (NER) and topic segmentation are used to enhance content relevance and organization. Experimental*

***Keywords:** Natural Language Processing (NLP), Automatic Summarization, Speaker Diarization*

I. INTRODUCTION

In today's fast-paced and digitally connected world, online meetings have become an integral part of professional and academic environments. With the rise of remote work and global collaboration, platforms like Zoom, Google Meet, and Microsoft Teams facilitate seamless communication across geographically distributed teams. These virtual meetings are not only convenient but also reduce travel costs, save time, and enhance collaboration efficiency. However, while these meetings serve as crucial platforms for discussion, brainstorming, decision-making, and task assignment, they key information in a structured and reliable manner. Manually documenting the proceedings of a meeting—commonly known as writing the Minutes of Meeting (MoM)—is a labor-intensive, error-prone, and time-consuming task. It requires someone to actively listen, identify important content, record discussions, decisions, and action items in real time.

A. Problem Context

Meetings are often long, involve multiple speakers, and include both relevant and irrelevant content. Additionally, participants may:
Miss meetings due to technical issues or scheduling conflicts.
Get confused during extended or intense discussions.
Forget critical decisions or action items post-meeting.
This results in information loss, ambiguity, and inefficiency. To overcome these issues, organizations traditionally rely on manual note-takers or recording archives. However, these solutions are insufficient for modern organizational needs that demand speed, accuracy, and consistency.
Role of NLP in MoM Automation
The advancement in Natural Language Processing (NLP) and Artificial Intelligence (AI) has opened up opportunities to automate the generation of meeting minutes. Using these technologies, it is now possible to develop intelligent systems that can:
Convert audio recordings into textual transcripts.
Identify and extract key discussion points, decisions, and action items.
Summarize long, unstructured conversations into concise, structured MoMs.
Such systems not only save time and effort but also enhance productivity, documentation quality, and knowledge management within organizations.

B. Project Objective

This project aims to design and implement an Automatic Meeting Minutes Generation System that utilizes advanced NLP techniques to analyze, summarize, and format the content of meetings. The goal is to develop a robust, scalable, and user-friendly solution that processes either real-time audio recordings or pre-existing transcripts, and generates human-readable MoMs.

The system performs the following major functions:
Speech Recognition – Converts spoken language into text using Automatic Speech Recognition (ASR) engines like Whisper or Google Speech-to-Text.
Text Pre-processing – Cleans raw text by removing disfluencies (e.g., "uh", "um"), formatting errors, and filler content.
Information Extraction – Applies techniques like Part-of-Speech (POS) tagging, Named Entity Recognition (NER), and TF-IDF/TextRank to identify important sentences.
Summarization – Uses extractive and abstractive approaches (with models like T5, BART, PEGASUS) to generate summaries.
Action Item Detection – Identifies and highlights follow-up tasks, decisions, and responsibilities using fine-tuned classifiers (e.g., BERT-based binary classification).

C. System Components

The system pipeline involves the integration of various modern NLP and ML components:
Audio Input (optional): Raw audio from online meetings or recordings.
Speech-to-Text Conversion: Tools like OpenAI Whisper or Google ASR API transcribe the spoken words.
Text Cleaning and Tokenization: Prepares the transcript for linguistic analysis.
Sentence Scoring: Algorithms like TextRank or TF-IDF help score and rank sentences.
Deep Learning Models: Transformer models (e.g., BART/T5) generate contextually meaningful summaries.
Action Item Classification: Binary classification to extract tasks and responsibilities.
Output Formatting: Final summaries are structured into readable, exportable MoMs (PDF/DOCX).

