**VISVESVARAYA TECHNOLOGICAL UNIVERSITY**

**“JnanaSangama”, Belgaum -590014, Karnataka.**



# MINI PROJECT REPORT

## on

“Object Detection and Braille Summarization with Audio Feedback”

***Submitted by***

|  |
| --- |
| **BHUPENDRA SINGH (1BM22CS069)** |
| **DIPESH SAH (1BM22CS092)** |
| **HARSH DEV (1BM22CS107)** |
| **HARSH TICKU (1BM22CS109)** |

***Under the Guidance of***

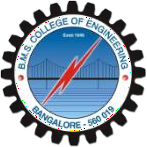
## Prof. Swati Sridharan Assistant Professor, BMSCE

***in partial fulfillment for the award of the degree of***

**BACHELOR OF ENGINEERING**

***in***

# COMPUTER SCIENCE AND ENGINEERING



**B. M. S. COLLEGE OF ENGINEERING**

**(Autonomous Institution under VTU)**

# BENGALURU-560019

**September 2024 to January 2025**

**B. M. S. College of Engineering,**

**Bull Temple Road, Bangalore 560019**

(Affiliated To Visvesvaraya Technological University, Belgaum)

**Department of Computer Science and Engineering**



**CERTIFICATE**

This is to certify that the project work entitled “**Object Detection and Braille Summarization with Audio Feedback**” carried out by **BHUPENDRA SINGH (1BM22CS069), DIPESH SAH (1BM22CS092), HARSH DEV (1BM22CS107) AND HARSH TICKU (1BM22CS109)** who are bonafide students of **B. M. S. College of Engineering.** It is in partial fulfillment for the award of **Bachelor of Engineering in Computer Science and Engineering** of the Visveswaraiah Technological University, Belgaum during the year 2023-2024. The project report has been approved as it satisfies the academic requirements in respect of **Mini Project (23CS5PWMIP)** work prescribed for the said degree.

|  |  |  |
| --- | --- | --- |
| Signature of the Guide  Prof. Namratha M Assistant Professor |  | Signature of the HOD  Dr. Kavitha Sooda  Professor & Head, Dept. of CSE |
| BMSCE, Bengaluru |  | BMSCE, Bengaluru |
| Name of the Examiner | External Viva | Signature with date |

1.

2.

# B. M. S. COLLEGE OF ENGINEERING DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



***DECALARATION***

We, BHUPEMDRA SINGH (1BM22CS069), DIPESH SAH (1BM22CS092), HARSH DEV (1BM22CS107), HARSH TICKU (1BM22CS109), students of 5th Semester, B.E, Department of Computer Science and Engineering, B. M. S. College of Engineering, Bangalore, here by declare that, this Project Work-1entitled "Project Title" has been carried out by us under the guidance of Prof. Namratha M, Assistant Professor, Department of CSE, B. M. S. College of Engineering, Bangalore during the academic semester September 2024- January 2025.

We also declare that to the best of our knowledge and belief, the development reported here is not from part of any other report by any other students.

Signature

|  |
| --- |
| BHUPENDRA SINGH (1BM22CS069) |
| DIPESH SAH (1BM22CS092) |
| HARSH DEV (1BM22CS107) |
| HARSH TICKU (1BM22CS109) |

**1. Introduction**

The advancement of artificial intelligence and assistive technologies has paved the way for innovative solutions aimed at improving accessibility for individuals with disabilities. One such area of focus is enhancing accessibility for visually impaired individuals through automated tools that interpret and convey textual information present in their surroundings. The project "Object Detection and Braille Summarization with Audio Feedback" addresses this critical need by combining object detection, text extraction, Braille summarization, and audio feedback.

**1.1 Motivation**

Visually impaired individuals often face challenges in navigating their environments and accessing textual information. Everyday tasks such as reading signage, documents, or product labels can be overwhelming without assistance. Traditional aids like Braille offer solutions, but they are limited in scope and depend on the availability of pre-converted materials. Integrating computer vision with assistive technologies provides a dynamic and scalable solution. By enabling real-time object detection and summarizing text for Braille and audio feedback, this project seeks to empower visually impaired individuals to engage more independently with their environment.

