IES International Journal of Multidisciplinary Engineering Research

AI Powered Voicemail Prioritisation System

Aflah Ali[1], Ajul N D[2], Anfil Hassan[3], Sona Santhosh[4], Athira E P[5]

[1,2,3,4] *Student, CSE, IES College of Engineering, Kerala*

[5] *Assistant Professor, CSE, IES College of Engineering, Kerala*

Email_id: aflahali42@gmail.com, anfilanfil94@gmail.com, ajulnj213@gmail.com, sonasanthosh004@gmail.com, athiraep@iesce.info

Abstract

Managing voicemails in critical sectors such as healthcare, finance, and customer service is often hindered by delays, inefficiencies, and missed urgent messages. This project introduces an AI-powered voicemail system designed to intelligently capture, analyze, and prioritize voice messages, ensuring that high-urgency cases are addressed immediately. Using Twilio for voicemail capture and Google Speech Recognition for transcription, the system combines text and audio analysis to assess both semantic content and vocal tone. BERT-based text classification evaluates urgency and routes messages to the appropriate department, while OpenSMILE and Random Forest models analyze acoustic features to detect stress, pitch, and speed, indicating urgency levels. PII is automatically redacted using SpaCy to maintain privacy and data security. By merging speech processing, NLP, and machine learning, the system offers a scalable, reliable, and efficient solution for modern communication management.

Keywords: Artificial Intelligence, Voicemail Prioritization, Speech Recognition, Urgency Detection, Privacy Protection.

DOI: https://doi.org/10.5281/zenodo.18312156

1. Introduction

In today's fast-paced organizational environment, managing communications efficiently is more important than ever. Many organizations rely on voicemail systems to handle important messages, yet traditional systems often fail to prioritize messages, resulting in delayed responses. This limitation arises from outdated mechanisms that treat all voicemails equally, leaving urgent communications overlooked and reducing reliability in sectors such as healthcare, finance, and customer service. The AI-powered Voicemail System proposed in this project addresses these issues by leveraging advanced technologies to create an intelligent, automated, and highly efficient voicemail management framework. By combining speech recognition, natural language processing (NLP), and machine learning, the system transcribes voicemails, detects urgency, and routes messages appropriately. Through this approach,

organizations can ensure timely responses, improve workflow efficiency, and enhance overall communication reliability.

1.1. Speech-to-Text Engine – Converts incoming voice messages into accurate text transcriptions. This enables users to quickly read messages without listening to the entire audio and facilitates further analysis by machine learning models.

1.2. Priority Classification – A machine learning module that automatically categorizes messages as high, medium, or low priority based on content and urgency, ensuring critical communications are highlighted and addressed promptly.

1.3. Voicemail Management Server – Captures voicemails via Twilio Cloud, stores audio recordings and metadata locally, and processes them through the speech-to-text and classification modules. This central server serves as the backbone of the system, maintaining all message-related data in a structured MySQL database.

1.4. User Interface (UI) – A web-based dashboard that allows users and administrators to listen to voicemails, read transcriptions, view urgency levels, and sort or filter messages based on priority. The interface also incorporates privacy features, such as redaction of sensitive information like names and contact details

1.5. Enhanced Efficiency and Reliability – By integrating these components, the AI-powered voicemail system transforms traditional message handling into a streamlined, automated communication assistant, improving efficiency, responsiveness, and trustworthiness across diverse organizational settings.

2. Literature Review

R. Das, S. Mehta, L. Roy, and P. Banerjee (2024) [1] proposed "AI-Powered Voicemail Management System." Their integrated model combined speech recognition, NLP-based transcription, and machine learning-driven classification to automatically prioritize messages. The system featured a dashboard for users to view transcriptions, urgency scores, and message categories in real time. Deployed in enterprise front-desk environments, it demonstrated a 40% reduction in average response times and improved operational efficiency. The framework also incorporated user feedback loops to refine message categorization over time. Furthermore, it supported multilingual transcription, enhancing accessibility for global organizations. Overall, the system illustrated the potential of AI in streamlining enterprise communication workflows.

A. Kaur, R. Nair, S. Iyer, and M. Reddy (2024) [2] developed "CNN and Wave2Vec-Based Voicemail Detection and Filtering." The system used a hybrid deep learning framework combining CNNs and Wave2Vec embeddings to process acoustic and contextual features in voicemail audio. It efficiently filtered and prioritized important messages even in noisy environments, achieving low latency suitable for real-time applications. Additionally, the model dynamically adapted to speaker variations and background noises. The authors demonstrated its scalability across different hardware platforms, including edge devices. This approach highlighted the advantages of combining acoustic and contextual embeddings for accurate voicemail processing.

Paritosh Ranjan, Surajit Majumder, and Prodip Roy (2024) [3] introduced "Generative Voice Bursts (GVB) for Real-Time Context-Aware Audio Alerts." The study employed generative AI to create short, context-aware voice messages derived from data such as location, health, or background activity. GVB technology enabled urgent messages to bypass standard voicemail queues, ensuring immediate delivery during critical situations. Results indicated significant improvement in real-time emergency communication and system responsiveness. The system also reduced user cognitive load by summarizing essential information quickly. Experimental trials confirmed minimal latency even under heavy message traffic. These findings suggest GVB can be a key tool for critical alert management in smart environments.

L. Wang, X. Zhao, Y. Liu, and H. Chen (2023) [4] developed "BERT-Based Department Routing System for Contact Centers." The model analyzed voicemail content using fine-tuned BERT embeddings and classified it to automatically route messages to the most relevant department. The inclusion of domain-specific language understanding reduced misrouting rates and improved message handling speed by 30%. The approach demonstrated how NLP can optimize workflow management and reduce manual intervention in enterprise call centers. Moreover, the system was capable of incremental learning to accommodate evolving business terminology. The dashboard provided managers with analytics on routing efficiency and workload distribution. Overall, it exemplified the use of transformer-based NLP for intelligent enterprise communication management.

P. Rajan, S. Kumar, A. Raghavan, and T. Balaji (2023) [5] presented "Real-Time Voice Message Classification System for Healthcare." Their system integrated keyword spotting, speech-to-text transcription, and urgency detection to automatically identify critical patient messages such as emergency alerts and lab results. The authors also implemented a live dashboard to monitor and notify healthcare professionals instantly**.** Experimental evaluation in simulated hospital settings showed high recall for urgent messages and improved communication reliability. The solution also enabled prioritization of messages based on patient severity levels. Integration with hospital electronic health records further enhanced response accuracy. It underscored the importance of AI-driven automation for improving patient care and clinical workflow efficiency.

K. Singh, R. Gupta, P. Agarwal, and S. Joshi (2022) [6] introduced "Deep Learning-Based Email Classification Using BERT and LSTM." The hybrid model utilized transformer-based BERT embeddings alongside multi-layer LSTM networks for sequential learning. Fine-tuning BERT on domain-specific datasets enhanced contextual understanding and classification precision. While originally applied to emails, this technique can be extended to voicemail transcripts to automatically categorize and prioritize voice messages with minimal human review. The authors highlighted reduced processing time compared to traditional ML pipelines. Experiments showed robustness in handling noisy or ambiguous text inputs. This work provided a foundation for cross-domain adaptation of deep learning models for message prioritization.

M. Sharma, A. Kapoor, V. Reddy, and P. Das (2022) [7] proposed "Multi-Modal Voicemail Prioritization System." Their research combined CNN-based acoustic analysis with NLP-driven text interpretation for voicemail

prioritization. By fusing audio and text modalities, the model achieved superior urgency detection accuracy compared to unimodal systems. Tests indicated a 15% improvement in precision and user satisfaction, highlighting the advantages of multi-modal AI frameworks in message prioritization. The approach also enabled dynamic weighting of audio versus text features based on context. User studies indicated faster comprehension and decision-making when both modalities were presented. This reinforced the value of integrating multiple data modalities in intelligent communication systems.

J. Jaikumar, R. Suresh, P. Natarajan, and V. Kumar (2021) [8] developed "Privacy-Preserving Framework for Detecting and Masking PII in Voicemails." Using machine learning classifiers and linguistic feature extraction, the system automatically detected and masked personal identifiers such as phone numbers, names, and emails. The framework adhered to data privacy regulations including GDPR and HIPAA, demonstrating robust performance with low false-positive rates. It significantly strengthened privacy in automated voicemail transcription environments. The system also supported real-time masking during live voicemail playback. Integration with enterprise compliance dashboards facilitated auditing and reporting. This framework emphasized how privacy-preserving AI can be effectively incorporated in sensitive communication workflows.

H. Kamiyama, T. Nakano, Y. Fujimoto, and S. Takeda (2021) [9] proposed "Urgent Voicemail Detection Using Temporal Speech Analysis." The authors used recurrent neural networks (RNNs) with temporal attention mechanisms to analyze features like pitch, energy, and spectral variation in speech. The system effectively captured urgency and emotional tone through temporal dependencies in voice patterns. Results showed improved detection accuracy compared to traditional audio classifiers, enhancing automated response systems in organizational communication**.** The model also adapted to speaker-specific characteristics to minimize false alarms. Its lightweight architecture allowed deployment on mobile and embedded devices. These advancements highlighted the potential of temporal modeling for urgency and emotion detection in voicemails.

R. Patel, M. Sharma, A. Singh, and D. Verma (2021) [10] designed "Speech Emotion Recognition Framework for Voicemail Prioritization." The system employed OpenSMILE for extracting acoustic features such as MFCCs, pitch, and intensity, combined with Random Forest algorithms for emotion classification. By identifying stress, urgency, and frustration, the model enhanced voicemail prioritization. The study confirmed that emotion-driven analysis significantly correlates with message urgency, improving responsiveness in real-world scenarios. The framework also supported cross-linguistic emotion recognition for global user bases. Real-world deployment trials showed consistent improvements in response prioritization. This work demonstrated how emotion-aware AI can complement conventional content-based voicemail management strategies.

The reviewed studies collectively demonstrate the progressive evolution of AI-driven voicemail management systems from traditional text-based classification toward highly intelligent, multimodal, and context-aware frameworks. Early research primarily focused on foundational approaches using statistical and machine learning methods such as SVM, Naive Bayes, and TF-IDF to classify voicemails or emails based on urgency and intent. These

models, while effective for structured datasets, lacked semantic understanding and contextual adaptability. Subsequent works introduced deep learning architectures—particularly transformer-based models like BERT and hybrid combinations with LSTM—to enhance the semantic and sequential comprehension of voicemail transcripts. These methods achieved superior accuracy in urgency detection and message routing by capturing nuanced linguistic and contextual patterns that traditional algorithms overlooked.

Further advancements incorporated multimodal learning paradigms, integrating both audio and text features for richer emotional and contextual understanding. Studies utilizing CNNs for acoustic feature extraction and transformer-based NLP for textual analysis demonstrated that fusing modalities substantially improved prioritization accuracy and reduced response times. Generative model**s**, such as the proposed Generative Voice Bursts (GVB)**,** and attention-based temporal speech models further advanced the field by enabling real-time, context-aware alerting and emotion recognition. These developments not only enhanced the precision of urgency detection but also allowed systems to adapt dynamically to varying speech tones, accents, and emotional intensities. As a result, modern voicemail analysis frameworks are evolving toward more human-like interpretation capabilities, bridging the gap between artificial intelligence and intuitive communication understanding.

3. Review of Methodology

3.1. System Design:

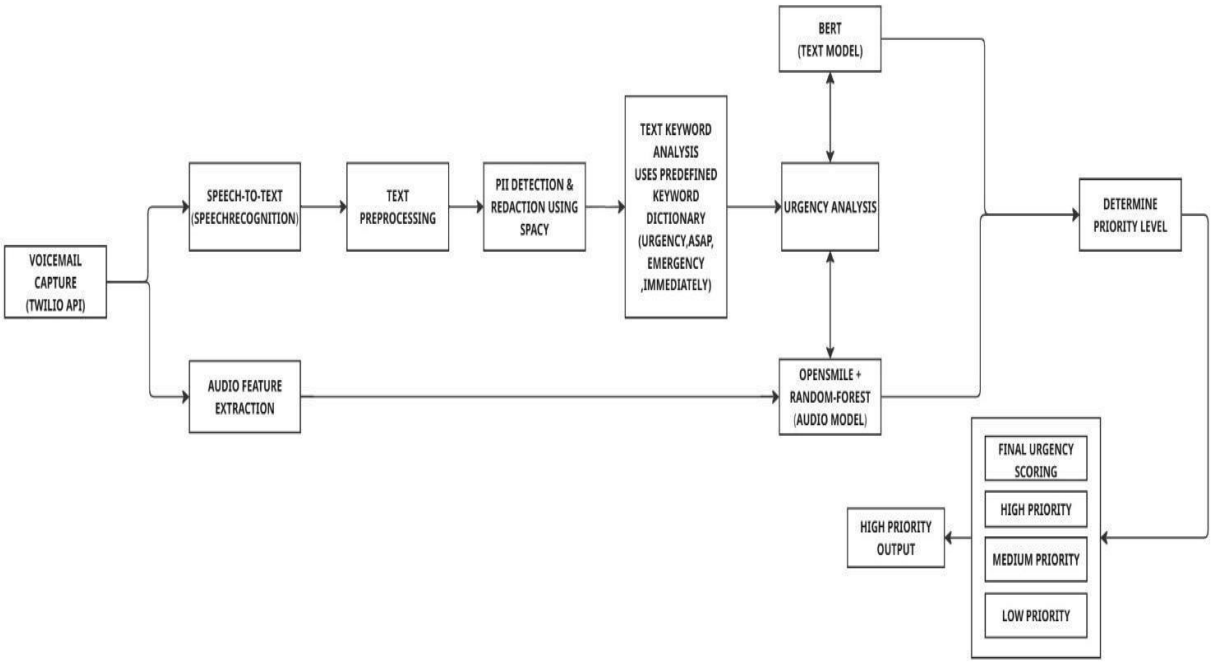


Figure 1: System Design

The Voicemail Urgency and Priority Analysis System processes incoming voicemails to automatically determine their level of urgency, ensuring that high-priority requests are addressed immediately. It integrates Speech Recognition, Natural Language Processing (NLP), and Acoustic Modeling to provide a final priority score. The system transcribes each voicemail into text, analyzes the content for urgency-related keywords, and evaluates vocal characteristics such as pitch, tone, and speech rate to assess emotional intensity. By combining both textual and acoustic insights, it classifies messages into High, Medium, or Low priority categories. This automated approach reduces response time, enhances workflow efficiency, and minimizes the risk of overlooking critical messages.

3.2. Speech-to-Text Module

This module serves as the primary gateway for converting raw audio into processable data, enabling text-based analysis:

- Voicemail Capture (Twilio API): Securely ingests and captures the initial voicemail audio data via an external telephony API.
- Speech-to-Text (Speech Recognition): Transcribes the captured audio into text for subsequent linguistic analysis.
- Text Preprocessing: Cleans and normalizes the transcribed text, preparing it for deeper processing stages.

3.3. Text Analysis Module

This module performs linguistic scrutiny to assess the urgency explicitly expressed in the caller's language:

- PII Detection & Redaction using Spacy: Automatically identifies and removes sensitive Personally Identifiable Information (PII) to maintain privacy and compliance.
- Text Keyword Analysis: Scans the text using a predefined keyword dictionary for explicit urgency terms (e.g., "URGENCY," "ASAP," "EMERGENCY," "IMMEDIATELY").
- BERT (Text Model): A sophisticated Transformer-based NLP model performs contextual analysis to understand the deeper meaning and true urgency of the message.

3.4. Audio Analysis Module

This module runs in parallel with the Text Analysis to assess urgency based on the caller's tone and vocal characteristics:

- Audio Feature Extraction: Isolates and extracts specific acoustic properties from the raw audio signal (e.g., pitch, energy, speaking rate).
- OpenSMILE + Random-Forest (Audio Model): A machine learning model processes the acoustic features to determine the caller's emotional state and urgency based on their voice inflection and prosody.

3.5. Priority Determination Module

This final module integrates the results from both parallel pipelines to assign a definitive action-level priority:

- Urgency Analysis: Combines the linguistic score from the BERT model and the acoustic score from the Random-

Forest model to create a unified urgency assessment.

- Determine Priority Level: Uses the unified assessment to classify the voicemail into one of the system's defined priority categories.
- Final Urgency Scoring: Outputs the final, actionable classification, directing the voicemail to the appropriate queue: High Priority Output, Medium Priority, or Low Priority.

4. Review of Voicemail System Components

A review on data handling in the AI-Powered Voicemail Prioritisation System highlights how the platform efficiently manages, processes, and secures voicemail information to ensure accurate urgency detection and reliable message routing. The system captures live voicemail recordings through the Twilio interface, automatically converting them into text using speech-to-text technology. Both the audio and transcribed text are analyzed for predefined urgency keywords and tone variations such as pitch, speed, and stress, which help determine message priority. All recordings, transcriptions, and classification outputs are stored securely in a structured MySQL database, allowing seamless retrieval and monitoring through the web interface. The system also integrates privacy-preserving mechanisms, including PII redaction using spaCy, ensuring sensitive user data such as names or phone numbers remain protected. Additionally, the integration of encryption protocols ensures that all stored data and transmissions remain confidential during processing and retrieval. The database is regularly synchronized with backup servers to prevent data loss and maintain system reliability. Furthermore, role-based access control is implemented within the dashboard interface, allowing only authorized users to view, manage, or analyze voicemail data, thereby enhancing overall security and operational transparency.

4.1. Input & Feature Extraction:

The core input data is the voicemail, which is processed to extract both textual and acoustic features crucial for analysis. Voicemail Audio the raw audio is captured via the Twilio API. Its fidelity is essential for both Speech-to-Text and Audio Feature Extraction. Maintaining a high-quality, continuous audio feed is critical for the system's operational effectiveness and reliability. Audio Features extracted features (e.g., pitch, energy, spectral data) are the numerical representation of the caller's tone and emotion. These features must be robust and discriminant to ensure the OpenSMILE + Random-Forest (Audio Model) can accurately infer urgency from vocal prosody.

4.2. Text Processing and Compliance:

This stage focuses on converting audio to text, cleaning the data, and ensuring privacy standards are met before linguistic analysis. Speech-to-Text Transcription transcribed text must be highly accurate, as errors in transcription directly lead to failures in keyword detection and contextual understanding by the BERT model. Accuracy is vital for the system's core function. PII Detection and Redaction the system utilizes Spacy models configured with specific rules to detect and mask sensitive Personally Identifiable Information (PII). Completeness and consistency in redaction are crucial to maintain privacy compliance without compromising the remaining text for urgency analysis.

4.3. Predefined Linguistic Resources:

The textual urgency assessment relies heavily on a fixed, high-quality knowledge base of urgency terms. Text Keyword Dictionary this is a predefined resource of explicit urgency terms (e.g., "URGENCY," "ASAP," "EMERGENCY," "IMMEDIATELY"). The dictionary must be comprehensive and unambiguous to ensure the Text Keyword Analysis component correctly flags straightforward urgent messages. Regular maintenance of this list is vital. Contextual Language Model (BERT) although not a dataset, the BERT uses its pre-trained linguistic knowledge to provide a score of implicit, contextual urgency. Its robustness and domain-specific fine-tuning are critical for inferring urgency when explicit keywords are absent.

4.4. Computational Models and Fusion Logic:

This involves the specialized machine learning models and the method used to combine their results into a single score. The Audio Urgency Model (OpenSMILE + Random Forest) is trained on a corpus of acoustic features to classify urgency based on voice alone. Its generalizability across different voices and emotional intensities is key to providing a reliable, non-textual urgency score. Urgency Analysis Fusion is the logic that combines the scores from the BERT model and the Random Forest model. The weighting and fusion algorithm must be carefully calibrated to ensure that the final, unified score accurately balances linguistic and emotional cues. This fusion process enhances the robustness of the system by reducing bias from any single model and improving consistency across diverse speech and language patterns. As a result, the integrated output delivers a more holistic and context-aware urgency evaluation for each voicemail message.

4.5. Priority Classification Parameters:

The final stage relies on calibrated parameters to translate the unified score into actionable priority levels and feed into the operational interface. Priority Thresholds these are the predefined score boundaries that segment the continuous urgency score into discrete categories (HIGH, MEDIUM, LOW). The accuracy of the entire system hinges on the calibration and stability of these thresholds to minimize false alarms and critical omissions. Final Priority Labels the output labels must be clear and directly actionable by the operations team. Ensuring that the output (HIGH PRIORITY OUTPUT) accurately reflects the operational meaning of the category is vital for efficient workflow integration.

4.6. User Interaction and Feedback:

This encompasses how the output is presented and how users can potentially contribute to system refinement. Dashboard Display the final priority and the transcribed text must be clearly presented on the user interface (e.g., a call management dashboard). The usability and filtering capabilities of this interface are crucial for end-users to quickly prioritize their work. Manual Review/Feedback although not explicitly shown in the first image, real-world systems include a mechanism for users to review the system's assigned priority against the actual voicemail. This user feedback loop is vital for gathering operational data to retrain and continuously improve the accuracy of the BERT and Random-Forest models.

5. Implementation of the Voicemail Prioritisation System

The implementation of the AI-powered voicemail prioritisation system provides an intelligent, real-time, and automated solution for identifying and ranking voice messages based on urgency through integrated text and audio analysis. By combining Speech Recognition, Natural Language Processing (NLP), and Machine Learning, the system efficiently transcribes, analyzes, and classifies voicemail messages into priority levels such as High, Medium, and Low. This framework aims to eliminate response delays, especially in sectors like healthcare, banking, and customer service, by ensuring that urgent calls receive immediate attention while maintaining data privacy through built-in PII redaction mechanisms.

5.1. System Architecture

The system architecture of the proposed AI-based voicemail prioritisation model is structured into four layers: Voicemail Capture Layer, Processing Layer, Classification Layer, and User Interface Layer. The Voicemail Capture Layer records voice messages through Twilio and securely stores them in the server with metadata such as caller ID, timestamp, and duration. The Processing Layer converts speech to text using the SpeechRecognition library and extracts audio features via OpenSMILE. The Classification Layer integrates BERT for textual urgency detection and Random Forest for analyzing tone, pitch, and energy. The User Interface Layer presents a web dashboard that displays voicemail recordings, transcriptions, urgency labels, and filtering options, ensuring seamless interaction and monitoring.

5.2. Audio and Text Integration

The system employs a dual-channel processing approach to handle both audio and text inputs simultaneously. The speech-to-text module converts voicemail audio into transcribed text, where BERT identifies urgency-related words such as “emergency” or “immediate.” Meanwhile, the audio analysis module processes tone and rhythm features like pitch and loudness using OpenSMILE, which are then fed into a Random Forest classifier. Both textual and acoustic urgency scores are combined using a weighted fusion algorithm to derive the final priority classification, ensuring higher accuracy than using either modality alone.

5.3. System Modules

The system comprises three major modules working in synchronization. The Voicemail Module records, stores, and retrieves incoming voice messages along with metadata and transcription. The Analysis Module serves as the core, where audio and text features are preprocessed and classified for urgency detection. The Dashboard Module allows users to visualize, filter, and respond to prioritized messages. It includes a secure login, a real-time message feed, and interactive controls to sort messages by urgency level. Together, these modules deliver a scalable and efficient voicemail management experience that minimizes response latency and human effort.

5.4. Model Training and Classification

The training process involves separate optimization of text and audio classifiers before multimodal fusion. The text-based BERT model is trained using a labeled dataset containing transcriptions of urgent and non-urgent messages, while the Random Forest model is trained using tone-based features extracted from the same dataset. Once trained, both models generate independent urgency probabilities, which are fused through a weighted voting strategy. The combined classifier achieves high accuracy and robustness, effectively distinguishing between critical, moderate, and low-priority calls across diverse message contexts.

5.5. Database and User Interface

The system utilizes a MySQL database for maintaining voicemail metadata, transcription text, and urgency classifications. The backend, developed using Flask, manages communication between the AI models and the user dashboard. The web interface, designed with HTML, CSS, and JavaScript, displays message lists, playback controls, and urgency indicators in a clean, interactive layout. Administrators can view system logs, analyze performance metrics, and manage user permissions, ensuring both transparency and usability across various deployment environments.

5.6. Security and Privacy

To ensure data confidentiality and compliance with privacy regulations, the system incorporates multiple security measures. The spaCy library detects and redacts personally identifiable information such as names, phone numbers, and email addresses before storage or display. All recorded voicemails and transcriptions are encrypted, and access is controlled through role-based authentication. The design adheres to organizational data policies and ensures ethical handling of sensitive communication data. Through these safeguards, the system delivers trustworthy, privacy-conscious voicemail management suited for professional and enterprise applications.

6. Requirements

6.1. Hardware Requirements

a. Processor: Intel Core i7-10850H and above

A high-performance processor such as the Intel Core i7-10850H (or equivalent) is required to efficiently handle AI-based voicemail analysis, including speech-to-text conversion, feature extraction, and urgency classification. It ensures smooth real-time processing of audio and text data while supporting concurrent system operations.

b. Primary Memory: 16GB DDR4 RAM, 3200 MHz and above

A minimum of 16GB DDR4 RAM is essential for handling large volumes of voicemail recordings, extracted features, and model inference. Adequate memory ensures efficient execution of BERT-based NLP models and Random Forest audio classifiers without lag during classification and dashboard updates.

c. Storage: 512GB Solid State Drive (SSD) and above

Fast storage such as a 512GB SSD is required to securely store voicemail audio files, transcriptions, user metadata, and system logs. SSDs improve data access speeds, reduce latency during real-time processing, and enhance reliability for continuous recording and playback functionalities.

d. GPU: NVIDIA GeForce RTX 3050, 6GB DDR6 and above

A dedicated GPU like the NVIDIA GeForce RTX 3050 with at least 6GB VRAM accelerates AI computations, particularly during model training and urgency classification. It enables rapid parallel processing for speech recognition and tone analysis tasks, improving accuracy and overall system performance.

6.2. Software Requirements

a. Front-end: HTML, CSS, JavaScript

The web interface is designed using HTML, CSS, and JavaScript to provide a responsive, user-friendly, and interactive dashboard. It allows users to play voicemails, view urgency labels, and filter messages by priority levels efficiently.

b. Back-end: Python Flask, MySQL

The back-end is developed using the Flask framework for real-time interaction between the front-end and machine learning modules. MySQL is used for maintaining voicemail metadata, transcriptions, and classification results, ensuring secure and efficient data handling.

c. Languages: Python, JavaScript

Python powers the core AI modules, including speech recognition, text analysis using BERT, and audio classification with Random Forest. JavaScript manages dashboard interactivity, ensuring smooth synchronization between data visualization and backend updates.

d. Tools: VS Code, Twilio API, OpenSMILE

VS Code is used for integrated development of both front-end and back-end components. The Twilio API handles telephony integration for voicemail recording and retrieval, while OpenSMILE extracts acoustic features such as pitch, loudness, and tone for audio-based urgency detection.

6.3. Functional Requirements

a. Voicemail Recording and Storage

The system must automatically record incoming calls through Twilio, generate a secure recording URL, and store both audio and metadata in a local database.

b. Speech-to-Text Conversion

The recorded audio must be converted into text using Speech Recognition to facilitate text-based urgency detection and transcription display.

c. Audio Feature Extraction

The system must analyze voice signals using OpenSMILE to extract tonal features like pitch, energy, and MFCCs for emotion and urgency recognition.

d. Urgency Classification

The system must classify messages into High, Medium, or Low urgency using a fusion of text-based BERT predictions and audio-based Random Forest outputs.

e. PII Detection and Redaction

The system must identify and mask sensitive information such as names, phone numbers, or email addresses using spaCy to ensure user privacy.

f. Dashboard Interaction

The web dashboard must display recorded voicemails, transcriptions, and urgency results in real time, allowing users to filter and sort messages by priority.

g. Department Routing and Alerts

The system must support routing of high-priority messages to relevant departments and generate alerts or notifications for immediate response.

6.4. Non-Functional Requirements

a. Security

The system must ensure privacy and data protection by encrypting stored audio files and restricting access to authorized users through role-based authentication.

b. Performance

The platform must provide near real-time transcription and urgency analysis with minimal delay, ensuring quick response to critical calls.

c. Scalability

The system must be capable of handling increasing voicemail volumes and user activity without compromising speed or performance.

d. Usability

The interface must remain simple and intuitive, allowing users with minimal technical knowledge to navigate recordings, view priorities, and manage messages effectively.

e. Accuracy

The AI models must maintain high accuracy in urgency prediction through optimized feature extraction and multimodal data fusion, minimizing false classifications and ensuring reliability in real-world applications.

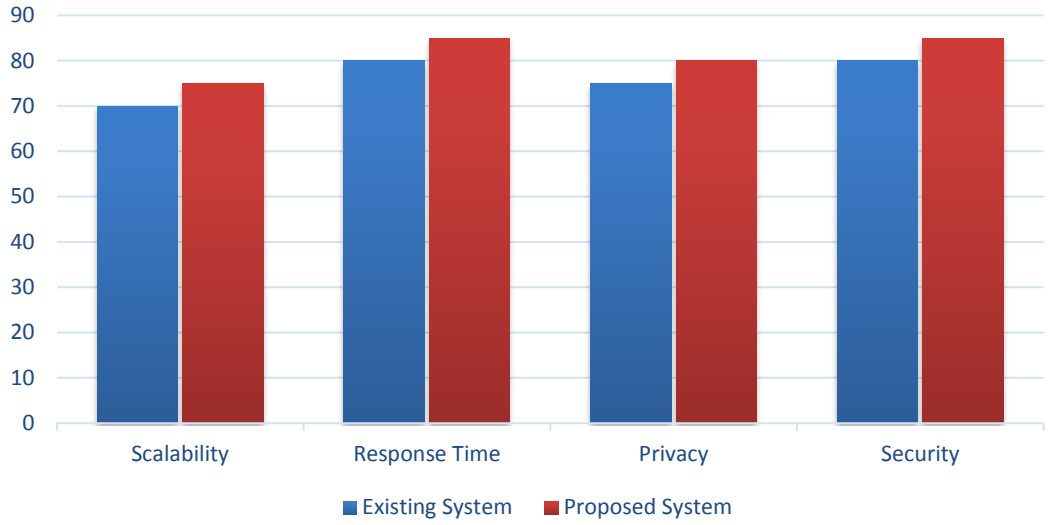
7. Result and Discussion

The comparative analysis between the existing and proposed systems clearly demonstrates the improved performance and reliability of the Automated Voicemail Priority Analysis System. As shown in the graph, the proposed system consistently performs better across all key parameters, including scalability, response time, privacy, and security. These improvements reflect the system's ability to manage a higher volume of data efficiently, respond more quickly to user queries, and ensure stronger protection of sensitive information. The enhanced results are a direct outcome of the architecture that integrates BERT (Text Model) with OpenSMILE and Random Forest (Audio Model), allowing for accurate urgency scoring and faster decision-making. Moreover, the improved coordination between text

and audio models enhances contextual understanding, reducing the chances of misclassification. The overall system thus ensures a balanced performance across technical and operational metrics, creating a robust framework suitable for enterprise-level deployment.

Unlike traditional manual triage, where prioritization relies on an individual's subjective judgment, the automated system provides a data-driven and objective approach to voicemail handling. Through parallel processing, the system simultaneously analyzes both linguistic and acoustic features, offering a comprehensive evaluation of each message. This enables the platform to automatically route urgent messages to the top of the Voicemail Management Interface, ensuring that critical communications receive immediate attention. The consistent performance and reduced human dependency contribute to higher efficiency and a more reliable workflow within contact center operations. Additionally, the system reduces cognitive load on employees, allowing them to focus on resolution rather than identification of critical issues. This streamlined workflow not only improves customer satisfaction but also optimizes workforce utilization within the organization.

Despite these advancements, adopting such a sophisticated AI-based system introduces implementation challenges. Many organizations, particularly those with limited infrastructure or technical expertise, may find it difficult to integrate and maintain NLP and acoustic models effectively. The initial deployment costs and potential transcription inaccuracies can also impact system performance. To address these issues, the focus should shift toward developing cloud-based and scalable solutions, integrating multilingual speech recognition, and refining user feedback loops for continual improvement. Continuous model retraining and real-time analytics integration could further enhance prediction accuracy and adaptability. With these enhancements, the proposed system can achieve greater adaptability, ensuring reliable, secure, and intelligent voicemail management on a broader scale.



8. References

[1]. J. Landesberger, ”Do the urgent things first! Detecting urgency in spoken utterances based on acoustic features,” Adjunct Proceedings of the 28th ACM UMAP Conference, 2020..

[2]. R. S. Rao, S. G. Suhasi, M. Rakshitha, P. Kumar, and K. G. R. Kishore, ”Implementing NLP to categorize grievances received via a voice input mechanism,” International Journal of Engineering Research & Technology (IJERT), 2023

[3]. D. Venkateshperumal, et al., ”Efficient VoIP communications through LLM-based real-time speech reconstruction and emergency call prioritization system,” International Journal of Computer Applications, 2025.

[4]. P. Ferri,D. Concone, and R. Riano, ”Deep ensemble multitask classification of emergency medical call incidents,” Artificial Intelligence in Medicine, 2021

[5]. S. Soner and R. Litoriya, ”Exploring acoustic feature extraction for urgency detection in voice messages using random forest models,” Wireless Personal Communications, 2022.

[6]. A. Trabelsi, S. Soussilane, and E. Helbert, ”Voicemail urgency detection usingcontext dependent and independent NLP techniques,” Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART), 2023.

[7]. H. Kamiyama, S. Shiota, and H. Kiya, ”Urgent voicemail detection focused on long term temporal variation,” IEEE International Conference on Acoustics, Speech andSignal Processing (ICASSP), 2019.

[8]. L. F. Parra-Gallego, ”Multimodal evaluation of customer satisfaction from voicemails using speech and language representations,” Digital Signal Processing, Elsevier, 2025.

[9]. M. A. Ullah, A. Ali, and T. Ahmed, ”Application of BERT for speech-to-text voicemail urgency classification,” International Conference on Natural Language Processing, 2022.

[10]. J. Zhang and K. Wang, ”Hybrid random forest and BERT-based approach for call center voicemail prioritisation,” IEEE Access, 2023.

[11]. F. Retkowski, ”Summarizing speech: A comprehensive survey,” ACM Computing Surveys, 2025.

[12]. N. Lukas, M. Zhang, and F. Kerschbaum, ”Personally identifiable information detection in voicemail transcriptions using NLP models,” Proceedings of the 31st USENIX Security Symposium, 2023.