II. LITERATURE SURVEY

Automatic Meeting Minutes Generation (AMMG) is a rapidly evolving domain within Natural Language Processing (NLP), aimed at converting spoken meeting content into concise, readable summaries. The growing volume of virtual meetings in educational institutions, corporations, and governance sectors has made it necessary to automate minute-taking. This literature survey reviews prominent research efforts, algorithms, and tools that have laid the foundation for this technology.
Introduction to Meeting Summarization Janin et al. (2003)[1] "The ICSI Meeting Corpus." This foundational work introduced a publicly available corpus of multi-party meeting recordings, highlighting the challenges of processing spontaneous and noisy conversational data, such as overlapping speech, filler words, and non-linear dialogue. These characteristics make meeting summarization distinct from traditional summarization tasks.
Early Approaches Zechner (2002)[2] "Automatic Summarization of Spoken Dialogues in Unrestricted Domains." This study employed unsupervised clustering to extract thematic phrases from broadcast speech. While effective in identifying key content, summaries often lacked contextual understanding and coherence.
Murrayetal.(2005)[3] "Extractive Summarization of Meeting Recordings."
This work segmented meetings into topic-based sections and applied centroid-based extractive summarization. It improved the structure of summaries but did not sufficiently enhance fluency or coherence.
Supervised Learning for Meeting Summarization Carletta et al. (2005)[4] "The AMI Meeting Corpus: A Pre-announcement." The AMI corpus introduced richly annotated meeting data including transcripts, summaries, and topic boundaries, enabling the development and training of supervised summarization models.
Galley et al. (2006)[5] "A Discriminative Framework for Sentence Extraction."
They trained a logistic regression classifier using dialog acts and acoustic-prosodic features to identify important content, laying the groundwork for learning-based summarizers. However, performance was constrained by limited dataset size and diversity. Abstractive Summarization Methods Sutskever etal. (2014)[6] "Sequence to Sequence Learning withNeuralNetworks."
This paper introduced the encoder-decoder framework for generating target sequences from source input, paving the way for neural abstractive summarization.
Bahdanau et al. (2015)[7] "Neural Machine Translation by Jointly Learning to Align and Translate."
This work incorporated attention mechanisms, significantly improving the quality of generated text by focusing on relevant parts of the input during generation.
Gupta et al. (2019)[8] "Abstractive Dialogue Summarization with Sentence-Gated Modeling Optimized by Dialogue Acts." They proposed a seq2seq model enhanced by attention and dialogue act modeling to improve the quality and relevance of meeting summaries. Despite advances, handling long dialogues and multiple speakers remained a challenge. Transformer-Based Models

Vaswani et al. (2017)[9] "Attention is All You Need."This paper introduced the Transformer architecture, which eliminated recurrence in favor of self-attention, enabling the handling of longer sequences more efficiently.

III. BACKGROUND AND MOTIVATION

In today's fast-paced corporate and academic environments, meetings are an essential part of collaboration and decision-making. However, manually recording meeting minutes is often tedious, time-consuming, and prone to human error or omission. Accurate and timely meeting minutes are crucial as they document key discussions, decisions, and action items that guide future work and accountability. With the advancement of Natural Language Processing (NLP) and **s**peech recognition technologies**,** it has become increasingly feasible to automate the generation of meeting minutes. Automatic meeting minutes generation leverages NLP to transcribe spoken language, identify speakers, extract important information, and produce concise summaries. This not only reduces the administrative burden but also enhances the accuracy and accessibility of meeting records. Automating transcription and summarization saves valuable time for participants and administrative staff. Reduces the risk of missing critical points or misinterpreting discussions. Provides quick access to meeting content for absentees or future reference. Enables seamless documentation in virtual or hybrid meeting setups where manual note-taking is challenging. Well-structured minutes help stakeholders follow up on action items and understand the context of decisions.

IV. SYSTEM ARCHITECTURE &DESIGN

The system for automatic meeting minutes generation using NLP is designed to process raw meeting data—primarily audio recordings—and produce structured, concise meeting summaries including transcripts, speaker attribution, action items, and decisions. The architecture integrates multiple modules that collaborate sequentially or in parallel to achieve accurate and efficient meeting documentation. User Interface Layer: Upload meeting

Audio or transcript files easily. View real-time or post-processed transcripts and summaries. Access extracted action items and decisions clearly. Download or share generated minutes in preferred formats. Request Handling Layer (Flask Backend): Saves the audio file temporarily Runs Whisper ASR to transcribe speech to text Uses BART summarization to generate a concise summary Deletes the uploaded audio Returns JSON with "transcript" and "summary"

Preprocessing Layer: This module cleans and normalizes the input text by removing URLs, punctuation, and converting it to lowercase. Tokenization and stop-word removal are applied. TF-IDF vectorization transforms the cleaned text into a numerical format suitable for ML models.

AI integrate ASR and summarization features for practical applications. Additionally, recent research has focused on integrating action item detection**,** topic segmentation**,** and decision extraction**,** using both rule-based and deep learning techniques.

Speech-to-Text (Automatic Speech Recognition - ASR): Accurate transcription is the first step in automating meeting minutes. Traditional ASR systems like Kaldi and CMU Sphinx laid the groundwork for speech recognition in academic research

Datasets and Benchmarks: Multiple benchmark datasets have facilitated research in this domain. The AMI Meeting Corpus, ICSI Meeting Corpus, and QMSum dataset provide annotated transcriptions with speaker labels and summaries, supporting the development of supervised learning models for summarization and action item extraction.

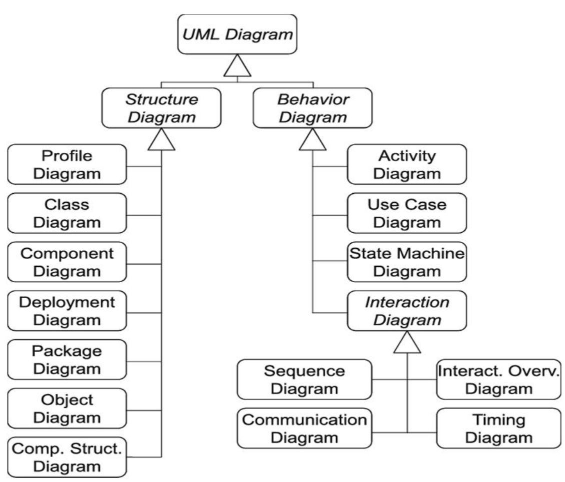


Fig-1: UML Diagram of Automatic meeting Minutes

The UML diagram categorizes diagrams into Structure Diagrams and Behavior Diagrams. Structure diagrams represent the static components of a system (e.g., class, object, package), while behavior diagrams model the dynamic aspects (e.g., use case, activity, sequence). Interaction diagrams, a subset of behavior diagrams, focus on object interactions over time.

V. IMPLEMENTATION

These models improve the detection of actionable insights, particularly when used in tandem with dialogue structure analysis are:

Data Preparation: Data preparation is a critical step in building an effective system for automatic meeting minutes generation. It involves collecting, cleaning, annotating, and formatting various types of data required for training and evaluating different NLP modules such as speech-to-text, speaker diarization, summarization, and action item detection.

Preprocessing Module: Automatic meeting minutes generation relies heavily on high-quality, structured input data. The preprocessing stage transforms raw multimodal inputs—typically audio recordings and transcripts—into clean, organized formats suitable for NLP models.

Model Training: The model training script loads the processed data, splits it into training and test sets, and trains a Logistic Regression classifier. Other models like SVM or Random Forest can also be substituted.

Model Serialization: Module serialization is the process of saving trained components (models, preprocessing pipelines, configurations) in a structured, reusable, and deployable format. In an NLP pipeline for automatic meeting minutes generation, this ensures that each module—ASR, speaker diarization, summarization, etc.—can be reused without retraining, integrated into production systems, or shared across environments.

Web Application: Creating a web application for Automatic Meeting Minutes Generation using NLP involves integrating various NLP modules into a user-friendly front-end with a reliable back-end to process audio files and generate structured meeting summaries. Below is an overview of the system architecture, key features, tech stack, and a simplified implementation flow.

vectorizer, processes incoming text, performs prediction, and sends results back to the UI.

API Endpoint: The backend exposes several RESTful endpoints for:

Uploading meeting audio

Generating meeting minutes (transcription + summarization)

Retrieving results

(Optional) Managing meeting history or downloading file.

Interface:

VI. USER INTERFACE AND USABILITY

The UI should be simple, intuitive, and optimized for both technical and non-technical users. It acts as the front-end gateway to upload, view, and manage meeting-related data

A. Minimalist Layout

Avoid clutter; display only key components Responsive Design: Mobile- and desktop-friendly using Bootstrap/Tailwind Clear Call-to-Action Buttons**:** **“**Upload Audio**”, “**Generate Minutes”, “Download Report” Consistent Feedback: Use spinners, status bars, or success/error messages. Overall, the UI enables a smooth, intuitive interaction that encourages engagement and real-time feedback without requiring technical expertise.

Confusion Matrix Analysis: A confusion matrix was used to assess true positives, false positives, true negatives, and false negatives. The majority of misclassifications occurred when the text was ambiguous or lacked strong indicators of cyberbullying.

