CompSci Journal
Double Blind Peer Reviewed, Open Access International Research Journal.
ISSN: 3067-3089| Website: https://csjour.com | Volume-1, Issue-1 | February - 2025

Original Article

MediTalker: Transforming Medicine Text into Speech

Dhanashri Bhosale[1], Rutuja Gaikwad[2], Dhnayata Phule[3], Pranjali Shinde[4], Shreya Somadale[5]
[1,2,3,4,5] *Department of computer Science and Engineering, SKN Sinhgad College of Engineering, Korti, Pandharpur, India*

Manuscript ID:
CSJ-2025-010101

Volume 1

Issue 1

Pp. 1-6

February 2025

ISSN: 3067-3089

Submitted: 15 Jan. 2025

Revised: 20 Jan. 2025

Accepted: 15 Feb. 2025

Published: 28 Feb. 2025

Correspondence Address:
Bhosale, D., Gaikwad, R., Phule, D., Shinde, P., & Somadale, S. (2025). MediTalker: Transforming Medicine Text into Speech. Compsci Journal
Email:
dhanashri.surwase@sknscoe.ac.in

Quick Response Code:

Web. https://csjour.com/

DOI:
10.5281/zenodo.17324265

DOI Link:
https://doi.org/10.5281/zenodo.17324265

Abstract

MediTalker is a smart healthcare solution designed to enhance accessibility for individuals with weak eyesight by converting medication text into audible information. Utilizing Optical Character Recognition (OCR) and Speech Synthesis technologies, MediTalker aims to provide clear and understandable medication details that are otherwise inaccessible due to unreadable packaging. The system processes printed text on medication labels and transforms it into speech, enabling users to hear critical medication information, such as dosage, instructions, and warnings. This solution empowers individuals with visual impairments to independently manage their medications, improving safety and quality of care. This paper discusses the architecture of MediTalker, its implementation, and its impact on visually impaired individuals, highlighting key use cases and future developments in assistive technologies.

Keywords: *Optical Character Recognition (OCR), Speech Synthesis, Smart Healthcare, Assistive Technology, Medication Accessibility, Visual Impairment, Natural Language Processing (NLP).*

Introduction

MediTalker is designed to help individuals with weak eyesight access essential medication information by converting printed text into speech. Using Optical Character Recognition (OCR) and Speech Synthesis, the system scans medication labels and audibly communicates details such as dosage, usage, and warnings.

This project aims to provide a simple, effective solution to improve medication management for visually impaired individuals, promoting safety and independence through accessible technology.

a) *What is Natural Language Processing?*

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on enabling computers to understand, interpret, and respond to human language. It involves analyzing text or speech data to enable machines to process, extract meaning, and generate responses in a way that mimics human communication. NLP is used in various applications like speech recognition, chatbots, translation, and sentiment analysis..

b) *How does NLP work?*

NLP works by processing and analyzing human language using techniques like tokenization (breaking text into smaller units), parsing (analyzing sentence structure), and semantic analysis (understanding meaning). It involves pre-processing text to remove unnecessary elements, applying machine learning models to recognize patterns, and extracting meaning. The system then uses this understanding to generate responses or perform tasks like translation, sentiment analysis, or text summarization.

What is OCR?

Optical Character Recognition (OCR) is a technology that enables the conversion of different types of written or printed text, whether in images, scanned documents, or photographs, into machine-readable and editable digital text. It essentially allows computers to recognize and interpret characters (letters, numbers, and symbols) from a scanned image of a physical document, making it possible to process and manipulate the text in digital form. Here's how OCR works and its key components

HOW OCR WORKS

i. **Image Acquisition:**

The first step in OCR is capturing the image that contains the text. This can be done using a scanner, camera, or other image-capturing devices. The image could be of a printed book, a receipt, a sign, or handwritten notes.

How to cite this article:
Bhosale, D., Gaikwad, R., Phule, D., Shinde, P., & Somadale, S. (2025). MediTalker: Transforming Medicine Text into Speech. Compsci Journal, 1(1), 1–6. https://doi.org/10.5281/zenodo.17324265

CompSci Journal
Double Blind Peer Reviewed, Open Access International Research Journal.
ISSN: 3067-3089| Website: https://csjour.com | Volume-1, Issue-1 | February - 2025

ii. **Preprocessing**:
Before the OCR engine processes the image, several preprocessing steps are performed to improve recognition accuracy.

iii. **Character Recognition**:
Once the image is preprocessed, OCR software uses pattern recognition or feature extraction to identify the characters

iv. **Post-Processing**:
After recognizing the characters, OCR applies additional processing to improve accuracy

v. **Output**:
The final output of OCR is the digital text that can be stored, searched, edited, or otherwise manipulated. This digital text can be used in various applications such as word processing, translation, indexing, and text-to-speech conversion.

Methodology:

The OCR process begins with **image acquisition**, where a document is scanned or captured via a camera. Next, the image undergoes **preprocessing**, which involves converting it to black-and-white (binarization), reducing noise, and segmenting the text into lines, words, and individual characters. In the **character recognition** phase, OCR software uses pattern matching or feature extraction to identify characters by comparing them with a database of known text patterns. Finally, in **post-processing**, the system corrects any recognition errors through contextual analysis, such as using language models or spell-checkers, to deliver accurate, editable digital text.

a) Architecture Design

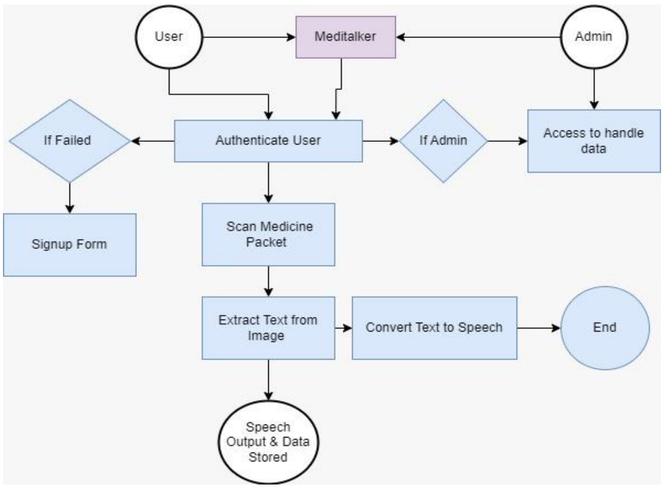


Fig 1: Block Diagram for MediTalker

b) Data Collection

For the MediTalker project, effective data collection includes compiling a diverse medical text corpus from trusted sources, such as medication names, dosages, and usage instructions. It's essential to gather user input from individuals with visual impairments to understand their specific needs and preferences regarding medication information. Additionally, recording clear speech samples of this information will help train the speech synthesis model, while collecting user feedback on the clarity and comprehensibility of the generated speech will enable continuous improvements to the system.

c) Data Preprocessing

Data processing for the MediTalker project involves several key steps: first, text preprocessing is performed to clean and normalize the medical text by removing unnecessary characters, correcting spelling errors, and standardizing terminology. Next, tokenization breaks down the processed text into manageable units (words or phrases) for easier analysis and transformation. Semantic analysis is then conducted using NLP techniques to extract meaningful information, such as medication instructions and context, enhancing overall understanding. Finally, the processed text is converted into a format suitable for speech synthesis, including prosody and pronunciation adjustments to ensure natural-sounding output..

d) Feature Extraction

Feature extraction for the MediTalker project involves identifying and selecting relevant characteristics from the processed medical text to improve the performance of the speech synthesis model. This includes extracting key features such as medication names, dosages, administration routes, and specific instructions. Additionally, linguistic features such as part-of-speech tags, named entity recognition, and syntactic structures are derived to enhance context understanding. By focusing on these critical features, the model can generate more accurate and contextually relevant speech output, ensuring that users receive clear and comprehensive medication information.

e) Model Training

Model training for the MediTalker project involves creating a dataset of audio recordings paired with text to facilitate

CompSci Journal
Double Blind Peer Reviewed, Open Access International Research Journal.
ISSN: 3067-3089| Website: https://csjour.com | Volume-1, Issue-1 | February - 2025

supervised learning. The model uses algorithms like Tacotron or WaveNet to learn the mapping from text to speech, optimizing parameters for natural output. Techniques such as data augmentation enhance robustness, while evaluation metrics like Mean Opinion Score (MOS) ensure the generated speech is clear and intelligible for users. Fine-tuning based on feedback helps refine the model's performance.

f) User Interface Design

Web Templates: Designed and implemented web templates using HTML, CSS, and MERN Stack templating engine. User Interaction: Provided user-friendly interfaces for adding new users, viewing attendance records, and initiating the face recognition process.

g) Future Enhancements

Deep Learning Integration: Considered the integration of deep learning models for improved accuracy and robustness. Optimization: Explored optimization techniques to enhance the system's real-time performance. This methodology outlines the step-by-step process undertaken to develop and implement the face recognition attendance system. Adjustments and additional steps may be necessary based on specific project requirements and constraints.

Algorithms

A. Optical Character Recognition (OCR) process

1. Image Acquisition: Capture a digital image of the text.
2. Preprocessing: Convert to grayscale, reduce noise, binarize, and deskew the image.
3. Text Detection: Segment the image into individual characters or words.
4. Feature Extraction: Extract relevant features from the segmented characters.
5. Character Recognition: Use a classification model to recognize characters.
6. Post-processing: Correct misrecognized text and restore formatting.
7. Output: Generate the final recognized text.
8. Evaluation: Assess accuracy and refine the model as needed.

B. text preprocessing Algorithm

1. Input Text: Start with the raw text.
2. Lowercasing: Convert text to lowercase.
3. Remove Punctuation: Eliminate punctuation marks.
4. Remove Numbers: Optionally remove digits.
5. Tokenization: Split text into tokens (words).
6. Remove Stop Words: Remove common non-essential words.
7. Stemming/Lemmatization: Reduce words to their base form.
8. Remove Extra Whitespace: Clean up excess spaces.
9. Output Processed Text: Generate the final text.

Result and Analysis

Upon inputting parameters corresponding to Medicine Wrapper, the system promptly verifies and matches text in real-time. Validation procedures ensure that input parameters adhere to specified ranges; deviations or omissions trigger appropriate error handling. Python programming forms the basis for implementing Machine Learning and NLP classification techniques, including Decision Tree, Random Forest, SVM, OCR, augmenting the system's accuracy and efficacy in facial recognition.

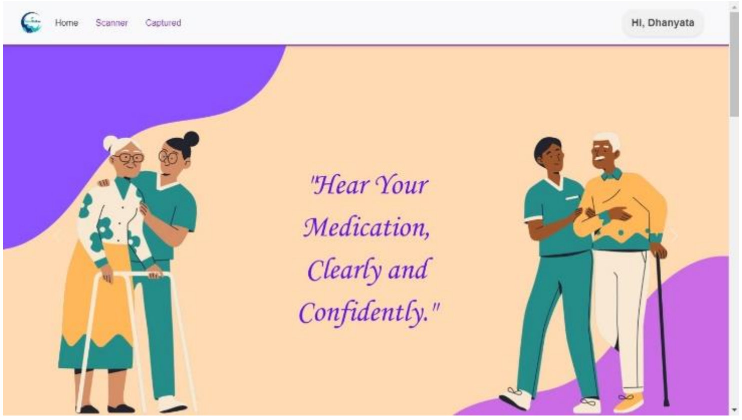
Conclusion

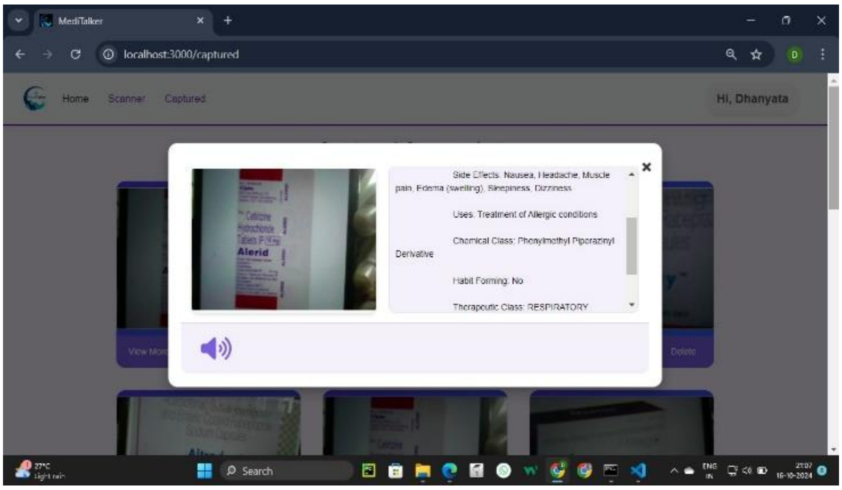
In conclusion, the development and implementation of In conclusion, the MediTalker project harnesses advanced NLP and speech synthesis techniques to transform medical text into audible information, significantly benefiting individuals with visual impairments. By systematically collecting and processing data, extracting relevant features, and training robust models, the project aims to deliver clear and comprehensible medication details. Continuous evaluation and user feedback are integral to refining the system, ensuring that it meets the needs of its users effectively. Ultimately, MediTalker strives to enhance accessibility to critical health information, improving the quality of life for those who rely on it.

Result

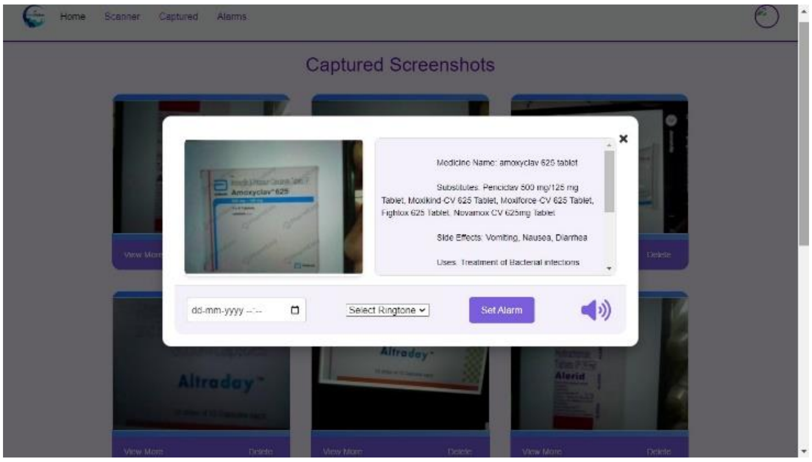
MediTalker Working Flow:

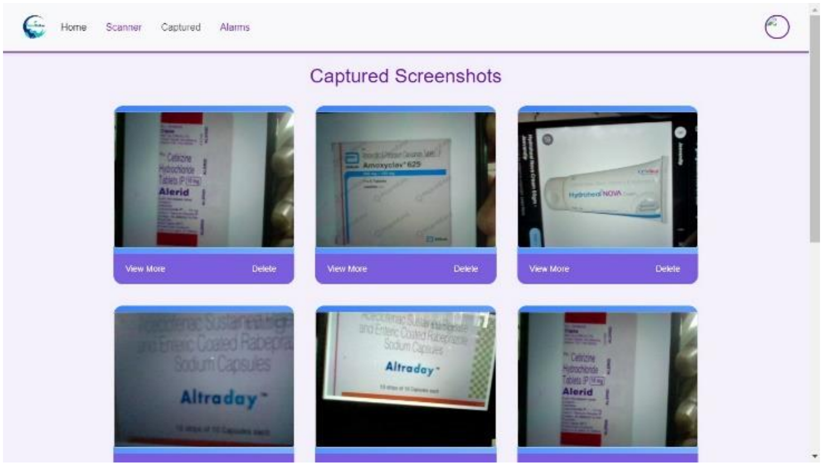
CompSci Journal
Double Blind Peer Reviewed, Open Access International Research Journal.
ISSN: 3067-3089| Website: https://csjour.com | Volume-1, Issue-1 | February - 2025

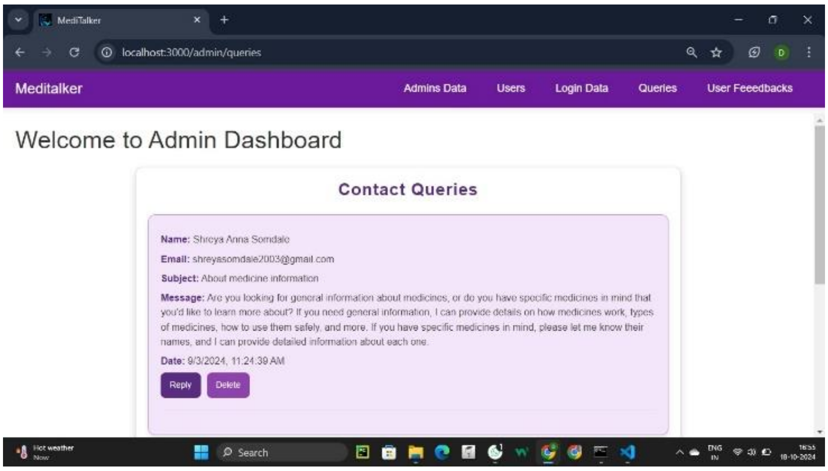


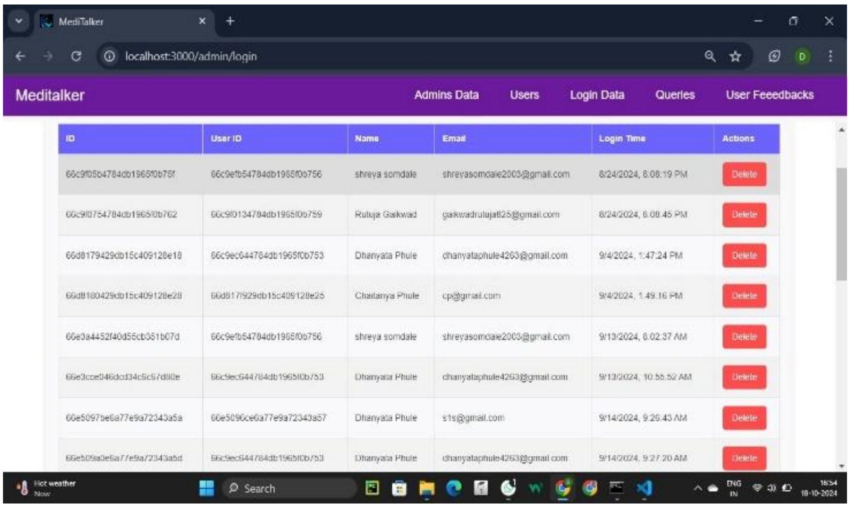


ISSN: 3067-3089| Website: https://csjour.com | Volume-1, Issue-1 | February - 2025









Acknowledgments

The authors thankful to Principal Dr. K. J. Karande, , SKN Sinhgad College of Engineering, Pandharpur for granting permission to carry out the work.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Reference

1. **Title: "Optical Character Recognition: A Survey and Techniques"**
   Authors: Noman Islam, Zeeshan Islam, Nazia Noor
   Journal: Journal of Information & Communication Technology-JICT Vol
2. **MT3S: Mobile Turkish Scene Text-to-Speech System for the Visually Impaired**
   Authors: Mehmet Yilmaz, Yusuf Sinan Akgul
   Journal: Available on arXiv (preprint server for scientific papers)
   Link: arXiv:1608.05054
3. **Machine Learning in Healthcare**:
   Esteva, A., et al. "A guide to deep learning in healthcare." Nature Medicine 25, no. 1 (2019): 24-29.
   Rajpurkar, P., et al. "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." arXiv preprint arXiv:1711.05225 (2017).
4. **Diagnosis and Predictive Algorithms**:
   Chen, T., & Guestrin, C. "XGBoost: A scalable tree boosting system." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
5. Natural Language Processing for Smart Healthcare October 2021
   Authors: Binggui Zhou, Guanghua Yang, Zheng Shi, Shaodan Ma
6. MT3S: Mobile Turkish Scene Text-to-Speech System for the Visually Impaired
   Authors:Muhammet Ba¸stan, Hilal Kandemir, and B¨u¸sra Cant¨urk
7. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends A
   Authors: lexy Bhowmick1,2 · Shyamanta M. Hazarika1
8. A Survey of Deep Learning Approaches for OCR and Document Understanding
   Authors: Nishant Subramani ,Alexandre Matton Scale AI,Malcolm