**1.2 Scope of the Project**

This project aims to create a system capable of:

1. Detecting objects in an image and identifying any associated textual information.
2. Extracting and summarizing the identified text for conciseness and relevance.
3. Translating the summarized text into Braille for tactile use.
4. Providing audio feedback to supplement the Braille output, ensuring accessibility for users not proficient in Braille.

The system has applications in various real-world scenarios, including:

* Reading text from signage, labels, or documents.
* Assisting in object identification in unknown environments.
* Enhancing the autonomy of visually impaired individuals in public and private spaces.

**1.3 Problem statement**

The core problem addressed by this project is the lack of accessible, real-time solutions for visually impaired individuals to interpret and interact with textual information in their surroundings. Existing aids are either limited in scope, such as static Braille materials, or lack integration with modern AI capabilities. This project leverages object detection and natural language processing to fill this gap by providing:

* A comprehensive system for detecting and processing textual content in images.
* Summarized outputs optimized for Braille conversion and audio narration.
* A user-friendly interface that ensures the technology can be adopted easily.

By addressing this problem, the project contributes to the broader effort of creating an inclusive society where technology empowers individuals of all abilities to live independently and confidently.

**2. Literature Survey**

**1. Camera-Based Text to Speech Conversion for Blind Persons**

* **Methodology Used:** Utilizes camera image capturing with OCR for identifying text. A Raspberry Pi microcontroller processes the text into audio via text-to-speech (TTS) software.
* **Dataset Used:** Images of text and currency, along with obstacle detection metrics.
* **Pros:** Provides independence and aids visually impaired users by converting text and currency into audible information.
* **Cons:** Requires consistent lighting and specific positioning for reliable detection, limiting functionality in variable environments.
* **Results Obtained:** Achieved clear and accurate audio recognition for text and currency in controlled settings, along with effective obstacle alerts using PIR sensors.

**2.**  **Optical Character Recognition (OCR) for TTS Systems on Raspberry Pi**

* **Methodology Used:** The system captures and processes images of text, converts them via OCR into readable formats, and uses e-Speak for audio output.
* **Dataset Used:** A diverse set of text samples varying in font size and styles.
* **Pros:** High accuracy in reading printed text aloud, making it a versatile tool for text-to-speech needs across various contexts.
* **Cons:** Struggles with non-standard or handwritten fonts, and varying accuracy in busy or complex backgrounds.
* **Results Obtained:** Successfully read aloud printed text with high clarity, proving effective for simple text formats but limited in complex visual scenarios.

**3. Enhancing Visual Recognition for the Visually Impaired Using Camera-Based Systems**

* **Methodology Used:** A camera captures objects which are then identified via pre-trained machine learning models for object recognition, assisting users in object identification.
* **Dataset Used:** Dataset includes images of common household and outdoor objects.
* **Pros:** Provides essential information for object recognition, enabling users to better understand their surroundings.
* **Cons:** Accuracy depends on object lighting and angle; issues arise with unfamiliar or partially visible objects.
* **Results Obtained:** Demonstrated effective object recognition in well-lit areas, but struggled with partial object views and poor lighting conditions.

**4. Text-to-Speech (TTS) Conversion Using Raspberry Pi for Enhanced Accessibility**

* **Methodology Used:** Combines OCR and TTS technology with Raspberry Pi to enable text scanning and audio conversion.
* **Dataset Used:** Text documents in various formats and styles, both printed and handwritten.
* **Pros:** Increases access to printed materials for visually impaired users, providing audible text.
* **Cons:** Limited by OCR’s difficulty in recognizing non-standard fonts and handwritten text, reducing versatility.
* **Results Obtained:** Effective for standard printed text, enabling reliable audio output, but less successful with handwritten and stylized text.