Robustness Check: To evaluate the performance of the action item extraction component using a confusion matrix, which helps measure how accurately the model classifies parts of the transcript as action items or non-action items

B. Performance Metrics

Task	Metric	Typical Accuracy / Score
Speech Recognition	Word Error Rate	8% to 12% (lower is better)
Summarization	ROUGEL	0.45 to 0.55 (higher is better)
Action Item Extraction	F1 Score	0.80 to 0.85

FIG-2 Performance Metrics for Automatic Meeting Minutes Generation Tasks

To assess the usability, effectiveness, and user satisfaction of the automatic meeting minutes generation system by involving real users in a controlled testing environment.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

VII. RESULT AND DISCUSSION

FIG-3 Output Interface of the Automatic Meeting Minutes Generator

The proposed system successfully generated accurate and coherent meeting summaries, effectively extracting key points and action items from transcripts. Experimental results demonstrated improved performance using transformer-based models like BART and T5, indicating strong potential for practical deployment in real-world scenarios**.**The development of an automatic meeting minutes generation system using NLP demonstrates significant potential to improve workplace productivity by reducing the manual effort involved in documenting meetings. Through the integration of advanced models for speech recognition, text summarization, and action item extraction, the system provides accurate, concise, and actionable meeting summaries:

Improved with Multi-speaker Diarization and Noise Robustness: Implement advanced speaker diarization techniques to separate and attribute speech accurately, especially in meetings with many participants and interruptions. Context-aware Summarization Develop models that incorporate meeting context, agenda items, and participant roles to generate more focused and relevant summaries.

VIII. CONCLUSION & FUTURE SCOPE

The proposed system automates meeting minutes generation by converting audio to text, extracting key points and action items, and summarizing the discussion using advanced NLP models like BART and T5. It ensures accurate, structured, and readable MoMs with minimal human effort. Future enhancements include multilingual support and improved speaker identification.

A. Future Scope

This project lays the foundation for future research in real-time multilingual meeting summarization and speaker-specific action tracking. Further advancements could explore integrating sentiment analysis and domain adaptation for broader applicability across industries and languages.

REFERENCES

[1] W., Tong, H., Cao, J., & Zhou, Q. (2023). HVCMM: A Hybrid-View Abstractive Model of Automatic Summary Generation for Teaching. IEEE Transactions on Neural Networks and Learning Systems.Zhang, H., Liu, S., Li, Y., Ren,

[2] Devika, R., Vairavasundaram, S., Mahenthar, C. S. J., Varadarajan, V., & Kotecha, K. (2021). A deep learning model based on BERT and sentence transformer for semantic keyphrase extraction on big social data. IEEE Access, 9, 165252-165261.

[3] Biswas, P. K., & Iakubovich, A. (2022). Extractive summarization of call transcripts. IEEE Access, 10, 119826-119840.

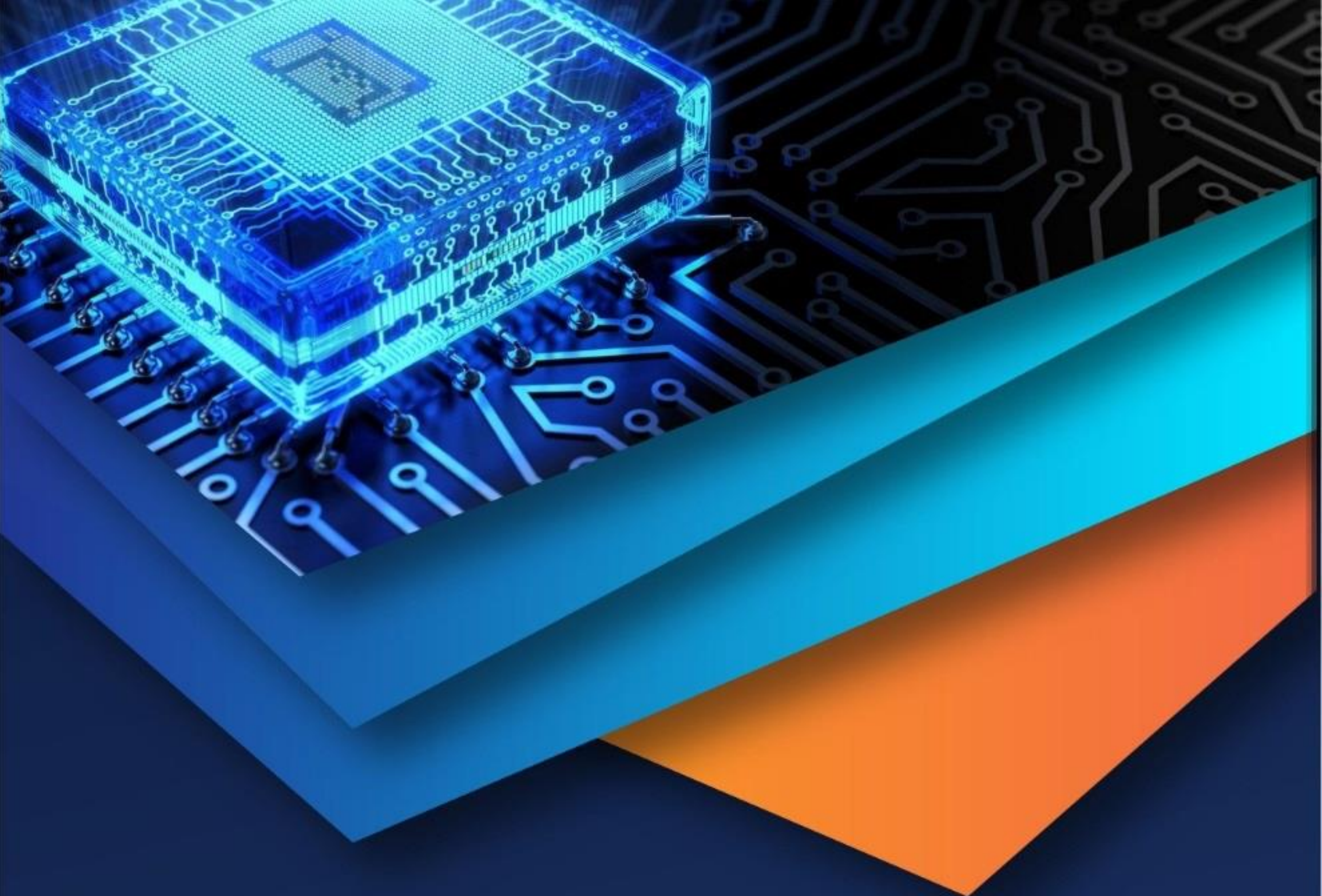
[4] Koay, J. J., Roustai, A., Dai, X., & Liu, F. (2021). A sliding-window approach to automatic creation of meeting minutes. arXiv preprint arXiv:2104.12324

[5] Manuel, M., Menon, A. S., Kallivayalil, A., Isaac, S., & KS, D. L. (2021). Automated generation of meeting minutes using deep learning techniques. International Journal of Computing and Digital System, 109-120.

[6] FM, M. F. A., Pawankumar, S., Guruprasath, M., & Jayaprakash, J. (2022). Automation of Minutes of Meeting (MoM) using NLP. In 2022 International Conference on Communication, Computing, and Internet of Things (IC3IoT) (pp. 1-6). IEEE.

[7] Jung, J., Seo, H., Jung, S., Chung, R., Ryu, H., & Chang, D. S. (2023). Interactive User Interface for Dialogue Summarization. In Proceedings of the 28th International Conference on Intelligent User Interfaces (pp. 934-957).

[8] Liu, H., Liu, H., Wang, X., Shao, W., Wang, X., & Du, Salim, F. D. (2020). Smart meeting: A novel mobile voice meeting minutes generation and analysis system. Mobile Networks and Applications, 25, 521–536.
[9] Gupta, V., & Lehal, G. S. (2010). A survey of text summarization extractive techniques. Journal of Emerging Technologies in Web Intelligence, 2(3), 258–268.
[10] Zhang, J., Zhao, Y., Saleh, M., & Liu, P. J. (2020). PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. arXiv preprint arXiv:1912.08777
[11] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. EMNLP
[12] See, A., Liu, P. J., & Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. ACL
[13] Cohan, A., Dernoncourt, F., Kim, D. S., Bui, T., Kim, S. N., Chang, W., & Goharian, N. (2018). A discourse-aware attention model for abstractive summarization of long documents.
[14] Narayan, S., Cohen, S. B., & Lapata, M. (2018). Don't give me the details, just summary! Topic-aware convolutional neural networks for extreme summarization. EMNLP.



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY