e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

SURVEY PAPER ON AUDIBLE INSTANT ALERT

Prof. Asmita Kamble[*1], Saurabh Pawar[*2], Kshitij Deshpande[*3],
Tejas Muluk[*4], Suyog Pisal[*5]
[*1,2,3,4,5]Sinhgad Institute Of Technology And Science, Pune 411041, Maharashtra, India.
DOI : https://www.doi.org/10.56726/IRJMETS51956

ABSTRACT

The "Vocalize Alert" system is a mobile application designed to empower blind individuals by providing a voice to their digital world. This innovative app serves as a lifeline for the visually impaired, transforming text-based chat notifications into spoken messages through text-to-speech (TTS) technology. Beyond accessibility, Vocalize Alert offers the flexibility to prioritize apps and chats, ensuring that critical information is promptly delivered through speech. This inclusive and user-centric application bridges the accessibility gap and enables seamless interaction with smart-phones. Vocalize Alert is a beacon of independence, connecting the blind community to the digital realm with clarity and convenience.

Keywords: Text-to-Speech (TTS), Mobile Application, Automated Voice Alerts.

I. INTRODUCTION

In a world increasingly driven by digital interactions, accessibility and inclusivity are paramount. The "Vocalize Alert" project emerges as a beacon of technological empowerment, designed to serve a specific yet significant community those with visual impairments. This innovative mobile application seeks to bridge the digital accessibility gap by providing auditory notifications for chat messages, ensuring that the visually impaired can seamlessly navigate the digital landscape.

At its core, Vocalize Alert harnesses the power of text-to-speech (TTS) technology to convert text-based chat notifications into clear and intelligible spoken language. This transformation offers a lifeline to individuals who rely on auditory cues to engage with their smartphones. Moreover, the app goes beyond mere accessibility; it empowers users by granting them the ability to prioritize apps and chats, ensuring that essential messages are promptly and distinctly communicated. In an era where digital interactions have become integral to our daily lives, addresses the specific needs of individuals with visual impairments, granting them not only access but also control over their digital world. The system significance lies in its commitment to making technology more inclusive and user-centric, providing a voice to those who have often been underserved in the digital realm. with the visually impaired can navigate their smartphones with newfound independence and clarity. It is this intuitive approach that allows to offer users with visual impairments a seamless and accessible experience. The Vocalize Alert system, while distinct in its goals, shares a fundamental principle with many cutting-edge technologies: the ability to make digital content more inclusive and user-friendly.

The methodology employed in "Vocalize Alert" draws inspiration from a two-phase principle:

1. In the first phase, the application scores each incoming chat notification, assessing the similarity and dissimilarity between messages.
2. The second phase involves decision-making, where the scores are evaluated, and notifications are prioritized and pronounced when considered important.

In the subsequent phase, a decision-making process unfolds, wherein the calculated scores are thoroughly assessed. Notifications are then prioritized based on their importance, and the selected ones are pronounced to the user. This two-phase approach ensures a systematic and efficient handling of incoming messages, allowing the application to provide users with a clear and prioritized spoken alert system tailored to their needs.

This transformative approach addresses the limitations of traditional notifications by providing a seamless auditory experience. The technical underpinnings involve the dynamic adaptation of TTS systems to diverse linguistic nuances, accommodating multilingual support and even addressing colloquial expressions and idiomatic language through machine learning-driven adaptability. In addition to its technical prowess, TTS notification systems adhere to stringent data privacy and ethical usage guidelines, promoting responsible application in various domains. The exploration of novel pricing models further underscores the commitment

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

to making this technology accessible to a broader audience, fostering inclusivity in the realm of digital communication.

Text-to-Speech (TTS) notification systems mark a significant advancement in human-computer interaction, particularly in the domain of digital communication and accessibility. convert NLP to Artifical Spoken language.

II. LITERATURE SURVEY

Farha Munmun et al. [1] and Diksha Khurana et al. [7] contribute to the field of Natural Language Processing (NLP) by applying advanced techniques and exploring various aspects of language-related technologies. This papers specifically focus on the application of Natural Language Processing speech synthesis, introducing a taxonomy for deep learning-based models, while also delving into Text-to-Speech (TTS) datasets and discussing key metrics for speech quality evaluation.

In a distinct study, Ahmed Tammaa and Ramy et al. [2] propose Neural TTS synthesis, comparing various vocoders. High Fidelity GAN (HiFi-GAN) surpasses competitors with a Mean Opinion Score (MOS) of 4.36 (vs. ground truth 4.45). WaveNet emerges as the closest-performing network with an MOS of 4.02, providing insights into vocoder efficacy in Neural TTS synthesis.

Dr. S.A. Ubale and colleagues highlight the internet's profound impact, particularly in communication, emphasizing Text-to-Speech (TTS) synthesis as a transformative tool. The TTS process involves Natural Language Processing and Digital Signal Processing technologies to convert written text into audible speech. On a related note, Akshay.A et al.[4] address low technology utilization in various sectors by proposing a system using an orange Pi(π) to extract text from images. The process involves capturing text images with a camera, processing them through image processing and noise reduction units, and utilizing Optical Character Recognition (OCR). The orange Pi(π) then converts the detected text into an audio signal, overcoming limitations of existing systems.

In [5], Mrs. Vijaya Bhosale et al. delve into Intelligent Virtual Assistants (IVAs), employing voice and contextual data to offer assistance, encompassing language understanding, social interaction, speech generation, and context for user responses. The term IVA is used interchangeably with expressions like Individual Digital Assistants, Informal Agents, Virtual Personal Assistants, Voice Activated Personal Assistants, or Voice-Enabled Assistants. Notable voice-activated IVAs such as Siri, Google Assistant, Microsoft Cortana, and Amazon Alexa are widely adopted in smartphones and households. In a related study by Amrita S. Tulshan et al.[6], virtual assistants like Siri, Google Assistant, Cortana, and Alexa are celebrated as a 21st-century boon, facilitating human-like interactions with machines. The paper addresses challenges in voice recognition and contextual understanding through a survey involving 100 users, aiming to enhance virtual assistant usage by tackling these issues.

Xu Tan, Tao Qin et al.[8] focus on Text-to-Speech (TTS), a prominent research topic in speech, language, and machine learning communities. TTS, or speech synthesis, aims to produce intelligible and natural speech from given text, with extensive applications across various industries. The paper highlights the substantial advancements in TTS quality achieved through neural network-based approaches, driven by developments in deep learning and artificial intelligence.

Chenshuang Zhang et al.[9] Generative AI has demonstrated impressive performance in various fields, among which speech synthesis is an interesting direction. With the diffusion model as the most popular generative model numerous works have attempted two active tasks: text to speech and speech enhancement. This work conduct a survey on audio diffusion model.

III. GAP ANALYSIS

According to literature review reveals a lack of well- defined evaluation metrics and a scarcity of training data for the domain-specific Text-to-Speech (TTS) task. Despite its strengths in accessibility and language diversity, TTS technology faces challenges such as emotionless speech, mispronunciations, and limited adaptability to colloquial language. Recognizing these limitations, our research aims to enhance TTS voices by refining pronunciation algorithms, improving context comprehension, and addressing ambiguity. We plan to invest in further development to improve TTS with emotional nuances, adaptability to idiomatic expressions, and slang through machine learning. Our strategy also includes ensuring data privacy and ethical usage

guidelines, promoting responsible application and exploring more accessible pricing models.

Gap Analysis based on following statements :-

- **Semantic Segmentation Techniques**: In order to guarantee the precise identification and isolation of damaged regions in road-surface photographs, the study highlights the necessity for more research into improving semantic segmentation approaches.
- **Computational Efficiency:** There is a need to overcome the shortcomings of current auto- encoding methodologies, which are frequently sluggish when processing big datasets, by developing quicker computing techniques while maintaining high accuracy**.**
- **Standardization and Benchmarking:** The lack of established benchmarks and performance metrics related to road-surface-damage object detection creates a gap in the evaluation and comparison of different algorithms, impeding the formation of a uniform performance standard.
- **Real-World Application Validation:** To verify the Proposed Net model's usefulness and resilience in actual applications, more research is required to validate it in real-world scenarios and different environmental circumstances.
- **Dataset Diversity:** Improving the diversity and volume of the datasets used for training and validation might close the gap in terms of complete data representation, allowing for more accurate and dependable model performance over a wide range of road-surface conditions.

IV. PROPOSED ARCHIECTURE

The Flask-based backend infrastructure guarantees smooth communication between the application and the CNN model. The solution also includes geo-location data processing and database connectivity to securely store and manage the gathered road damage data. A specialized website is built for local governments to view and update the status of maintenance and repaired places, allowing for more efficient communication and collaboration. To assure the system's functioning and dependability, rigorous testing and validation processes are carried out. The detailed reporting and documentation of the whole development process further explain the methodology, technical requirements, and system architecture, stressing the project's contribution to efficient road maintenance and management methods.

The specialized website designed for local governments can offer advanced analytics and reporting tools, enabling comprehensive insights into road conditions and maintenance needs. Continuous optimization efforts, such as machine learning model updates and adaptive algorithms, contribute to the system's longevity and effectiveness. The emphasis on user-friendly interfaces and accessibility features further enhances the collaboration and communication efficiency for local government officials.

Furthermore, the emphasis on user-friendly interfaces goes beyond mere accessibility, encompassing intuitive design elements that streamline navigation and functionality. These features not only make the platform more user-friendly for local government officials but also contribute to heightened collaboration and communication efficiency. Officials can easily access and interpret data, fostering quicker decision-making processes and facilitating prompt responses to road maintenance issues. The integration of modern communication tools may enhance collaboration, allowing officials to share real-time updates and coordinate efforts effectively, ultimately contributing to more efficient and proactive road maintenance strategies.

The platform's features facilitate quick and informed decision-making by providing officials with easily accessible and comprehensible data related to road conditions and maintenance needs. Real-time updates, made possible through the integration of modern communication tools, enable officials to stay informed about ongoing maintenance activities, road conditions, and emerging issues. This real-time collaboration fosters a more proactive approach to road maintenance, allowing officials to coordinate efforts effectively and address challenges promptly. The use of these modern tools contributes to an efficient and responsive road maintenance strategy that aligns with the dynamic needs of the infrastructure.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

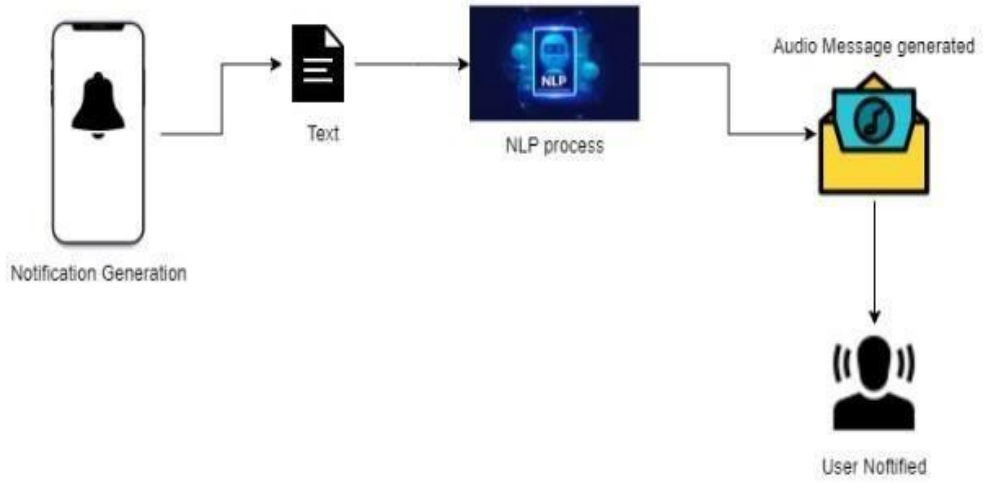

Figure 1: System Architecture of Vocalize Alert System.

Moreover, the suggested approach highlights the CNN model's careful architecture, which incorporates essential architectural components to improve its ability to detect road damage accurately. The CNN model is designed to extract complex characteristics and patterns from a variety of road photos by carefully configuring its numerous convolutional layers. Efficient feature extraction with minimal computing cost is possible with the use of suitable activation functions and pooling layers.

Furthermore, the presence of fully linked layers enables thorough learning of intricate correlations within the retrieved characteristics, enabling the model to identify minute differences between various kinds of road damage. To reduce overfitting and improve the model's capacity for generalization, the methodology makes use of regularization strategies such batch normalization and dropout.

The survey paper aims to provide a thorough understanding of the model's architecture and its crucial role in the accurate identification and classification of road damages for effective and timely road maintenance practices by going into detail about the CNN model design within the proposed methodology.

Additionally, the incorporation of the CNN model into the suggested technique highlights its flexibility and scalability, enabling the effective analysis of various road photos in a range of geographical locations and climatic circumstances. With the use of deep learning, the CNN model is able to identify and classify many sorts of defects, such as cracks, potholes, and surface deformations, with accuracy.

It also shows a strong capacity to detect subtle patterns and abnormalities within the road photos. The CNN model's practical applicability in real-world scenarios is highlighted by its seamless integration with user-friendly applications and backend infrastructures. This facilitates seamless communication between end users and local authorities, resulting in streamlined data collection, analysis, and timely decision-making.

This survey paper highlights the potential of advanced deep learning techniques in revolutionizing the field of road infrastructure management, paving the way for safer and more resilient transportation networks globally. It does this by clarifying the complex interactions between the CNN model and the larger system architecture.

V. CONCLUSION

This paper based on NLP and Deep Learning methods examines the Vocalize Alert system, a pioneering effort to improve digital accessibility for people with visual impairments. The author use Natural Language Process(NLP) and Deep Learning(DL) by applying advanced techniques and exploring various aspects of language-related technologies. It highlights the importance of making digital platforms accessible in an era dominated by smartphones and digital communication. Vocalize Alert, a text-to-speech application, aims to enable visually impaired users to access digital text in spoken form and manage notifications effectively. The study, based on 10 papers, outlines the necessity for digital inclusivity and the project's response to this need, including its objectives formulated from identified gaps. At the heart of Vocalize Alert is an advanced text-to-speech engine that uses neural networks and natural language processing to transform text into expressive speech. It features cutting-edge AI for adaptive, multilingual translation and personalized notification management through machine learning, which improves with user feedback. The system's technical

architecture is designed for real-time, context-aware language processing, offering a natural, human-like voice output, underscoring its innovative approach to fostering digital inclusivity.

VI. REFERENCES

[1] Taylor, P. Black, A. (2017). Recent Developments in Speech Synthesis. In The Oxford Handbook of Multimodal Analysis.

[2] Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... Sotelo, J. (2017). Natural vocalize alert synthesis by conditioning WaveNet on mel spectrogram predictions. In 2018 IEEE International Conference on Acoustics, Speechand Signal Processing (ICASSP) (pp. 4779- 4783).

[3] Tjandra, A., Vo, Q. N., Li, L. (2017). Listening while speaking: Speech recognition for whispered speech. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5670-5674).

[4] Ribeiro, A., Sampaio, J., Oliveira, L. S. (2016). A new text-to-speech synthesis system based on the HTS Platform is In the Processing of 9th ISCA speech Synthesis Workshop.

[5] Epstein, R., Golan, O., Simmons, R. (2016) Synthesized speech prosody and the acoustic realization of focus In Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPEECH) (pp.3161-3165).

[6] Diksha Khurana , C.Purnima “Assistive System for Product Label Detection with Voice Output For Blind Users” International Journal of Research in Engineering Advanced Technology 2020.

[7] Chenshuang Zhan, “Review on Text-To-Speech Synthesizer,” International Journal of Advanced Research in Computer and Communication Engineering, 2015.

[8] Wu, Z., Watts, O., King, S.Fitt, J. (2016). Merlin: An open source neutral network speech synthesis system. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASS).