A PRELIMINARY REPORT
ON

Image Captioning Application Using Deep Learning Algorithms With Audio Description For Visually Impaired

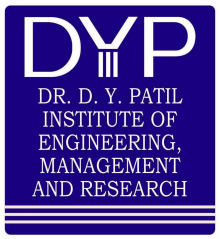
SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING (ARTIFICIAL INTELLIGENCE AND DATA SCIENCE)

SUBMITTED BY

STUDENT NAME	EXAM SEAT NO:
Payal Malviya	**BEAD21220**
Anushka Pote	**BEAD21222**
Akanksha Pawar	**BEAD21235**



DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
DR. D. Y. PATIL INSTITUTE OF ENGINEERING, MANAGEMENT & RESEARCH
AKURDI, PUNE- 411044

SAVITRIBAI PHULE PUNE UNIVERSITY
2024 -2025

DR. D. Y. PATIL INSTITUTE OF ENGINEERING, MANAGEMENT AND RESEARCH

CERTIFICATE

This is to certify that the project report entitles

"Image Captioning Application Using Deep Learning Algorithms With Audio Description For Visually Impaired"

Submitted by

STUDENT NAME	EXAM SEAT NO
Payal Malviya	**BEAD21220**
Anushka Pote	**BEAD21222**
Akanksha Pawar	**BEAD21235**

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Mrs. Deepali Hajare** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Artificial Intelligence and Data Science).

Mrs. Deepali Hajare
Guide
Department of AI&DS

Dr Suvarna Patil
Head
Department of AI&DS

Dr. Anupama V. Patil
Principal,
Dr. D. Y. Patil Institute of Engineering, Management & Research, Akurdi, Pune – 411044

Place : Pune
Date :

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on 'Blockchain-Enhanced Zero-Trust for Metaverse Security'.

I would like to take this opportunity to thank my internal guide **Mrs. Deepali Hajare** for giving me all the help and guidance I needed. I am really grateful to them for their kind support. Their valuable suggestions were very helpful.

I am also grateful to **Dr Suvarna Patil**, Head of AI&DS Department, Dr. D. Y. Patil Institute of Engineering, Management & Research for her indispensable support, suggestions.

In the end our special thanks to **Mrs. Sneha Kanawade** for providing various resources such as laboratory with all needed software platforms, continuous guidance, for Our Project.

Name of Student Sign

Payal Malviya
Anushka Pote
Akanksha Pawar

(BE Artificial Intelligence and Data Science)

ABSTRACT

In recent years, technological advances have opened up new ways to help visually impaired people see their surroundings and access digital content. This review explores development in effectiveness of various technological aids designed for such individuals. Focusing on image-to-sound conversion, community engagement, and immediate environmental feedback. Modern tools such as IoT sticks and Braille displays are always crucial, but they are often restricted by only offering superficial solutions to handle problems. Emergence of deep learning models, especially in image captioning and audio description, aims to provide more immersive yet content-rich experience. Convolutional neural networks (CNN) and short-term (LSTM) networks are used to improvise accuracy of image captioning, while text-to-speech (TTS) system provides useful suggestions. This review explains current state of vision-free applications, identifies their strengths and limitations. It highlights ongoing challenges in implementing innovative technology in a way that is cheaper, highly accurate and easily accessible to users.

Keywords:

Image Captioning, Deep Learning Algorithms, Visually Impaired Assistance, Audio Description, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Text-to-Speech (TTS), Environmental Feedback, Accessibility Technology, IoT-based Assistive Devices

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABBREVIATION	ILLUSTRATION
HTTP	HyperText Transfer Protocol
CSS	Cascading Stylesheets
HTML	Hypertext Markup Language
CNN	Convolutional Neural Network
LSTM	Long Short Term Memory
TTS	Text-to-Speech
API	Application Programming Interface
AI	Artificial Intelligence
GPU	Graphics Processing Unit

LIST OF FIGURES

01. INTRODUCTION

1.1 OVERVIEW

Focusing on image-to-sound conversion, community engagement, and immediate environmental feedback. Modern tools such as IoT sticks and Braille displays are always crucial, but they are often restricted by only offering superficial solutions to handle problems. Emergence of deep learning models, especially in image captioning and audio description, aims to provide a more immersive yet content-rich experience. Convolutional neural networks (CNN) and short-term (LSTM) networks are used to improvise accuracy of image captioning, while text-to-speech (TTS) system provide useful suggestions. This review explains the current state of vision-free applications, identifies their strengths and limitations. It highlights ongoing challenges in implementing innovative technology in a way that is cheaper, highly accurate and easily accessible to users.

1.2 MOTIVATION

Affordable AI-driven audio descriptions provide an innovative and cost-effective alternative for visually impaired individuals, enhancing their understanding and engagement with their surroundings. By leveraging advanced technology, these solutions deliver real-time audio feedback that not only improves access to information but also fosters inclusivity and social interaction. With a focus on affordability, these tools democratize access to essential resources, ensuring that everyone can participate fully in various experiences, whether in entertainment, education, or daily activities. This approach empowers users, allowing them to navigate their environments with greater confidence and autonomy.

Moreover, these solutions promote social interaction and connectivity, breaking down barriers that often isolate individuals with visual impairments. With easily accessible audio descriptions, users can participate in conversations, share experiences, and enjoy activities alongside their sighted peers.

1.3 PROBLEM STATEMENT AND OBJECTIVE

Problem Statement:

The primary challenge faced by visually impaired individuals lies in the limited effectiveness and accessibility of existing assistive technologies. Solutions like obstacle-avoiding IoT sticks and basic Braille displays are often expensive, offer only basic support, and focus primarily on navigation,

leaving users with insufficient means to engage meaningfully with their surroundings or digital content. This highlights the need for more advanced, affordable, and user-friendly solutions that go beyond navigation to enhance independence and improve quality of life. This study addresses these gaps by exploring deep learning models for image captioning and audio description, aiming to create immersive and practical tools for visually impaired individuals.

Objectives:

- Implement real-time audio notifications and voice commands for environmental updates.
- Utilize natural language processing to generate detailed and cost-effective audio descriptions.
- Integrate social media features for sharing experiences and fostering community engagement.

1.4 PROJECT SCOPE

Project Scope:

The project scope involves an in-depth investigation and development of cutting-edge assistive technologies powered by artificial intelligence. These technologies aim to significantly improve environmental awareness and accessibility for individuals with visual impairments. Specifically, the focus will be on creating systems that provide detailed image captioning and comprehensive audio descriptions. By harnessing the capabilities of advanced AI, the project seeks to empower visually impaired users, enabling them to better navigate their surroundings and engage more fully with visual content, ultimately enhancing their overall quality of life and independence.

02. LITERATURE SURVEY

Sr.No.	Paper Title	Journal Name	Authors & Publication Date	Methodology
1	Android based application for visually impaired using deep learning approach	IAES International Journal of Artificial Intelligence (IJ-AI)	Kunal pahwa, Neha Agarwal 16 Feb 2019	Android app with DL, TensorFlow Object Detection API CNN model (SSD, MobileNet V2), Audio output for user notification
2.	EyeRis: Visual Image Recognition using Machine Learning for the Visually-Impaired	IEEE Access	S. Naveen Balaji, P. Victer Paul, R. Saravanan . 2018	TensorFlow Lite, CNN, Programming: Dart, Flutter, Process: Select, preprocess, train, classify, Model: Pre-trained TensorFlow v1
3	Object Detection and Localization for Visually Impaired People using CNN	International Research Journal of Engineering and Technology (IRJET)	Gozde Sismanoglu, Mehmet Ali Onde, Furkan Kocer, Ozgur Koray Sahingoz. 2018	Auto-assistance system with CNNs for object detection, Camera module for image acquisition, Text-to-speech using pyttsx3, Custom CNN trained with indoor dataset, Integrated camera, microcontroller, and output devices.
4	Voice Guided Object Detection: Enabling Independence for the Visually Impaired	2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)	Penglei Gao , Rui Zhang , Xi Yang. 18 August 2020	Utilizes ESP32 CAM Module to capture images of objects, employs OCR (Optical Character Recognition) and TTS (Text-to-Speech) technology.

5	Real Time Object Detection with Speech Recognition using Tensorflow Lite	IEEE Access	Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, Victor Chang. 2018	Deep Learning, Convolutional Neural Network, Object Detection, Object Classification, Tensoflow Lite
6	Object Detection with voice output for visually impaired	2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)	Shom Prasad Das, Sudarsan Padhy. March 2017	Computer vision, machine learning, Torchscript, Voice output, YOLOv5.
7	Image Recognition Technology Based on Machine Learning	IEEE Access	Lijuan Liu1, Yanping Wang, and Wanle Chi3	Machine learning, Image recognition
8	An effective obstacle detection system using deep learning advantages to aid blind and visually impaired navigation	Ain Shams Engineering Journal 15 (2024) 102387	Özgür İcan, Taha Buğra Çelik. October 15, 2017	Develop a deep learning-based obstacle detection system for blind and visually impaired navigation.
9	Image Captioning using Deep Learning Model for Visually Impaired People	International Journal for Multidisciplinary Research (IJFMR)	Mujibur Rahman Majumder, Imran Hossain, Mohammad Kamrul Hasan. 9 February, 2019	Apply an encoder-decoder model with supervised and unsupervised learning for generating image captions, combining computer vision and NLP.
10	Real-time Object Detection and Voice Labeling for Enhanced Accessibility and Visual Interaction	Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)	K. R. Madhavi et al.	Use deep learning for object detection by enclosing objects in bounding boxes and addressing variations in object counts within images. Used YOLO version 7.

03. SOFTWARE REQUIREMENT SPECIFICATION

3.1 INTRODUCTION

3.1.1 PROJECT SCOPE

- Evaluate strengths and limitations of existing assistive technologies.
- Develop deep learning models for image captioning and audio descriptions.
- Create user-friendly interfaces for easy interaction.
- Ensure cost-effective and accessible solutions.

3.1.2 USE CLASSES AND CHARACTERISTICS

Our system is divided into two class/modules:

1) user

2) system

3.1.3 ASSUMPTIONS AND DEPENDENCIES

1. User must have knowledge of Android based applications.
2. User must have knowledge of English.
3. User must have all required software to run the application.

3.1.4 MATHEMATICAL MODELING

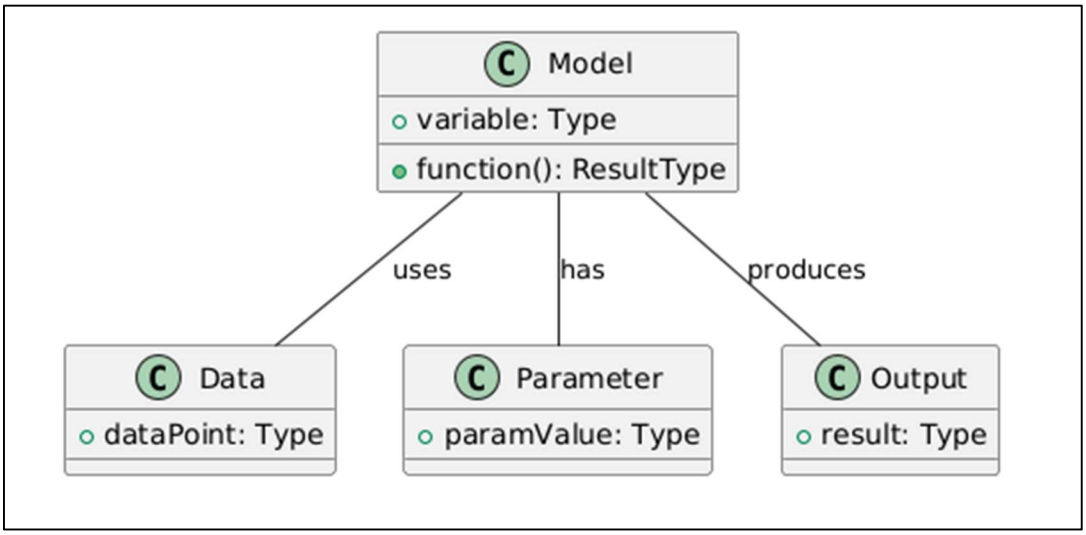


Figure 1 : Mathematical Modelling Diagram

1. System Definition: Let SSS be the system that processes input images.

$$S = \{In, P, Op\}$$

2. Input Identification: Identify the input

$$In \text{ as: } In = \{Q\}$$

Where:

Q = User-entered input image (dataset)

4. **Process Identification:** Identify the process PPP as:

$$P = \{CB, C, PR\}$$

Where:

- CB = Preprocess the data
- C= Apply deep learning algorithm (e.g., LSTM)
- PR = Preprocess request evaluation

5. **Output Identification:** Identify the output Op as:

$$: Op = \{UB\}$$

Where:

- UB= Predict outcome, including success and failure conditions

5. Success and Failure Conditions:

Failures:

1. **Time Consumption**: A large database may lead to significant delays in retrieving information.
2. **Hardware Failure**: Potential breakdowns in hardware can disrupt processing.
3. **Software Failure**: Bugs or issues in the software can prevent correct processing.

Success:

1. **Efficient Data Retrieval**: The system successfully searches and retrieves the required information from the available datasets.
2. **Rapid Results**: Users receive results quickly, aligned with their needs.

6. Complexity Analysis:

- **Space Complexity**: The space complexity is influenced by the presentation and visualization of the discovered patterns. As the data storage increases, the space complexity rises proportionally. If DDD is the dataset size, space complexity can be represented as O(D).

- **Time Complexity**: Let n be the number of patterns available in the datasets.
    - If n>1n > 1n>1, retrieving information can become time-consuming. Thus, the time complexity of this algorithm can be represented as: O(n)

  This indicates that as the number of patterns increases, the time taken for retrieval will also increase linearly.

3.2 FUNCTIONAL REQUIREMENTS

- **User Image Upload**: The system shall enable users to easily upload images for the purpose of generating captions. This feature will facilitate seamless interaction, allowing users to engage with the technology without barriers.
- **Speech Conversion of Captions**: The system shall have the functionality to convert the generated captions into audible speech, providing users with audio feedback. This feature enhances accessibility, allowing visually impaired individuals to comprehend the content of images through auditory means.
- **Cost-Effectiveness and Accessibility**: The system shall prioritize cost-effectiveness to ensure that the solution is affordable for all users. Additionally, it shall focus on accessibility, aiming to provide equitable access to assistive technologies for individuals from diverse backgrounds and with varying levels of ability.
- **Accurate Caption Generation**: The system shall utilize advanced deep learning models to produce precise and contextually relevant captions for the uploaded images. This capability ensures that users receive informative descriptions that accurately reflect the visual content.
- **User-friendly interface:** The system shall be designed with a user-friendly interface that promotes intuitive navigation and interaction. This design consideration ensures that users of all skill levels can easily access and utilize the system's features without frustration.

3.3 EXTERNAL INTERFACE REQUIREMENTS

1. USER INTERFACES

The user interface will feature a simple and intuitive design tailored for visually impaired users, enabling easy image uploads and accessible navigation. It will include audio feedback for actions and results, ensuring smooth interaction. The interface will be optimized for screen readers, allowing users to effortlessly access previous captions and audio descriptions. Additionally, customizable settings will be available, such as adjusting voice speed, to cater to individual preferences and enhance the overall user experience.

2. HARDWARE INTERFACES

The following hardware requirements are essential to ensure the effective operation of the online-based framework for security management:

- **Processor**: A minimum of Pentium IV 2.4 GHz or equivalent to handle the processing demands of image analysis and captioning.
- **Processor Speed**: At least 1.5 GHz or higher to ensure efficient performance and responsiveness of the system.
- **Camera Type**: High-resolution digital camera or webcam suitable for capturing clear images in various lighting conditions**.**
- **Resolution**: Minimum resolution of 1080p (1920 x 1080 pixels) to ensure high-quality image inputs for effective captioning.
- **RAM**: A minimum of 8 GB is required to support multitasking and manage the memory needs of various applications running simultaneously.
- **Hard Disk Space**: At least 220 GB of storage capacity to accommodate the operating system, application software, and stored images and data for processing.
- **Display**: A monitor with a resolution of at least 1920 x 1080 pixels (Full HD) to provide clear visuals for user interaction and content review.
- **Network Connection**: A stable and high-speed internet connection (preferably broadband) to enable seamless interaction with the online framework and support real-time data processing and uploads.

3. SOFTWARE INTERFACES

- **Operating System**: The project is designed to operate on the Windows platform, ensuring compatibility with a wide range of hardware and providing a user-friendly environment.
- **Front End Technologies**: The user interface is built using a combination of HTML, CSS, TypeScript, and JavaScript. HTML provides the structural foundation, CSS is used for styling and layout, while TypeScript and JavaScript handle dynamic interactions and functionality, enhancing user engagement.
- **Development Tool**: Visual Studio Code (VSCode) has been chosen as the primary Integrated Development Environment (IDE) for this project. Its robust features, including syntax highlighting, debugging capabilities, and a rich ecosystem of extensions, facilitate efficient coding and development processes.
- **Database Management**: MySQL serves as the database management system, enabling effective storage, retrieval, and management of data. Its relational structure supports complex queries and ensures data integrity, making it a reliable choice for the project's backend operations.

4. COMMUNICATION-INTERFACES

- **Wi-Fi Connectivity**:

  The application shall support Wi-Fi connectivity, enabling seamless data transfer and access to cloud services. This feature allows users to upload images and retrieve captions effortlessly, ensuring a smooth user experience.
- **Bluetooth-Technology**:

  The application shall utilize Bluetooth technology to facilitate connections with external devices, such as speakers or Braille displays. This capability enhances accessibility, allowing users to receive audio descriptions and tactile feedback wirelessly.
- **Text-to-Speech-APIs**:

  The application shall incorporate text-to-speech (TTS) APIs to convert generated captions into audio descriptions, providing real-time feedback to users. This feature ensures that visually impaired individuals can understand the content of images through auditory means, enriching their interaction with the application

3.4 NON-FUNCTIONAL REQUIREMENTS

1. PERFORMANCE REQUIREMENTS

- High Speed:

The process begins with reading the dataset, followed by data preprocessing, applying the deep learning algorithm, evaluating the request, and finally predicting the outcome.

- Accuracy:

System should correctly execute process, display the result accurately. System output should be in user required format.

2. SAFETY REQUIREMENTS

The safety requirements for the application include ensuring data privacy through secure storage and processing of user data, implementing user authentication to prevent unauthorized access, and designing robust error handling for unexpected inputs. Physical components will be made safe for user interaction, and the application will comply with accessibility standards to ensure usability for individuals with various disabilities, including visual impairments.

3. SECURITY REQUIREMENTS

- **User Authentication**: Implement secure login processes, including multi-factor authentication, to verify user identities.
- **Data Encryption**: Ensure all sensitive user data is encrypted during transmission and storage to protect against unauthorized access.
- **Access Control**: Establish role-based access controls to limit user permissions based on their roles within the application.
- **Regular Security Audits**: Conduct regular security audits and vulnerability assessments to identify and address potential weaknesses in the system.
- **Incident Response Plan**: Develop a clear incident response plan to quickly address and mitigate security breaches or data leaks.

4. SOFTWARE QUALITY ASSURANCE

- Availability [related to Reliability]
- Modifiability [includes scalability, flexibility]
- Performance
- Security
- Testability

3.5 SYSTEM REQUIREMENTS

1. DATABASE REQUIREMENTS

The database requirements for the project include the use of a relational database management system (RDBMS) to store and manage user data, image metadata, and generated captions. The database should support secure connections and allow for efficient querying to retrieve information quickly. It will need to handle concurrent access from multiple users while ensuring data integrity and consistency. Regular backups should be implemented to prevent data loss, and appropriate indexing strategies should be employed to optimize performance and improve response times during data retrieval. Additionally, the database must comply with relevant data protection regulations to safeguard user information.

2. SOFTWARE REQUIREMENTS

DL Algorithm : CNN, LSTM

Coding Language : Flask (Python), ReactJS

Dataset : Flickr 30k

3. HARDWARE REQUIREMENTS

System : Intel I3 Processor and above.

Hard Disk : 200 GB.

Monitor : 15 VGA Color.

Ram : 4 GB.

3.6 ANALYSIS MODELS : SDLC MODEL TO BE APPLIED

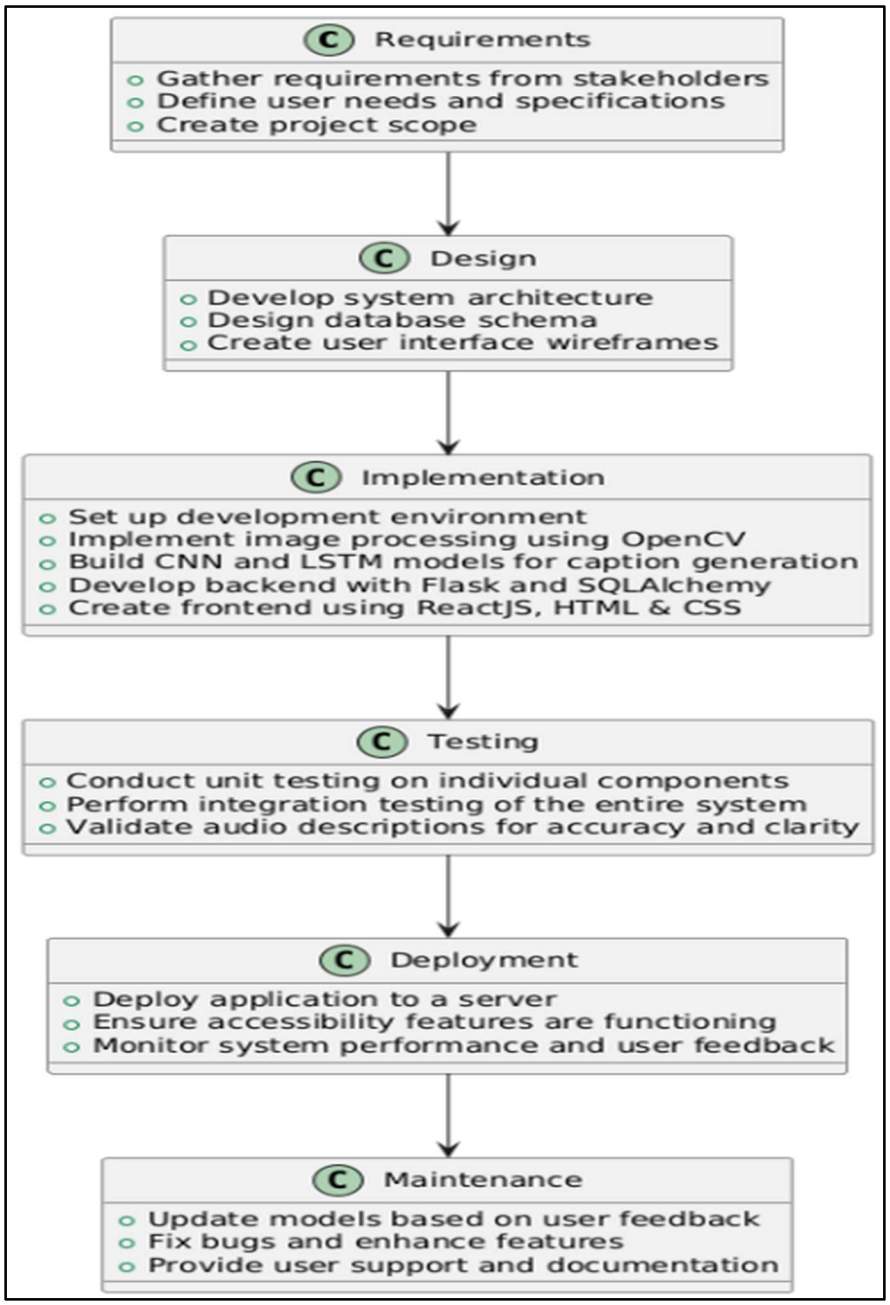


Figure 2: Linear Waterfall Model

3.6 SYSTEM IMPLEMENTION PLAN

1. Requirement gathering and analysis:

In this step, all necessary requirements for the project are identified, including software, hardware, database, and interfaces. A detailed analysis is conducted to ensure the project components align with the needs of the system.

2. System Design:

The system is designed to include a user management layer for identity verification, a secure backend for handling data, and a real-time monitoring layer for performance. The focus is on creating a user-friendly interface and ensuring seamless interaction between the frontend and backend components.

3. Implementation:

The backend is developed using Python and Flask to manage data processing and communication. The frontend is implemented with HTML and CSS to create an accessible user interface. All components are integrated to ensure smooth functioning, and features such as image captioning and audio descriptions are tested for accuracy.

4. Testing:

Various test cases are executed to ensure each module performs as expected. Unit testing validates individual functions, such as image captioning and audio output. User Acceptance Testing (UAT) is performed to ensure the system meets end-user requirements and functions effectively from the user's perspective.

5. Deployment of System:

After successful testing, the system is deployed in the target environment. Any final adjustments are made to ensure it runs smoothly and meets performance expectations.

6. Maintenance:

Maintenance involves regular updates to the software, continuous monitoring for bugs and performance issues, and timely backups of critical data. User feedback is collected to improve system functionality over time, and security patches are applied to ensure ongoing data protection.

04. SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

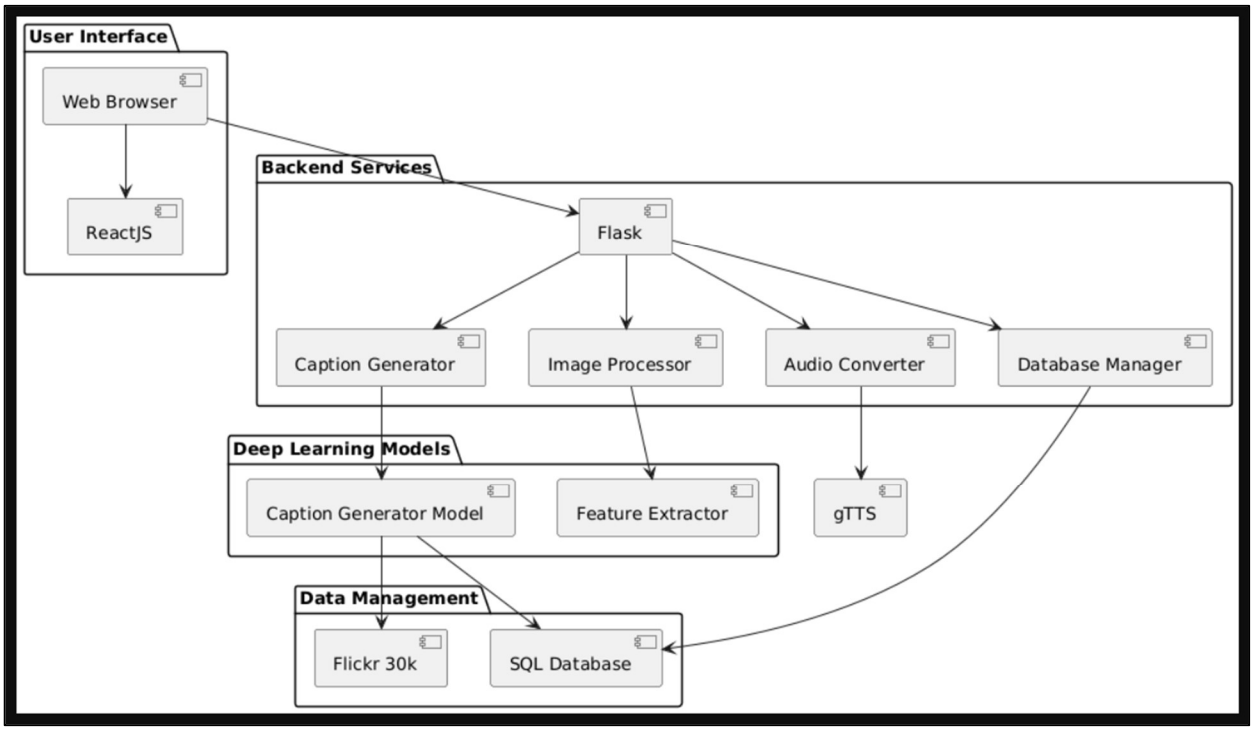


Figure 3 : System Architecture

4.2 DATA FLOW DIAGRAMS

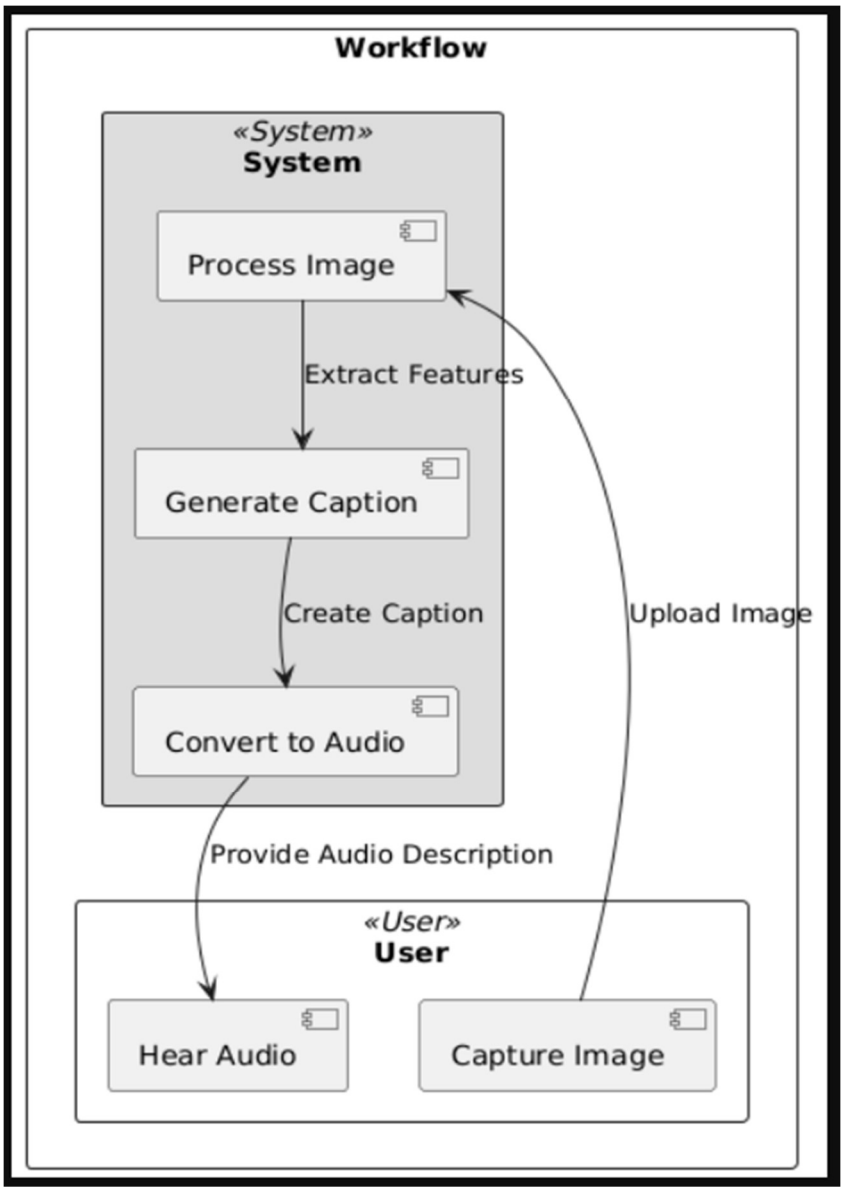


Figure 4.1 : DFD 01

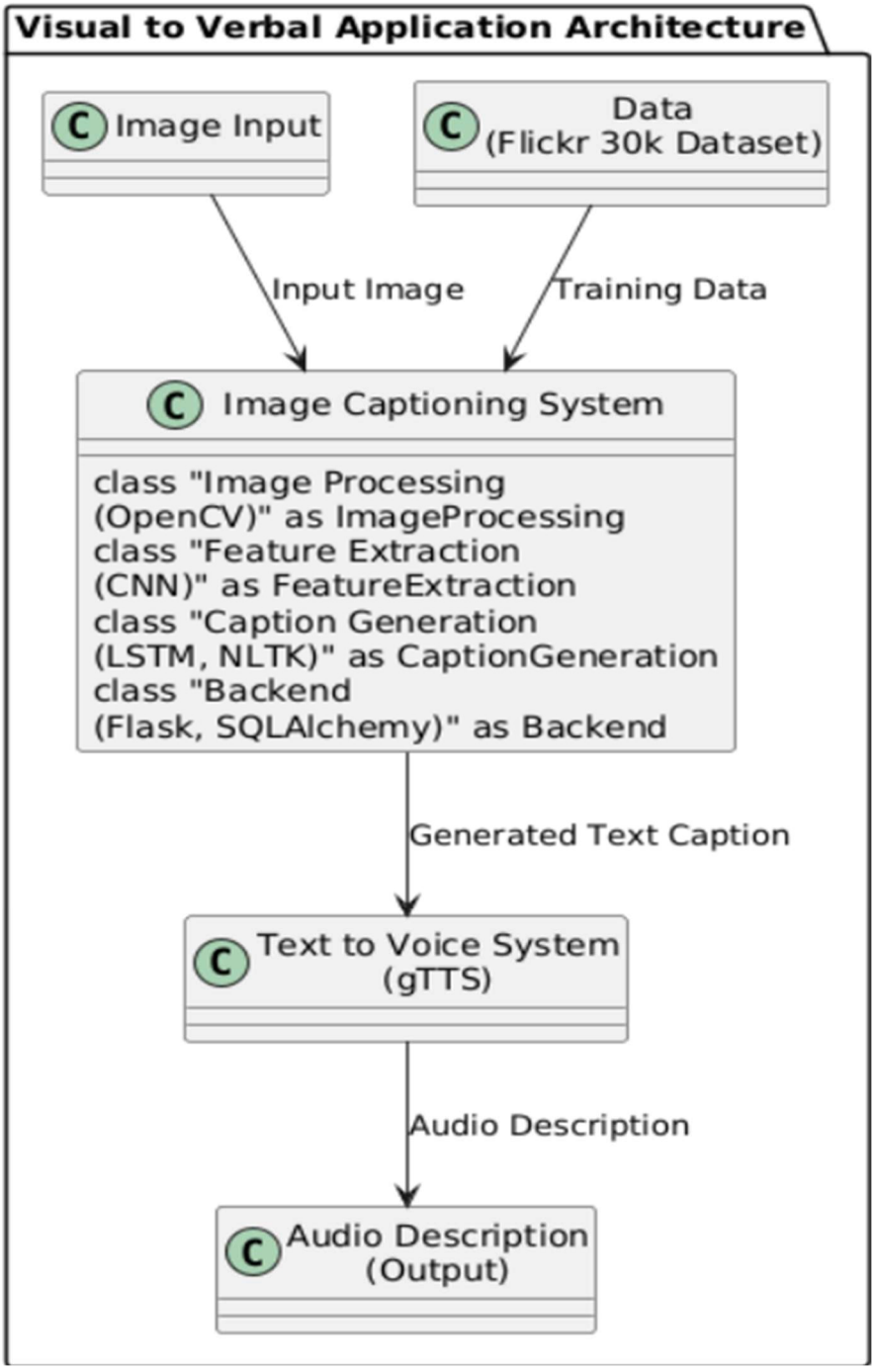


Figure 4.2: DFD 02

4.3 ENTITY RELATIONSHIP DIAGRAMS

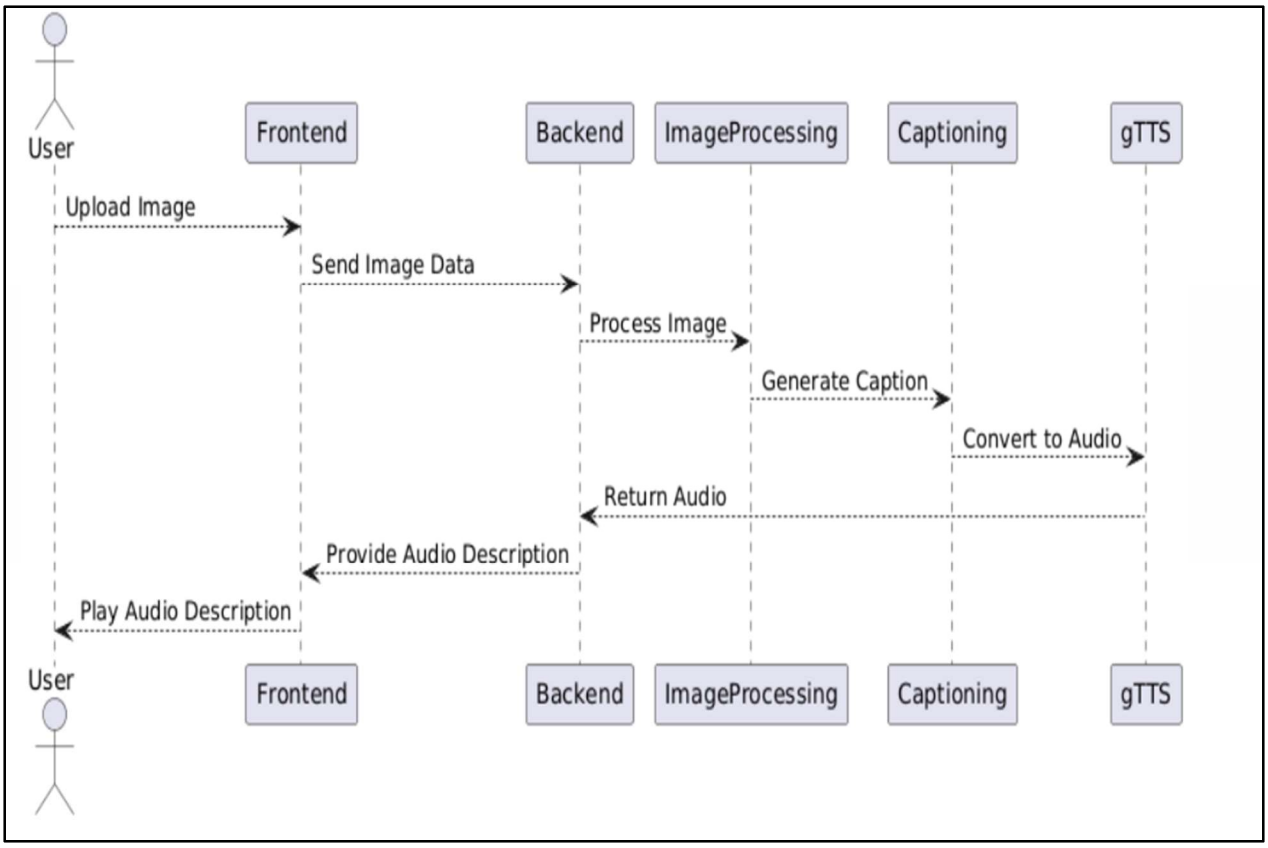


Figure 5 : Entity Relationship Diagram

05. OTHER SPECIFICATION

5.1 ADVANTAGES

- Empowers visually impaired users to understand images through captions and audio.
- Simplifies interaction for users with disabilities.
- Promotes wider access and independence.
- Utilizes deep learning for improved accuracy in image captioning.

5.2 LIMITATIONS

1. Image captioning may take time, affecting real-time usability.
2. Captions may sometimes be incorrect or lack detail.
3. Requires stable internet for certain features, limiting offline use.
4. Performance may vary across devices with low processing power.

06. CONCLUSION & FUTURE WORK

This project is focused on significantly enhancing the independence and quality of life for visually impaired individuals through the provision of accurate image captioning and audio descriptions, all powered by advanced deep learning algorithms. By developing a solution that is both user-friendly and cost-effective, the application seeks to address and bridge the gaps left by traditional assistive technologies, which often fall short in meeting the diverse needs of users.

While there are challenges to overcome—such as processing delays and hardware limitations—the ongoing integration of innovative technologies promises to ensure continuous improvement and increased accessibility for visually impaired individuals. The aim is to create a seamless user experience that allows for real-time feedback and interaction with visual content.

This project not only serves as a foundation for future innovations in assistive technology but also promotes inclusivity and empowerment for users. By enabling them to engage meaningfully with their environment, the application aspires to create a more equitable world where visually impaired individuals can navigate their surroundings with confidence. Ultimately, the initiative is designed to enrich their experiences, facilitating deeper connections with both digital content and the physical world around them.

Future scope:

- **Enhance Processing Speed**: Focus on optimizing algorithms to significantly improve processing speed, enabling instant image-to-audio conversion. This enhancement will facilitate better real-time usability, allowing users to receive immediate auditory feedback for a more interactive experience.

- **Support Multiple Languages**: Expand audio description capabilities to include multiple languages, ensuring that the application is accessible to a diverse, global audience. This inclusivity will empower users from various linguistic backgrounds to benefit from the technology.

- **Integrate with Wearable Devices**: Develop connectivity options for smart glasses or other wearable devices, enabling hands-free interaction. This integration will allow users to access audio descriptions and information seamlessly while on the move, enhancing their independence.

- **Implement Offline Functionality**: Create offline capabilities to ensure usability even without an internet connection. This feature will allow users to access essential functionalities and audio descriptions in areas with limited or no connectivity, increasing reliability.

- **Incorporate User Feedback Mechanisms**: Introduce built-in user feedback mechanisms to gather insights on usability and performance. This will enable continuous improvement and customization of features based on user needs and preferences, fostering a more personalized experience.

- **Enhance Image Recognition Accuracy**: Invest in refining image recognition algorithms to improve accuracy in identifying various objects and scenes. This will ensure that users receive precise and relevant audio descriptions, enhancing their overall experience.

- **Develop Community Engagement Features**: Create features that promote community engagement, such as user forums or sharing platforms where visually impaired individuals can exchange experiences and tips, fostering a sense of belonging and support.

- **Integrate AI for Contextual Understanding**: Utilize artificial intelligence to provide contextual understanding of images, allowing the system to generate richer audio descriptions that convey not just what is present, but also the significance or context of the visual content.

- **Create Tutorials and Support Resources**: Develop comprehensive tutorials and support resources to assist users in navigating the application and maximizing its features. This will enhance user confidence and ensure they can fully utilize the technology.

References:

[1] H. M. Nasir, N. M. A. Brahin, M. M. M. Aminuddin, M. S. Mispan, and M. F. Zulkifli, "Android based application for visually impaired using deep learning approach," IAES International Journal of Artificial Intelligence, vol. 10, no. 4, pp. 879–888, Dec. 2021, doi: 10.11591/ijai.v10.i4.pp879-888.

[2] A. M. Eugenio et al., "EyeRis: Visual Image Recognition using Machine Learning for the Visually-Impaired," in 2023 International Conference on Electronics, Information, and Communication, ICEIC 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi:10.1109/ICEIC57457.2023.10049927.

[3] S. Shah, J. Bandariya, G. Jain, M. Ghevariya, and S. Dastoor, "CNN based auto-assistance system as a boon for directing visually impaired person," in Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019, Institute of Electrical and Electronics Engineers Inc., Apr. 2019, pp. 235–240. doi: 10.1109/ICOEI.2019.8862699.

[4] R. R. Subramanian, L. Ravikiran, K. V. P. Teja, K. V. Reddy, and K. N. Reddy, "Voice Guided Object Detection: Enabling Independence for the Visually Impaired," in 2024 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS 2024 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/INCOS59338.2024.10527768.

[5] G. Khekare and K. Solanki, "REAL TIME OBJECT DETECTION WITH SPEECH RECOGNITION USING TENSORFLOW LITE," 2022.

[6] D. Das and S. Roy, "Object Detection with voice output for visually impaired," in 2024 International Conference on Communication, Computing and Internet of Things, IC3IoT 2024 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/IC3IoT60841.2024.10550247.

[7] L. Liu, Y. Wang, and W. Chi, "Image Recognition Technology Based on Machine Learning," IEEE Access, 2021, doi: 10.1109/ACCESS.2020.3021590.

[8] A. Ben Atitallah, Y. Said, M. A. Ben Atitallah, M. Albekairi, K. Kaaniche, and S. Boubaker, "An effective obstacle detection system using deep learning advantages to aid blind and visually impaired navigation," Ain Shams Engineering Journal, vol. 15, no. 2, Feb. 2024, doi: 10.1016/j.asej.2023.102387.

[9] S. S. Patil and P. J. Patel, "Image Captioning using Deep Learning Model for Visually Impaired People."

[10] M. Swathi, R. Supraja, M. L. Prasanna, S. Sameer, and G. R. K. Reddy, "Real-time Object Detection and Voice Labeling for Enhanced Accessibility and Visual Interaction," 2024, pp. 721–733. doi: 10.2991/978-94-6463-471-6_70.

[11] S. Sharma and K. Guleria, "A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG 16 and Neural Networks," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 357–366. doi: 10.1016/j.procs.2023.01.018.

[12] S. Quazi, "Artificial intelligence and machine learning in precision and genomic medicine," Aug. 01, 2022, *Springer*. doi:10.1007/s12032-022-01711-1.

[13] A. Ben Atitallah, Y. Said, M. A. Ben Atitallah, M. Albekairi, K. Kaaniche, and S. Boubaker, "An effective obstacle detection system using deep learning advantages to aid blind and visually impaired navigation," Ain Shams Engineering Journal, vol. 15, no. 2, Feb. 2024, doi: 10.1016/j.asej.2023.102387.

[14] S. S. Patil and P. J. Patel, "Image Captioning using Deep Learning Model for Visually Impaired People."

[15] M. Swathi, R. Supraja, M. L. Prasanna, S. Sameer, and G. R. K. Reddy, "Real-time Object Detection and Voice Labeling for Enhanced Accessibility and Visual Interaction," 2024, pp. 721–733. doi: 10.2991/978-94-6463-471-6_70.