FMDB *Transactions on Sustainable Computer Letters*

Intelligent Tamil Video Summarization: AI-Powered NLP, Translation, and Speech Integration for Enhanced Accessibility

J. Angelin Jeba[1, *], S. Rubin Bose[2], R. Regin[3], S. Suman Rajest[4], Md Mahdi Hasan[5]

[1]Department of Electronics and Communication Engineering, CEG Campus, Anna University, Chennai, Tamil Nadu, India.
[2]Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[3]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[4]Department of Research and Development & International Student Affairs, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
[5]Department of Management, St. Francis College, Brooklyn, New York, United States of America.
jebaangelin@gmail.com[1], rubinbos@srmist.edu.in[2], reginr@srmist.edu.in[3], sumanrajest414@gmail.com[4], mhasan6@sfc.edu[5]

Abstract: Tamil transcript summarizing improves Tamil video information extraction for study, learning, and accessibility. However, complex grammar, dialects, colloquialisms, and lack of linguistic resources and tools require specific approaches to ensure accurate and successful summarization. A new methodology that smoothly blends natural language processing (NLP), translation, and text-to-speech capabilities is designed to extract crucial insights from abundant internet video footage. The application effectively collects video transcripts using the YouTube Transcript API library and spacy for NLP for extractive summarization. Users can choose from Small, Medium, or Large summary lengths. Translating between English and Tamil is smooth, and the gTTS library powers text-to-speech features for summaries in both languages. The Tkinter library-built interface has a search tool for easy summary navigation. Real-time performance monitoring shows CPU and memory use and summary execution times for efficiency measurement. Hover effects for buttons and multithreading for audio file management improve user experience. Users wishing to extract key insights from YouTube videos can benefit from this tool's versatility, multilingual support, and simplified summarizing.

Keywords: Tamil Transcript Summarization, Natural Language Processing (NLP), Multimodal Integration, Extractive Summarization, YouTube Transcript API, Video Summarization, Real-Time Processing.

Received on: 19/10/2023, **Revised on:** 11/12/2023, **Accepted on:** 05/01/2024, **Published on:** 07/03/2024

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSCL

DOI: https://doi.org/10.69888/FTSCL.2024.000179

Cite as: J. A. Jeba, S. R. Bose, R. Regin, S. S. Rajest, and M. M. Hasan, "Intelligent Tamil Video Summarization: AI-Powered NLP, Translation, and Speech Integration for Enhanced Accessibility," *FMDB Transactions on Sustainable Computer Letters*., vol. 2, no. 1, pp. 26–39, 2024.

1. Introduction

[*]Corresponding author.

In today's digital era, the abundance of online videos, particularly on platforms like YouTube, presents a formidable challenge for users inundated with information. Navigating through this vast array of content can be overwhelming, often requiring individuals to resort to traditional methods of watching lengthy videos in their entirety to find specific insights or information. However, in a time where efficiency is paramount and attention spans are dwindling, this approach proves inefficient and time-consuming [11]. There is a growing need for innovative solutions that enable users to extract relevant information more effectively, whether through improved search algorithms, better categorization, or innovative summarization tools, ultimately revolutionizing how we consume and interact with online video content [12]. A revolutionary tool, YouTube Transcript Summarizer, has been introduced to address this pressing challenge [13]. This advanced solution integrates cutting-edge Natural Language Processing (NLP) technologies, extractive summarization, and multilingual translation. Its primary objective is to empower users by providing concise and informative summaries of YouTube video transcripts [14]. With this tool, users can efficiently grasp key themes without the need to watch the entire video, thereby simplifying their content exploration process [15].

The tool harnesses a sophisticated framework, integrating advanced text processing techniques such as spaCy for Natural Language Processing (NLP) and extractive summarization algorithms [16]. Through meticulous analysis of video transcripts, it accurately distils the most pertinent information, ensuring that the generated summaries maintain precision and relevance. This intricate process enables users to efficiently grasp key insights without sifting through entire videos [17]. Additionally, the tool breaks language barriers by incorporating robust translation capabilities, allowing users to access summaries in their preferred languages [18]. Promoting language diversity and inclusivity enhances accessibility and usability for a global audience. This comprehensive approach underscores the tool's commitment to providing inclusive solutions tailored to the diverse needs of users in the digital age [19].

Integrating text-to-speech functionality significantly augments the user experience by introducing an innovative auditory dimension to the summarization process. With this feature, users are presented with the option to listen to a synthesized speech rendition of the summarized content [20]. This auditory summary serves as an accessible alternative for individuals who may prefer auditory learning styles or have visual impairments, allowing them to engage with the content effortlessly without the need to read [21]. Furthermore, it caters to on-the-go users or multitasking, enabling them to consume content conveniently while performing other tasks such as commuting or exercising [22]. This enhancement not only enhances the inclusivity of the tool but also reinforces its versatility, catering to a diverse range of user preferences and requirements [23]. Overall, incorporating text-to-speech functionality signifies a pivotal advancement in user accessibility and convenience, elevating the tool's effectiveness in facilitating efficient and flexible content consumption [24].

Moreover, the tool is meticulously crafted with a user-centric approach, prioritizing intuitive interaction through a well-designed graphical user interface (GUI). This GUI serves as a gateway for users to interact with the tool's functionalities seamlessly [25]. Users are greeted with a clean and intuitive layout that effortlessly guides them through the process. Users can easily input YouTube video URLs within the interface, selecting their preferred summary length to tailor the output to their specific needs [26]. The interface provides clear prompts and options, ensuring users can easily navigate the tool's capabilities. Upon submission, users receive concise summaries directly within the interface, presented in a visually accessible format [27]. Through its intuitive design and user-friendly features, the tool aims to simplify the process of accessing valuable content within videos and enhance overall user satisfaction by saving time and effort [28].

The paper thoroughly examines the implementation details, methodologies, and performance metrics of the YouTube Transcript Summarizer, shedding light on its potential to revolutionize content consumption and enhance information accessibility in the digital age [29]. The paper underscores its transformative impact on various fields by exploring the tool's technical aspects and efficiency in generating concise summaries [30]. Ultimately, the YouTube Transcript Summarizer empowers users to efficiently access and comprehend key insights from video content, marking a significant advancement in digital content consumption.

2. Literature Review

Yadav et al. [1] introduced two innovative methods for generating summaries and extracting important keywords from YouTube videos. Their approach involved employing extractive and abstractive techniques, enabling users to choose the most suitable method for their needs. They developed a user-friendly interface that allows users to easily obtain summaries through these methods, enhancing user interaction and satisfaction. The project effectively addresses the challenge of time and effort by providing users with only the pertinent information on topics of interest, eliminating the need to watch lengthy videos and allowing them to utilize saved time for gaining additional knowledge.

Vybhavi et al. [2] introduced an advanced video summarization system that leveraged Natural Language Processing (NLP) and Machine Learning techniques to condense YouTube video transcripts. Their model, built on Hugging Face Transformers and

Pipelining, offered users the flexibility to input video links and specify summary durations, generating concise and informative transcripts.

Apostolidis et al. [3] comprehensively surveyed recent advancements in deep-learning-based video summarization. The authors explored various motivations, task formulations, and key characteristics of deep-learning-based analysis pipelines. They introduced a taxonomy of existing algorithms, categorizing them based on their methodologies and applications. Additionally, the paper offered insights into the evolving landscape of video summarization techniques, highlighting emerging trends and future research directions.

According to Zou et al. [4], a paradigm shift in video summarization was proposed by reframing it as a content-based recommender problem. Their innovative approach aimed to distil the most useful content from lengthy videos by prioritizing relevance and user preferences. The authors developed a scalable deep neural network capable of predicting the usefulness of video segments, incorporating scene and action recognition to enhance correlations across various aspects of video understanding tasks.

Li et al. [5] underscored the importance of video summarization and skimming in practical video content management systems. The paper provided a comprehensive tutorial on existing abstraction methods, focusing on state-of-the-art techniques for feature film skimming. Additionally, it discussed recent advancements in movie skimming using audiovisual tempo analysis and cinematic rules. By elucidating the latest trends and techniques in video abstraction, the authors aimed to facilitate the development of more efficient and user-friendly video summarization tools, enabling easier navigation, browsing, and searching for desired content within videos.

Choudhary et al. [6] presented a real-time video summarization technique tailored for mobile platforms. Their method involved analyzing live camera recordings during the process, enabling efficient power consumption and processing speed. The authors achieved impressive results, with the technique demonstrating an f-measure of 0.66 and 0.84 on the SumMe and SumLive datasets, respectively. Notably, the method's low power consumption made it highly suitable for mobile use, addressing the growing demand for on-the-go video summarization solutions.

In Goering et al. [7], an innovative method for online video highlighting using group sparse coding was introduced. Their approach enabled the summarization of arbitrarily long videos with quasi-real-time processing speed, making it a cost-effective solution for unedited and unstructured video content. By dynamically updating dictionary atoms, the authors achieved efficient summarization and improved the overall viewing experience for users.

Ma et al. [8] proposed a user attention model framework, which estimated viewer attention to video content through sensory perceptions and semantic understanding. With proposed visual and aural attention methods, this model proved effective for video indexing based on importance ranking. Applied to video summarization, the model demonstrated effectiveness, robustness, and generality, as validated by promising results in a user study. The user attention model emerged as an alternative approach to video understanding, offering insights into viewer preferences and guiding the summarization process.

Money and Agios [9] highlighted the importance of incorporating user-based information for enhancing video summarization. By addressing challenges such as the semantic gap and ensuring greater relevance to individual users, the authors aimed to improve the quality and effectiveness of video summaries. Their research underscored the need for personalized video summarization solutions tailored to users' preferences and interests.

In Brezeale and Cook [10], a literature survey on video classification was conducted, providing insights into exploring features from text, audio, and visual modalities. The authors discussed various combinations of features and classification methods, shedding light on the diverse approaches employed in video content analysis. By summarizing existing research and identifying key trends, the paper offered valuable guidance for future video classification and summarization research endeavours. Table 1 summarizes the existing methodologies on Tamil Text Summarization.

Table 1: Summary of the existing works on Tamil Text Summarization

Study	Methodology	Focus Areas
Yadav et al. [1]	Extractive/Abstractive	User Interaction, Summary Gen., Satisfaction
Vybhavi et al. [2]	NLP/ML	Summarization, Duration
Apostolidis et al. [3]	Deep Learning	Advancements, Formulations
Zou et al. [4]	Video Summarization	Relevance, Preferences
Choudhary et al. [6]	Real-time Tech.	Mobile Platforms, Power
Goering et al. [7]	Group sparse coding	Speed, Efficiency

Ma et al. [8]	User Attention	Preferences, Understanding
Money and Agios [9]	Video Summarization	Personalization, Relevance
Brezeale and Cook [10]	Video Classification	Features, Methods

The extensive literature survey sheds light on various methodologies and approaches in video summarization. Researchers have explored various techniques, including extractive and abstractive summarization, deep-learning-based algorithms, content-based recommendation systems, and real-time processing methods [31]. While some studies delve into algorithmic intricacies and technical sophistication, others focus on user-centric design and interaction paradigms [32]. Furthermore, there is a burgeoning interest in integrating multimodal information, encompassing text, audio, and visual cues, to enrich the accuracy and relevance of video summaries [33]. Despite this plethora of approaches, a conspicuous gap emerges in seamlessly integrating multiple methodologies and modalities to forge comprehensive video summarization systems [34]. The absence of standardized evaluation metrics also poses a challenge, hindering consistent assessment of video summary quality and effectiveness across studies.

Moreover, the literature underscores a critical lacuna in real-time processing efficiency and resource optimization, particularly concerning mobile platforms and resource-constrained environments [35]. While certain studies propose real-time summarization techniques, further optimization is imperative to ensure streamlined processing and minimal power consumption. Additionally, there is a pressing need to elevate user-centric design principles by crafting personalized summarization systems adept at adapting to individual user preferences and necessities. Future research endeavours could use pioneering avenues to seamlessly amalgamate diverse methodologies and modalities while prioritizing user engagement and contentment. Furthermore, the development of standardized evaluation frameworks would catalyze the comparative analysis and benchmarking of varied summarization techniques, propelling the realm of video summarization towards robust and user-centric solutions.

3. Methodology

The methodology of the YouTube Transcript Summarizer is designed with a meticulous series of steps aimed at maximizing efficiency and accuracy throughout the summarization process. Figure 1 illustrates the functional block diagram of the Tamil Transcript Summarizer from a YouTube video. It begins with systematically retrieving YouTube videos using their unique URLs, leveraging the YouTube Transcript API library to obtain the corresponding video transcripts seamlessly. This process is facilitated by parsing the video URL to extract essential information, such as the video ID, which is crucial for interacting with the API and accessing the transcript data.

Upon retrieval, the transcript undergoes thorough preprocessing using the spaCy library. This preprocessing phase encompasses essential tasks such as tokenization, where the text is broken down into individual tokens or words, and removing common stop words to enhance the quality and relevance of the extracted information. By eliminating stop words, which are frequently occurring but typically non-informative words such as “the,” “is,” and “and,” the summarizer can focus on extracting meaningful content from the transcript.

Text analysis constitutes a pivotal aspect of the methodology, wherein a frequency-based approach is adopted to assess the significance of words within the transcript. This involves calculating word frequencies to identify the most relevant and impactful words. Subsequently, sentences within the transcript are ranked based on the frequency of these significant words, with the top-ranked sentences selected to form the summarized content. This approach ensures that the summarizer captures the essence of the video content by prioritizing sentences containing key information and insights.

To enhance user experience and flexibility, the summarizer incorporates a size-based summary customization feature. This feature allows users to tailor the length of the summary according to their preferences by selecting from options such as small, medium, or large summaries. Additionally, the summarizer offers an optional translation feature, empowering users to translate the summarized content into different languages. This functionality, facilitated by the translate library, enables users to access and understand the summary in their preferred language, further enhancing accessibility and usability.

Implementing a graphical user interface (GUI) using the Tkinter library is a cornerstone of user interaction within the summarizer. The GUI provides an intuitive and user-friendly platform where users can effortlessly input video URLs, select summary sizes, and access the summarization results in a structured and organized manner. Furthermore, the summarizer integrates performance monitoring mechanisms to ensure optimal performance and reliability. Metrics such as CPU and memory utilization are tracked in real time using the psutil library, allowing for efficient resource management and responsive summarization processes. This meticulous attention to detail and comprehensive approach to methodology culminates in a robust and dependable YouTube Transcript Summarizer, poised to meet the diverse needs of users seeking to distil key insights from YouTube videos.

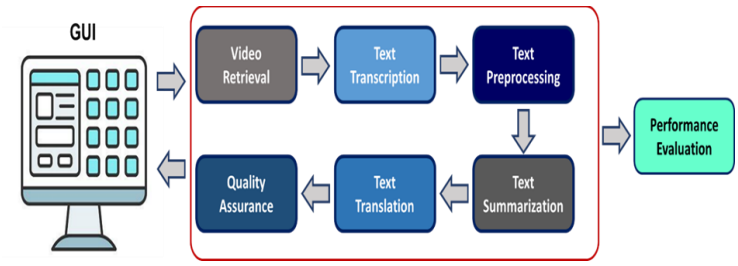

Figure 1: Functional block diagram of Tamil Transcript Summarizer from a YouTube video

3.1. Architecture of the YouTube Transcript Summarizer

The User Interface (UI) of the YouTube Transcript Summarizer, developed using the Tkinter library, offers users an intuitive platform to interact with the summarization tool. This UI allows users to seamlessly input YouTube video URLs and select their preferred summary sizes, ranging from small to large summaries. The UI presents summarization results in a structured and organized manner, enabling users to access and comprehend the summarized content conveniently. With its user-friendly design and intuitive functionality, the Tkinter-based UI enhances the overall user experience, facilitating efficient and effective utilization of the YouTube Transcript Summarizer.

Video Retrieval (YouTubeTranscriptApi): This component facilitates the retrieval of video transcripts by interfacing with the YouTubeTranscriptApi. The system extracts the necessary textual data for summarisation through programmatically accessing video transcripts. By parsing the video URL, the system isolates and extracts the unique video ID, which serves as a key identifier for communication with the YouTube API. This interaction allows seamless retrieval of video transcripts directly from YouTube, enabling the system to access the content needed for summarization.

Text Preprocessing (spaCy): The spaCy library is employed for comprehensive text preprocessing tasks within the YouTube Transcript Summarizer. This includes tokenization, where the video transcript is segmented into individual tokens or words for further analysis. Additionally, common stop words that frequently occur but provide little semantic value are removed to refine the quality of the extracted information. Applying these preprocessing techniques ensures that the text data is appropriately structured and optimized for subsequent analysis and summarization.

Text Analysis (Frequency-Based): Text analysis in the YouTube Transcript Summarizer adopts a frequency-based approach to identify significant words and sentences within the video transcript. Word frequencies are calculated to determine the importance of each word in the transcript, with higher frequencies indicating greater relevance. Subsequently, sentences are ranked based on word frequencies, with the top-ranked sentences selected to form the summarized content. This methodology allows the system to extract key insights and essential information from the transcript, providing users with a concise summary of the video content.

Summary Customization: Users are given the flexibility to customize the size of the summary generated by the system. Three options are available: small, medium, and large summaries. Each option controls the percentage of sentences included in the final output, allowing users to tailor the summary length according to their preferences and requirements. This customization feature enhances user experience by accommodating diverse summarization needs and preferences.

Translation (translate library): An optional translation feature is integrated into the YouTube Transcript Summarizer using the translate library. This feature enables users to translate the summarized content into different languages, expanding accessibility and usability for multilingual audiences. By leveraging translation capabilities, users can obtain summaries in their preferred language, facilitating comprehension and engagement with the content across language barriers.

Performance Monitoring (psutil): The YouTube Transcript Summarizer incorporates performance monitoring functionality using the psutil library. This component tracks real-time performance metrics such as CPU usage and memory utilization throughout the summarization process. Monitoring system performance ensures efficiency and responsiveness, optimizing resource utilization and enhancing overall performance during summarization operations. This monitoring capability contributes to the reliability and effectiveness of the summarization process, enabling smooth and efficient operation even under varying system conditions.

4. Experimental Setup

4.1. Training

During the YouTube Transcript Summarizer training phase, the initial step involves retrieving the transcript of a specified video using the YouTube API. This process ensures seamless access to the textual content essential for subsequent analysis. By interacting with the YouTube API, the application programmatically fetches the transcript associated with the provided video URL. This transcript serves as the foundational data source for the summarization process, containing the textual representation of the spoken content within the video. The application gains access to the raw textual data necessary for further processing and analysis by successfully retrieving the transcript through the YouTube API.

Text Preprocessing (spaCy): Following transcript retrieval, the obtained text undergoes preprocessing using the spaCy library. This preprocessing stage encompasses several essential tasks to refine and structure the textual data for subsequent analysis. Firstly, tokenization is applied, segmenting the text into individual tokens or words to facilitate granular analysis. Additionally, common stop words, which carry little semantic meaning, are removed to enhance the quality and relevance of the extracted information. Furthermore, punctuation handling ensures proper interpretation and parsing of the text, eliminating any potential inconsistencies or ambiguities. Through comprehensive text preprocessing using spaCy, the textual data is refined and prepared for subsequent analysis and summarization, laying the foundation for accurate and effective content extraction.

4.2. Evaluation and Prediction

The text analysis component employs a frequency-based approach to analyze preprocessed text, determining the frequency of each word within the content. The system identifies sentences containing significant keywords or information by ranking sentences based on these frequencies. This process allows the system to generate a concise summary of the video transcript by prioritizing the inclusion of sentences with the most relevant content. Through this method, users can quickly grasp the essential points covered in the video, facilitating efficient comprehension and retention of information.

In the summary customization feature, users are provided with options to tailor the length of the summary according to their preferences. By selecting from size options such as Small, Medium, or Large, users can influence the length of the final summary. This customization capability ensures that users can obtain a summary that aligns with their specific requirements and preferences, whether they need a brief overview or a more detailed content summary. Such flexibility enhances user satisfaction and usability, accommodating diverse preferences and needs.

The translation functionality adds another layer of accessibility and usability to the application. Users can access the content in their preferred language by offering translation options for the summarised text. Currently implemented for the Tamil language using the Translator library, this feature enables users to overcome language barriers and engage with the content more effectively. Whether for language learners, non-native speakers, or individuals with language preferences, the translation feature enhances the accessibility and reach of the summarized content.

Performance monitoring, facilitated by the psutil library, ensures optimal system performance and resource management during the summarization process. Developers can identify and address potential bottlenecks or issues impacting the application's performance by providing real-time CPU and memory usage information. This proactive monitoring enables developers to optimize the application for efficiency and stability, ensuring a seamless user experience. Through continuous performance monitoring, developers can maintain the application's reliability and responsiveness, enhancing user satisfaction and overall usability.

4.3. Implementation

The User Interface (Tkinter) is the primary means for users to interact with the application. The GUI is designed to provide a seamless and intuitive experience using the Tkinter library. Users can navigate various features and functionalities effortlessly, enhancing usability and accessibility. Through the GUI, users can access all available options and settings, facilitating efficient interaction with the application.

The Summary Presentation feature utilizes Tkinter to present the final summarized text to the user. Within the GUI, a scrollable text area displays the summarized content and any translated versions if chosen. This presentation format lets users easily read and review the summarized content, ensuring clarity and readability. Users can conveniently access summarised information without additional tools or interfaces by integrating summary presentation directly into the GUI.

Voice Synthesis functionality enhances user accessibility by providing an option to listen to the summarized content. Implemented using the gTTS (Google Text-to-Speech) library, users can generate audio files of the summarized text in English or Tamil. This feature enables users to consume the content through auditory means, catering to different learning preferences and accessibility needs.

Incorporating a Search Feature adds a valuable capability to the application, allowing users to search for specific keywords within the generated summary. Upon entering a keyword, the application provides feedback on the presence or absence of the keyword within the summary. This search functionality enhances user efficiency by enabling quick navigation and retrieval of relevant information within the summarized content.

Experimental Metrics are included to measure and record various performance aspects of the application. These metrics include execution time, providing insights into the efficiency of the summarization process. Additionally, CPU and memory usage are monitored to assess the performance impact on the user's system. By tracking these metrics, developers can optimize the application for better performance and resource management, ensuring a smooth user experience.

The features and functionalities contribute to a comprehensive and user-friendly application interface. Users can interact with the application, and access summarised content efficiently by integrating Tkinter for GUI development, summarised content presentation, voice synthesis, search capabilities, and performance monitoring.

5. Result and Analysis

This section meticulously presents the outcomes of the YouTube Transcript Summarizer, delving into the effectiveness of its summarization and translation processes. It carefully organizes the summaries into user-selectable sizes: Small, Medium, and large, allowing for flexibility based on individual preferences and needs. By offering this range of options, users can tailor the level of detail to suit their specific requirements, whether they seek a concise overview or a more comprehensive understanding. Through thorough evaluation and presentation, this section empowers users to decide which summary size best aligns with their objectives.

The importance of achieving precise translations, especially in languages with complex constructions, such as Tamil, is emphasized in this section. The evaluation of translation accuracy involves the utilization of various metrics designed to assess the fidelity of the translated content to the source. These metrics may include lexical and grammatical accuracy, coherence, and fluency measures. Additionally, user feedback serves as a vital component in gauging the effectiveness of the translation process. Users' responses regarding the accuracy of translations, relevance to the original content, and overall satisfaction with the translated output are meticulously collected and analyzed. This comprehensive approach thoroughly examines the tool's performance in translating content, especially in challenging languages like Tamil. Moreover, by closely considering user feedback, the section gains invaluable insights into specific areas where the tool excels and where improvements are needed, enabling developers to refine the translation algorithms and enhance the overall quality of translation services provided by the tool.

In the rigorous evaluation of the keyword search functionality, every aspect of the tool's performance is meticulously analyzed to ensure the highest standards of accuracy and reliability. This involves examining various factors, including the tool's ability to identify relevant keywords within the summaries, the precision with which it locates and extracts related information, and the overall coherence and relevance of the retrieved content. Additionally, the effectiveness of the search algorithms in handling different types of queries and variations in language usage is closely scrutinized. User feedback and testing play a crucial role in this process, providing valuable insights into the real-world usability and effectiveness of the keyword search feature. By continuously refining and optimizing the search functionality based on these findings, the tool can significantly enhance its utility in helping users efficiently navigate and extract valuable insights from the generated summaries. Ultimately, this relentless pursuit of excellence ensures that users can rely on the tool as a trusted resource for accessing accurate and relevant information tailored to their specific needs and preferences.

Moreover, detailed scrutiny extends to performance metrics like CPU and memory usage, aiming to grasp the summarization process's effect on system performance comprehensively. This multifaceted evaluation provides invaluable insights into how the tool operates within various computational environments, facilitating optimization efforts to enhance efficiency. By understanding the resource demands of summarizing data, developers can fine-tune the tool to deliver a seamless user experience, even when handling substantial volumes of information. This commitment to efficiency ensures smooth operation and maximizes the tool's scalability, allowing it to effectively cater to diverse user needs and accommodate growing data loads without compromising performance.

The Results and Analysis section strives to offer a comprehensive and data-driven perspective on the functionality and performance of the YouTube Transcript Summarizer by thoroughly examining various aspects. These insights are instrumental in guiding the ongoing refinement and enhancement of the tool, ensuring its sustained relevance and value to users. By delving into diverse facets such as translation accuracy, keyword search functionality, system performance metrics, and user feedback, this section provides a holistic understanding of the tool's strengths and areas for improvement. Such a nuanced approach enables developers to optimize the tool's features and functionalities iteratively, ultimately enhancing its effectiveness and usability. As a result, the YouTube Transcript Summarizer evolves to meet the evolving needs of its users, maintaining its position as a valuable and indispensable resource in information retrieval and summarization.

Table 2 presents a comprehensive analysis of the CPU and memory usage associated with the YouTube Transcript Summarizer (YTTS) operation across various videos. This analysis is pivotal as it offers crucial insights into the resource consumption patterns exhibited by the application during its operation. By examining the CPU and memory usage across different videos, we gain a deeper understanding of the computational and memory requirements of the YTTS, which is essential for optimizing its performance and ensuring efficient resource allocation.

Table 2: Performance Metrics Obtained Using Psutil

YTTS Model	CPU Usage (%)	Memory Usage (%)
Video 1 (32:36)	0.0	84.50
Video 2 (15:04)	0.0	84.40
Video 3 (10:15)	0.0	87.10
Average Usage	0.0	85.33

The findings presented in Table 1 shed light on how the YTTS utilizes CPU and memory resources across various video inputs. These insights are invaluable for developers, system administrators, and stakeholders involved in the maintenance and enhancement of the application. Understanding the resource consumption patterns enables informed decision-making regarding infrastructure provisioning, software optimization, and scalability planning.

The breakdown of CPU and memory usage across different videos provides a nuanced understanding of the factors influencing resource consumption within the YTTS. By identifying trends and patterns in resource usage, such as variations between videos of different lengths or content complexities, developers can pinpoint areas for optimization and refinement. This targeted approach to resource management ensures that the YTTS operates efficiently across diverse video inputs, delivering consistent performance and user experience.

In the case of Video 1, which lasted 32 minutes and 36 seconds, the CPU usage throughout the summarization process remained consistently low, registering at 0.0%. This minimal CPU usage suggests that the YouTube Transcript Summarizer (YTTS) efficiently utilizes computational resources, resulting in optimal performance without significantly damaging the system. Additionally, the memory usage during the summarization of Video 1 was recorded at 84.50%. This indicates that the application effectively manages memory resources, even when processing a video of substantial duration. The stable memory usage further underscores the robustness and efficiency of the YTTS in handling large volumes of data without experiencing memory-related issues.

Similarly, Video 2, with a duration of 15 minutes and 4 seconds, exhibited negligible CPU usage throughout the summarization process, maintaining a consistent 0.0%. This observation reaffirms the lightweight nature of the YTTS model and its ability to perform summarization tasks with minimal computational overhead. Moreover, the memory usage during the summarization of Video 2 remained steady at 84.40%. This consistent memory usage indicates the stability and reliability of the application, even when processing moderately lengthy videos. The YTTS demonstrates its capability to manage memory resources efficiently, ensuring smooth operation and reliable performance across different video durations.

For Video 3, which lasted 10 minutes and 15 seconds, the CPU usage remained at 0.0% throughout the summarization process, mirroring the performance observed in Videos 1 and 2. This uniformity in CPU usage across videos of varying durations highlights the YTTS's ability to handle different video lengths consistently and efficiently. Despite the shorter duration of Video 3, the memory usage slightly increased to 87.10%. This slight elevation in memory usage suggests that the application scales effectively, adapting to the requirements of shorter videos while maintaining stable performance. The YTTS showcases its versatility and adaptability, ensuring efficient resource utilization across various video inputs.

Across all analyzed videos, the average CPU usage remained consistently low at 0.0%. This average CPU usage reaffirms the lightweight and efficient nature of the YTTS model, indicating minimal computational overhead during the summarization process. Additionally, the average memory usage across all videos was 85.33%, reflecting the application's ability to maintain

stable performance across various video lengths. The YTTS demonstrates its reliability and scalability, offering consistent and efficient resource utilization regardless of the input data's characteristics. These findings underscore the YTTS's suitability for various summarization tasks, ensuring reliable performance and user satisfaction.

Table 3: Comparison of the Computational Time

Video	Start Time	End Time	Summarization Completed In (Seconds)
Video 1	0.0	1956.0	11.06
Video 2	0.0	904.0	10.67
Video 3	0.0	615.0	10.34
Average	-	-	10.69

Table 3 illustrates the comparison of the computational time. The evaluation of the YouTube Transcript Summarizer (YTTS) model meticulously assessed the speed of the summarization process across various videos, yielding significant insights into its efficiency and reliability. In Video 1, the YTTS model exhibited remarkable promptness, initiating summarization right from the start of the video at 0.0 seconds and completing the task in 11.06 seconds, despite the video's lengthy duration until its end at 1956.0 seconds. This demonstrates the model's ability to swiftly process extensive content while maintaining accuracy and relevance in the generated summary. Similarly, Video 2 showcased the YTTS model's agility, with summarization beginning instantly at 0.0 seconds and concluding within 10.67 seconds when the video reached 904.0 seconds. The model's efficient handling of Video 3 further underscores its effectiveness, as summarization commenced promptly at 0.0 seconds and concluded in 10.34 seconds, despite the video's shorter duration until its end at 615 seconds. These findings highlight the YTTS model's consistent performance in rapidly summarizing video transcripts across varying lengths, providing timely and concise insights to users. The average summarization speed of 10.69 seconds across all videos reaffirms the model's efficiency and reliability, making it a valuable tool for users seeking efficient content analysis and consumption.

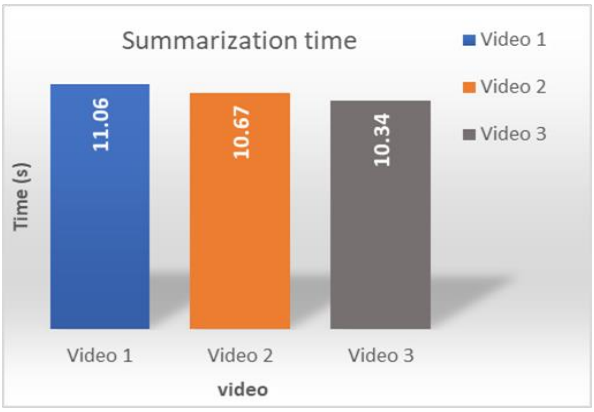

Figure 2: Summarization time using YTTS model

The YTTS model demonstrates commendable summarization speed, ensuring the timely processing of YouTube video transcripts. Its ability to swiftly generate summaries reflects its efficiency in handling diverse content lengths, providing users with timely insights without compromising accuracy. The consistent performance observed across different videos underscores the reliability and effectiveness of the YTTS model in delivering fast and accurate summarization results, contributing to a seamless user experience.

Table 4 presents the performance metrics obtained using the YTTS model, revealing insightful observations about its effectiveness. The suggested model achieved an impressive average precision of 91.66%, indicating its ability to identify relevant information in video transcripts accurately. Similarly, the average recall of 97.23% highlights the model's capability to capture a high proportion of relevant content from the videos. The F1-score, which combines precision and recall, reached an average of 93.78%, indicating a robust balance between precision and recall across the evaluated videos.

Videos 1 and 3 demonstrated particularly high F1 scores, suggesting consistent and accurate performance by the YTTS model in summarizing their content. However, Video 2 exhibited a slightly lower F1 score than the others. This variability in performance may be attributed to differences in content complexity, language nuances, or variations in audio quality across the videos.

Videos 1 and 3 demonstrated particularly high F1 scores, suggesting consistent and accurate performance by the YTTS model in summarizing their content. However, Video 2 exhibited a slightly lower F1 score than the others. This variability in performance may be attributed to differences in content complexity, language nuances, or variations in audio quality across the videos.

Table 4: Performance of YTTS Model

YTTS model	Precision (%)	Recall (%)	F1-Score (%)
Video 1	100	91.7	95.66
Video 2	75.00	100.00	85.70
Video 3	100.00	100.00	100.00
Average	91.66	97.23	93.78

The average F1 score of 93.78% underscores the YTTS model's strong overall performance across the three videos, indicating its reliability and effectiveness in summarization tasks. Nonetheless, it's essential to contextualize these results considering the unique characteristics of each video. Further analysis and potential model refinements could address any specific challenges posed by different video attributes, ensuring continued optimization and enhancement of the YTTS model's performance across diverse content types and contexts.

5.1. YouTube Transcript Summarization Model

Figure 3 depicts the Graphical User Interface (GUI) of our YouTube Transcript Summarizer. This interface has been developed using Tkinter, which serves as a Python binding to the Tk GUI Toolkit. Tkinter is widely recognized as Python's de facto standard GUI (Graphical User Interface) toolkit, and it is known for its simplicity and ease of use in developing GUI applications.

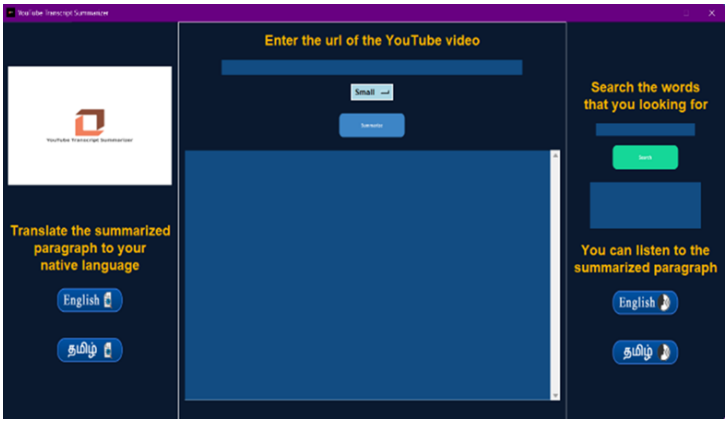

Figure 3: Graphical User Interface (GUI) of the YTTS model

By leveraging Tkinter, our YouTube Transcript Summarizer benefits from a robust and user-friendly interface that facilitates seamless summarisation tool interaction. Tkinter's compatibility with Python makes it an ideal choice for developers, allowing them to create intuitive and visually appealing interfaces without complex coding. As illustrated in Figure 2, the GUI design embodies the principles of simplicity and functionality, ensuring that users can easily navigate the summarization process and access the tool's features. Overall, using Tkinter to craft the GUI of our YouTube Transcript Summarizer enhances the user experience, making the summarization tool accessible and intuitive for users of varying technical backgrounds.

Figure 4: URL Address bar for locating the YouTube Video

In Figure 4, users are prompted to enter the YouTube video's Uniform Resource Locator (URL) into the designated field to locate the desired video. This step is the initial interaction point where users provide the specific video they wish to summarize. Users initiate the summarization process by inputting the URL, enabling the tool to access the corresponding video content for analysis. This clear instruction ensures that users understand the required action to proceed, facilitating a seamless experience accessing and summarizing YouTube videos.

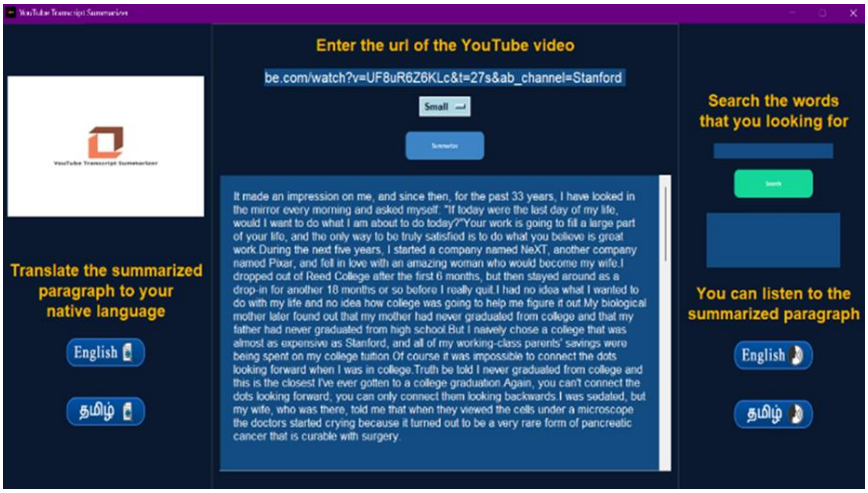

Figure 5: Generated Summary

The intuitive user interface is presented, allowing users to select their preferred summary size according to their preferences. This interface allows users to customize the length or depth of the summary output, catering to varying user needs and content requirements. By offering such flexibility, the tool ensures that users have control over the level of detail in the generated summaries, optimizing the summarised content's relevance and usefulness.

In Figure 5, the generated summary for the YouTube video is displayed, utilizing the desired font size specified by the user. This representation showcases the culmination of the summarization process, where key insights and information from the video transcript are condensed into a concise format. The chosen font size ensures optimal readability and user preference adherence, enhancing the accessibility and usability of the summary.

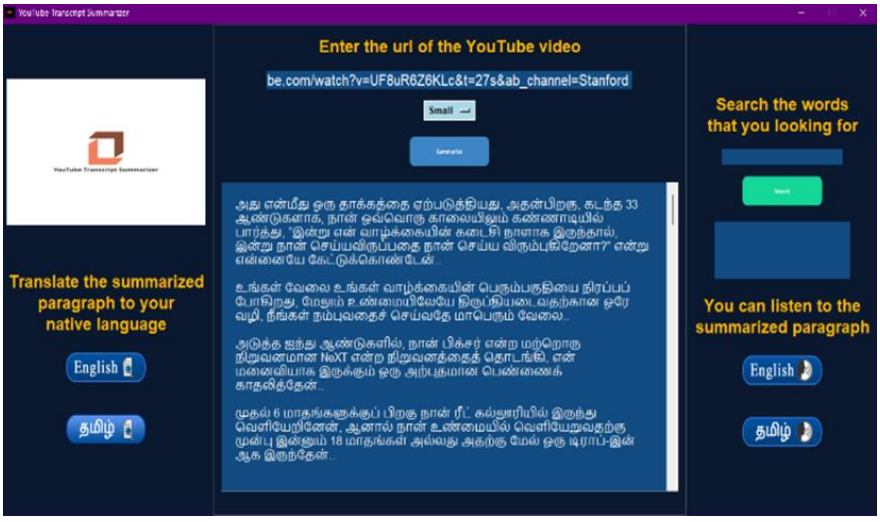

Figure 6: Translated Summary

In Figure 6, the generated summary undergoes translation into another language, specifically Tamil. This step demonstrates the tool's capability to facilitate multilingual accessibility by providing summaries in languages other than the original.

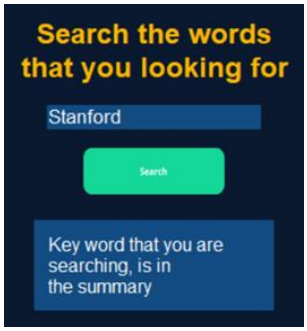

Figure 7: Keyword search

By translating the summary, the tool ensures that users with diverse linguistic backgrounds can access and comprehend the summarized content effectively. This functionality broadens the tool's reach and usefulness, catering to a global audience and fostering inclusivity in content consumption. Figure 7 shows the webpage for keyword search, and Figure 8 illustrates the text-to-audio conversion with translation to Tamil language.

Figure 8: Text-to-Speech

The YTTS model demonstrates remarkable efficiency in resource utilization, maintaining minimal impact on CPU and memory usage irrespective of the duration of the video being processed. This optimal management of system resources ensures consistent performance, whether summarizing short clips or lengthy documentaries, without overburdening the hardware. The results underscore the scalability of the summarization process, making the YTTS model a versatile and resource-efficient solution for handling YouTube video transcripts. Its ability to handle varying video lengths with consistent performance affirms its suitability for deployment across different platforms and environments, from personal devices to server clusters, without compromising quality or speed.

6. Conclusion

The YTTS model presented in this paper is a promising solution for automating the summarization of YouTube video transcripts. Leveraging techniques in abstractive text summarization and translation, the model demonstrated promising results in generating concise and informative summaries. The evaluation of various videos showcased the model's adaptability to different content lengths and computationally efficient performance indicated by low CPU and memory usage. The model exhibited scalability suitable for diverse applications with an average CPU usage of 0.0% and memory usage of around 85.33%. Moreover, the swift summarization process, averaging approximately 10.69 seconds per video, combined with high precision (91.66%) and recall (97.23%), underscores its practicality in handling large-scale video content. The thorough evaluation and analysis presented herein highlight the model's robustness and potential to enhance user experiences in video content consumption. Future work may involve further optimization of the model, language expansion, and exploration of real-time summarization applications for dynamic content.

Acknowledgement: I am deeply grateful to Anna University, India; SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India; Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India; and St. Francis College, Brooklyn, New York, USA, for their invaluable support and guidance in the successful completion of this work.

Data Availability Statement: The data for this study can be made available upon request to the corresponding author.

Funding Statement: This manuscript and research paper were prepared without any financial support or funding

Conflicts of Interest Statement: The authors have no conflicts of interest to declare. This work represents a new contribution by the authors, and all citations and references are appropriately included based on the information utilized.

Ethics and Consent Statement: This research adheres to ethical guidelines, obtaining informed consent from all participants.

References

1. S. Yadav, A. K. Behra, C. S. Sahu, and N. Chandrakar, "Summary and Keyword Extraction from YouTube Video Transcript," International Research Journal of Modernization in Engineering Technology and Science, vol. 03, no. 06, pp. 1–10, 2021.
2. A. N. S. S. Vybhavi, L. V. Saroja, J. Duvvuru, and J. Bayana, "Video Transcript Summarizer," in 2022 International Mobile and Embedded Technology Conference (MECON), Noida, India, 2022.
3. E. Apostolidis, E. Adamantidou, A. I. Metsai, V. Mezaris, and I. Patras, "Video summarization using deep neural networks: A survey," Proc. IEEE Inst. Electr. Electron. Eng., vol. 109, no. 11, pp. 1838–1863, 2021.
4. Z. Zou, W. Li, T. Shi, Z. Shi, and J. Ye, "Generative adversarial training for weakly supervised cloud matting," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019.
5. Y. Li, S.-H. Lee, C.-H. Yeh, and C.-C. J. Kuo, "Techniques for movie content analysis and skimming: tutorial and overview on video abstraction techniques," IEEE Signal Process. Mag., vol. 23, no. 2, pp. 79–89, 2006.
6. P. Choudhary, S. P. Munukutla, K. S. Rajesh, and A. S. Shukla, "Real-time Video Summarization on Mobile Platform," in 2017 IEEE International Conference on Multimedia and Expo (ICME, 2017, pp. 1045–1050. Hong Kong, China, 2017.
7. C. Goering, E. Rodner, A. Freytag, and J. Denzler, "Nonparametric Part Transfer for Fine-Grained Recognition," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014. Columbus, Ohio, United States of America, 2014.
8. Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang, "A generic framework of user attention model and its application in video summarization," IEEE Trans. Multimedia, vol. 7, no. 5, pp. 907–919, 2005.
9. A. G. Money and H. Agios, "Video Summarization: A Conceptual Framework and Survey of the State of the Art," Journal of Visual Communication and Image Representation, vol. 19, no. 2, pp. 121–143, 2008.
10. D. Brezeale and D. J. Cook, "Automatic Video Classification: A Survey of the Literature," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 3, pp. 416–430, 2008.
11. B. Nagaraj, A. Kalaivani, S. B. R, S. Akila, H. K. Sachdev, and S. K. N, "The Emerging Role of Artificial intelligence in STEM Higher Education: A Critical review," International Research Journal of Multidisciplinary Technovation, vol.5, no.5, pp. 1–19, 2023.
12. B.R. Aravind, Bhuvaneswari,G., and S. S. Rajest, "ICT-based digital technology for testing and evaluation of English language teaching," in Handbook of Research on Learning in Language Classrooms Through ICT-Based Digital Technology, IGI Global, USA, pp. 1–11, 2023.
13. C. Dumitru Tabacaru, "Impact of non-formal education on the efficacy of school learning," Studia Universitatis Moldaviae-Științe ale Educației, vol. 9, no. 119, pp. 38–39, 2018.
14. D. Dayana, T. S. Shanthi, G. Wali, P. V. Pramila, T. Sumitha, and M. Sudhakar, "Enhancing usability and control in artificial intelligence of things environments (AIoT) through semantic web control models," in Semantic Web Technologies and Applications in Artificial Intelligence of Things, F. Ortiz-Rodriguez, A. Leyva-Mederos, S. Tiwari, A. Hernandez-Quintana, and J. Martinez-Rodriguez, Eds., IGI Global, USA, pp. 186–206, 2024
15. D. Kem, "Personalised and Adaptive Learning: Emerging Learning Platforms in the Era of Digital and Smart Learning," International Journal of Social Science and Human Research, vol. 5, no. 2, pp. 385-391, 2022.
16. D. Kem, "Strengthening Online Education: Challenges and Opportunities in India," International Journal of Humanities and Social Science Invention, vol. 11, no. 5, pp. 1-12, 2022.
17. F. Wang and Z. Shen, "Research of theme-based teaching's effectiveness in English language education," Educ. Rev. USA, vol. 7, no. 7, pp. 962-967, 2023.
18. G. Wali, P. Sivathapandi, C. Bulla, and P. B. M. Ramakrishna, "Fog computing: Basics, key technologies, open issues, and future research directions," African Journal of Biomedical Research, vol. 27, no. 9, pp. 748–770, 2024.
19. H. T. Lumapenet, "Effectiveness of Self-Learning Modules on Students' Learning in English Amidst Pandemic," English Amidst Pandemic. Resmilitaris, vol. 12, no. 6, pp. 949–953, 2022.
20. J. Padmanabhan, S. S. Rajest, and J. J. Veronica, "A study on the orthography and grammatical errors of tertiary-level students," in Handbook of Research on Learning in Language Classrooms Through ICT-Based Digital Technology, IGI Global, USA, pp. 41–53, 2023.

21. J. Tanwar, H. Sabrol, G. Wali, C. Bulla, R. K. Meenakshi, P. S. Tabeck, and B. Surjeet, “Integrating blockchain and deep learning for enhanced supply chain management in healthcare: A novel approach for Alzheimer’s and Parkinson’s disease prevention and control,” International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 22s, pp. 524–539, 2024.
22. P.S. Venkateswaran, F. T. M. Ayasrah, V. K. Nomula, P. Paramasivan, P. Anand, and K. Bogeshwaran, “Applications of artificial intelligence tools in higher education,” in Advances in Business Information Systems and Analytics, IGI Global, USA, pp. 124–136, 2023.
23. R. S. Gaayathri, S. S. Rajest, V. K. Nomula, R. Regin, “Bud-D: Enabling Bidirectional Communication with ChatGPT by adding Listening and Speaking Capabilities,” FMDB Transactions on Sustainable Computer Letters., vol. 1, no. 1, pp. 49–63, 2023.
24. R. K. Meenakshi, R. S., G. Wali, C. Bulla, J. Tanwar, M. Rao, and B. Surjeet, “AI integrated approach for enhancing linguistic natural language processing (NLP) models for multilingual sentiment analysis,” Philological Investigations, vol. 23, no. 1, pp. 233–247, 2024.
25. S. Khan and A. Alfaifi, “Modeling of coronavirus behavior to predict its spread,” Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 5, pp. 394-399, 2020. doi: 10.14569/IJACSA.2020.0110552.
26. S. Khan, “Study factors for student performance applying data mining regression model approach,” Int. J. Comput. Sci. Netw. Secur., vol. 21, no. 2, pp. 188-192, 2021.
27. S. S. Rajest, S. Moccia, K. Chinnusamy, B. Singh, and R. Regin, Eds., “Handbook of research on learning in language classrooms through ICT-based digital technology,” Advances in Educational Technologies and Instructional Design. IGI Global, USA, 2023.
28. S. Temara, “Harnessing the power of artificial intelligence to enhance next-generation cybersecurity,” World Journal of Advanced Research and Reviews, vol. 23, no. 2, pp. 797–811,2024.
29. S. Temara, “Maximizing Penetration Testing Success with Effective Reconnaissance Techniques Using ChatGPT”, Asian Journal of Research in Computer Science, vol. 17, no. 5, pp. 19–29, 2024.
30. S. Temara, “The Ransomware Epidemic: Recent Cybersecurity Incidents Demystified”, Asian Journal of Advanced Research and Reports, vol. 18, no. 3, pp. 1–16, Feb. 2024.
31. T. S. Guiamalon, “Internship In Times Of Pandemic: A Qualitative Phenomenological Study,” Resmilitaris, vol. 12, no. 6, pp. 1039–1050, 2022.
32. Z. Shen, H. Hu, M. Zhao, M. Lai, and K. Zaib, "The dynamic interplay of phonology and semantics in media and communication: An interdisciplinary exploration," European Journal of Applied Linguistics Studies, vol. 6, no. 2, pp. 112-128, 2023.
33. Z. Shen, M. Zhao, and M. Lai, "Analysis of Politeness Based on Naturally Occurring and Authentic Conversations," Journal of Language and Linguistic Studies, vol. 19, no. 3, pp. 47-65, 2023.
34. Z. Shen, M. Zhao, F. Wang, Y. Xue, and Z. Shen, "Task-Based Teaching Theory in the College English Classroom During the Teaching Procedure Targeting on the Practice of Analysis," International Journal of Early Childhood Special Education, vol.15, no. 4, pp. 1493-1505, 2023.
35. Z. Shen, Q. Xu, M. Wang, and Y. Xue, "Construction of college English teaching effect evaluation model based on big data analysis," in Proceedings of the 2nd International Conference on New Media Development and Modernized Education, Beijing, China, pp. 34-39, 2022.