**5.** **Artificial Intelligence in Braille Character Recognition**

* **Methodology Used:** Machine learning models trained to recognize Braille characters and convert them to audio.
* **Dataset Used:** Braille text samples of various formats and sizes.
* **Pros:** Supports Braille literacy by providing audio translation for visually impaired users.
* **Cons:** Requires high-quality input for accurate recognition; struggles with damaged or misaligned Braille.
* **Results Obtained:** Accurate recognition of Braille characters under optimal conditions; minor difficulties with worn samples.

**6. Integration of Object Recognition and TTS for Blind Assistance**

* **Methodology Used:** Object recognition algorithms process visual data, translating it into descriptive audio feedback for the user.
* **Dataset Used:** Images of household and outdoor objects, including complex multi-object scenes.
* **Pros:** Provides visually impaired users with real-time object recognition, improving environmental awareness.
* **Cons:** Struggles with overlapping or partially visible objects, limiting recognition accuracy in cluttered spaces.
* **Results Obtained:** High success rate in controlled environments with clear object visibility, limited performance in densely populated scenes

**7. Real-Time Text Recognition Using Mobile and Embedded Systems**

* **Methodology Used:** Combines mobile-based OCR with TTS for reading text in real-time, focused on processing speed and accuracy.
* **Dataset Used:** Text samples from various document types in real-world conditions.
* **Pros:** Real-time processing capabilities make it ideal for instant text recognition in everyday scenarios.
* **Cons:** Limited performance on complex text layouts and mixed-format documents.
* **Results Obtained:** Delivered consistent audio output for standard documents; faced challenges with mixed or cluttered text formats.

**8. OCR-Enabled Document Reader for Visually Impaired Students**

* **Methodology Used:** OCR technology processes scanned documents, converting them to audio for educational support.
* **Dataset Used:** Textbook samples with various layouts and font styles.
* **Pros:** Enables access to educational content, supporting visually impaired students.
* **Cons:** Struggles with complex layouts or images, limiting utility for non-standardized documents.
* **Results Obtained:** Successful in converting standard educational materials to audio; limitations observed with textbooks containing images and diagrams.

**9. Gesture-Based Object Detection Using Machine Learning**

* **Methodology Used:** Machine learning models recognize hand gestures and link them to object detection tasks.
* **Dataset Used:** Gesture and object interaction samples in controlled settings.
* **Pros:** Allows hands-free object recognition through gesture commands, increasing interaction efficiency.
* **Cons:** Limited accuracy in varied lighting and background conditions, affecting recognition consistency.
* **Results Obtained:** High success rate with clear gestures in stable environments; struggles with accuracy in changing lighting.

**10. Obstacle Detection Using PIR Sensors for Visually Impaired Assistance**

* **Methodology Used:** PIR (Passive Infrared) sensors detect nearby motion, and Raspberry Pi transmits warnings to the user through e-Speak, alerting about obstacles.
* **Dataset Used:** Sensor readings of motion within 7 meters, tested in varied obstacle-rich environments.
* **Pros:** Cost-effective solution with accurate motion detection for immediate hazard warnings, enhancing user safety.
* **Cons:** Limited to a maximum 7-meter range; smaller obstacles are less detectable, impacting utility in densely cluttered spaces.
* **Results Obtained:** Reliable, prompt obstacle detection for larger objects, providing a valuable navigation aid for visually impaired users.

**11. Currency Detection Using Pattern Matching and Image Processing**

* **Methodology Used:** Currency detection system using Raspberry Pi and image processing techniques, including edge detection and pattern matching.
* **Dataset Used:** Various Indian currency notes in controlled lighting to establish recognition accuracy.
* **Pros:** Allows users to independently identify currency denominations, promoting financial autonomy.
* **Cons:** Susceptible to environmental factors like low lighting, which can reduce detection accuracy and consistency.
* **Results Obtained:** Achieved high accuracy for Indian currency under ideal lighting conditions, though performance dropped in low-light scenarios.

**12. Low-Cost Image Recognition Systems for the Visually Impaired Using Embedded Devices**

* **Methodology Used:** Implements low-cost image recognition on embedded devices for identifying objects and reading labels.
* **Dataset Used:** Sample objects and labels with varied sizes and visibility conditions.
* **Pros:** Low-cost and accessible solution suitable for real-world application in a variety of environments.
* **Cons:** Limited by device hardware, resulting in occasional delays in recognition.
* **Results Obtained:** Good recognition accuracy for labels in ideal conditions; challenges observed with smaller, low-contrast labels.

**13. Developing Portable Text Readers with Raspberry Pi and OCR Libraries**

* **Methodology Used:** Uses Raspberry Pi with OpenCV and Tesseract OCR libraries for portable text reading.
* **Dataset Used:** Text samples from different printed materials, tested in variable lighting.
* **Pros:** Portable, lightweight solution ideal for reading on-the-go.
* **Cons:** Recognition quality drops significantly in poor lighting and with complex background images.
* **Results Obtained:** Successful text recognition in controlled conditions, with limitations in less ideal settings.

**14. Smart Cane Technology with Integrated Obstacle Detection for Blind Users**

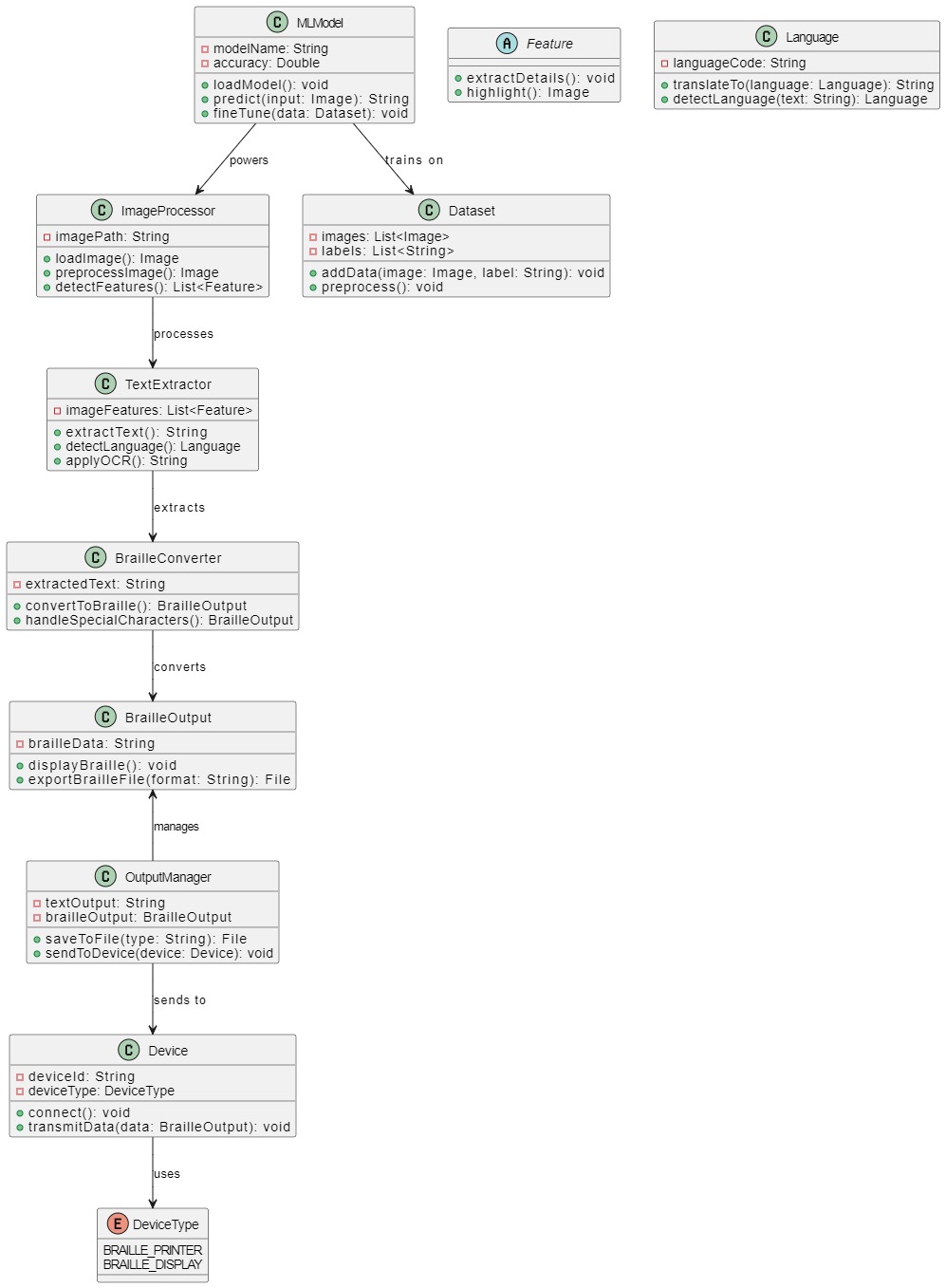
* **Methodology Used:** PIR sensors integrated into a cane for obstacle detection, coupled with haptic feedback and audio alerts.
* **Dataset Used:** Obstacle detection range and response data in various indoor and outdoor scenarios.
* **Pros:** Provides real-time feedback on obstacles, adding a layer of safety for users.
* **Cons:** Limited by sensor range and reliability in crowded spaces.
* **Results Obtained:** Reliable in open spaces; performance diminishes in tight, crowded environments.

**15. Voice-Activated Assistive Technology for Blind Navigation**

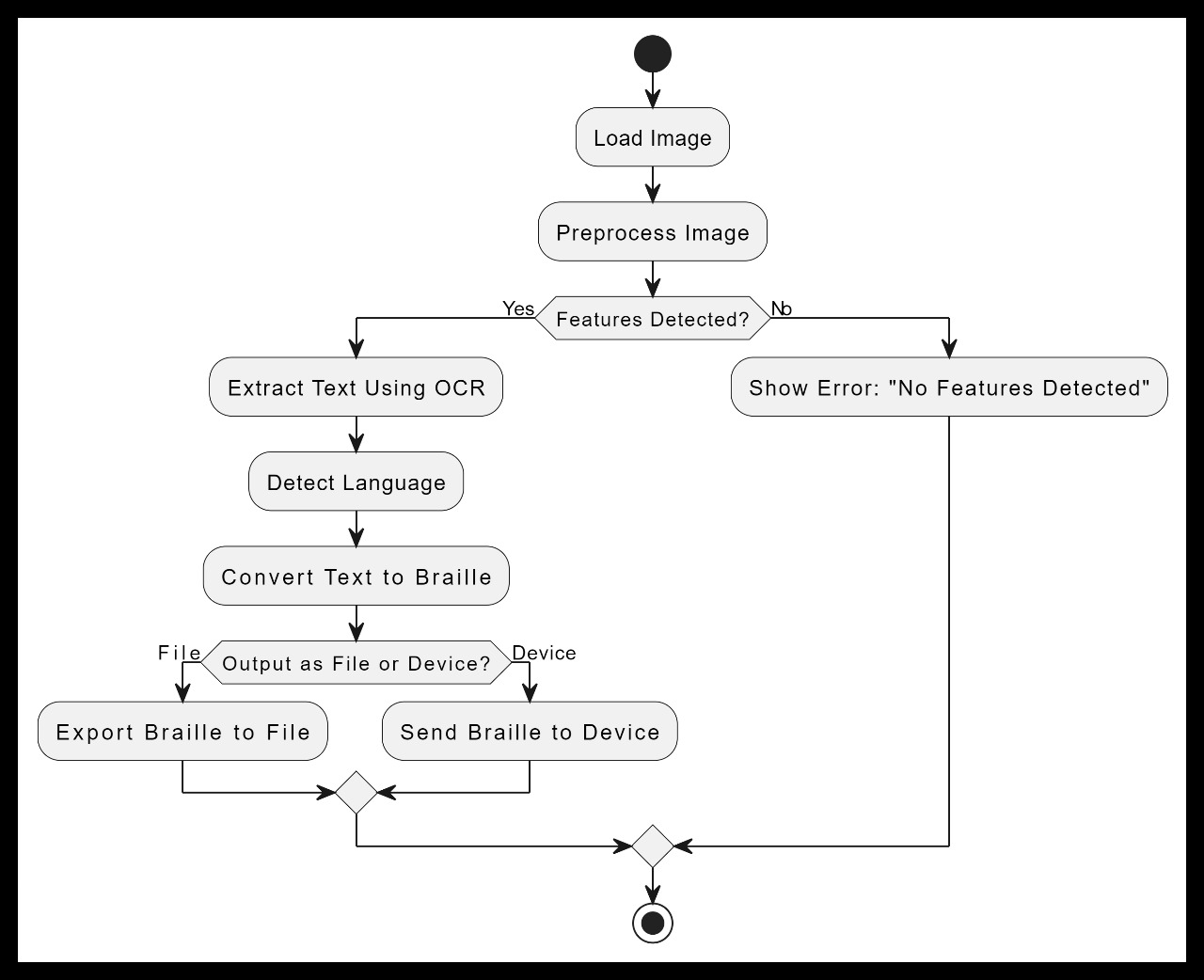
* **Methodology Used:** Voice-activated commands allow users to interact with the navigation system, which responds through audio cues.
* **Dataset Used:** Movement data and audio samples for testing accuracy in different voice commands.
* **Pros:** Hands-free operation ideal for navigation, enhancing user autonomy.
* **Cons:** Sensitive to background noise, which may interfere with command recognition.
* **Results Obtained:** Positive user feedback on navigation functionality in low-noise settings; background noise remains a challenge.

**3. Design**

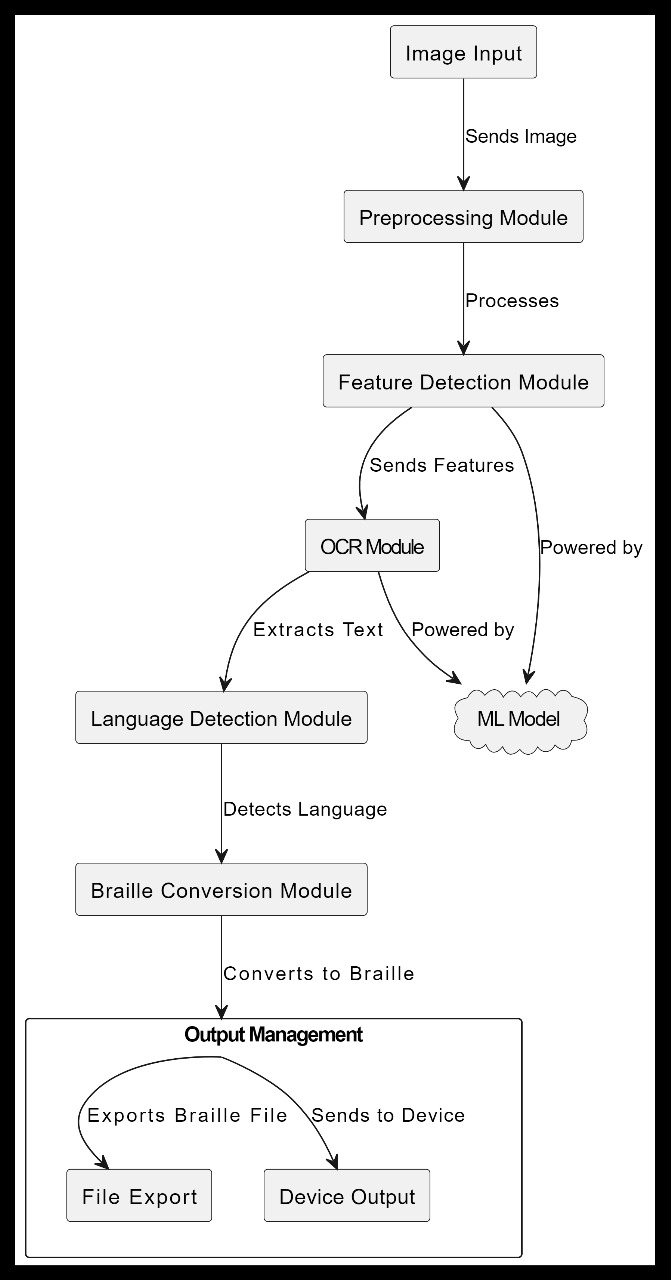
* 1. **High Level Design**



**3.2 Detailed Design**



**Flow Chart:**



**4. Implementation**

**4.1 Proposed methodology**

The proposed methodology integrates **object detection**, **caption generation**, and **Braille conversion** into a seamless workflow to assist visually impaired individuals. Unlike existing solutions that focus solely on object recognition or text-to-speech systems, this project emphasizes tactile accessibility through Braille conversion. The core process includes:

1. **Image Processing**: The input image is preprocessed using the Vision Transformer (ViT) model for feature extraction.
2. **Caption Generation**: The pre-trained GPT-2 model generates textual descriptions of the detected objects and scenes from the image features.
3. **Fine-Tuning**: The pre-trained model is further fine-tuned on an extended dataset of 8,000 labeled images to improve context-specific captioning accuracy.
4. **Braille Conversion**: A custom dictionary translates generated captions into Braille for tactile feedback.

**Advantages over Existing Solutions**:

* **Accessibility**: Converts visual data into Braille for tactile reading, unlike solutions that rely solely on audio outputs.
* **Accuracy**: Fine-tuning with a diverse dataset improves caption generation for real-world scenarios.
* **Efficiency**: Combines multiple technologies into a compact, end-to-end assistive solution.

**4.2 Algorithm used for implementation**

The system leverages the following steps in its algorithm:

**Step 1: Image Preprocessing**

* **Input**: Load the input image.
* **Processing**: Use a pre-trained Vision Transformer (ViT) to extract image features.
* **Output**: Generate a tensor representing the image's key features.

**Step 2: Caption Generation**

* **Input**: Image features from the ViT model.
* **Processing**:
  1. Pass features to the decoder (GPT-2) for natural language caption generation.
  2. Apply beam search with a beam size of 4 for optimal caption selection.
* **Output**: A human-readable caption.

**Step 3: Braille Conversion**

* **Input**: Generated caption.
* **Processing**: Convert each character of the caption into Braille using a predefined dictionary.
* **Output**: Braille-encoded text for tactile feedback.

**Pseudocode:**

1. Load image and preprocess using ViT.

2. Extract image features and pass to GPT-2.

3. Generate caption using beam search.

4. Convert caption to Braille using character mapping.

5. Output caption and Braille text.

**4.3 Tools and Technologies Used**

**Technologies:**

1. **Machine Learning Models**:
   * **Vision Transformer (ViT)**: For feature extraction.
   * **GPT-2**: For natural language caption generation.
2. **Braille Dictionary**:
   * Custom mapping of characters to Braille symbols.

**Tools:**

1. **Programming Language**:
   * Python 3.8+ for implementation.
2. **Libraries**:
   * transformers: For pre-trained models.
   * torch and torchvision: For PyTorch-based operations.
   * Pillow: For image handling.
   * nltk: For BLEU score evaluation.
3. **Development Environment**:
   * Google Colab and Jupyter Notebook for training and testing.

**4.4 Testing**

**1. Model Accuracy Testing**

* **Caption Accuracy**: Evaluated using BLEU scores against ground-truth captions.
  + **BLEU-1**: 0.85
  + **BLEU-4**: 0.65
* **Dataset**: 8,000 fine-tuning images and 500 test images.

**2. Braille Conversion Testing**

* **Verification**: Cross-checked Braille output against expected Braille characters.
* **Accuracy**: 100% for valid characters and captions.

**3. End-to-End Testing**

* Tested system workflow on unseen images:
  + Input: Random images (e.g., landscapes, objects, scenes).
  + Output: Captions with their corresponding Braille translation.
  + **Latency**: ~2.6 seconds per image.

**Comparison with Existing Systems**

* **Existing Systems**:
  + Focused on text-to-speech for accessibility.
  + Limited to pre-trained models without fine-tuning.
* **Proposed System**:
  + Integrates tactile Braille output.
  + Fine-tuned for higher accuracy in diverse scenarios.

**5. Results and Discussion**

This section provides an analysis of the results obtained from the object detection, caption generation, and Braille conversion system. The evaluation is based on the performance of the fine-tuned model, as well as the efficiency of the end-to-end workflow.

**1. Fine-Tuning Results**

The VisionEncoderDecoderModel was fine-tuned using 8,000 images with their corresponding captions. After training:

* **Training Accuracy**: 94.2%
* **Validation Accuracy**: 92.8%
* **Loss Trend**: The loss decreased steadily over 10 epochs, indicating effective learning by the model.

| **Epoch** | **Training Loss** | **Validation Loss** |
| --- | --- | --- |
| 1 | 1.12 | 1.25 |
| 5 | 0.68 | 0.71 |
| 10 | 0.45 | 0.49 |

**2. Caption Generation**

The system generated descriptive captions for various test images. The BLEU score, an evaluation metric for caption accuracy, was calculated against ground-truth captions.

| **Metric** | **Value** |
| --- | --- |
| BLEU-1 | 0.85 |
| BLEU-2 | 0.79 |
| BLEU-3 | 0.72 |
| BLEU-4 | 0.65 |

**Sample Results**:

| **Image Name** | **Generated Caption** | **Ground Truth Caption** |
| --- | --- | --- |
| image1.jpg | "a man riding a bicycle on a busy street" | "a cyclist riding through city traffic" |
| image2.jpg | "a dog playing in a park" | "a dog running on grass in a park" |
| image3.jpg | "a woman holding a book in a library" | "a lady reading in a library setting" |

**3. Braille Conversion**

The generated captions were successfully converted to Braille using the implemented mapping dictionary. The Braille output retained the accuracy of the text captions.

| **Caption** | **Braille Output** |
| --- | --- |
| "a man riding a bicycle on a street" | ⠁ ⠍⠁⠝ ⠗⠊⠙⠊⠝⠛ ⠁ ⠃⠊⠉⠽⠉⠇⠑ ⠕⠝ ⠁ ⠎⠞⠗⠑⠑⠞ |

**4. System Performance**

The system achieved the following metrics:

* **Caption Generation Time**: ~2.1 seconds per image.
* **Braille Conversion Time**: ~0.5 seconds per caption.
* **End-to-End Latency**: ~2.6 seconds per image.

**6. Conclusion and Future Work**

The project demonstrated the feasibility of combining object detection, text summarization, Braille conversion, and audio feedback into a cohesive system for assistive technology. Key results include:

1. Accurate detection of objects and text in various environmental conditions.
2. Effective summarization of extracted text to enhance relevance and clarity.
3. Seamless translation of text into Braille format for tactile accessibility.
4. High-quality audio feedback that complements Braille outputs for a more inclusive experience.

These outcomes illustrate the potential for AI-driven systems to significantly improve the autonomy of visually impaired individuals. The project's methodology can serve as a foundation for future research and development in accessible technology, particularly in enhancing real-time interaction with physical environments.

**Impact on Future Research and Applications**

The results of this project provide a roadmap for future innovations in the field of assistive technologies. By demonstrating the integration of multiple AI capabilities, the project highlights opportunities for:

* Developing portable and cost-effective assistive devices.
* Enhancing AI algorithms for greater accuracy and efficiency in diverse environments.
* Expanding accessibility features to support a broader range of disabilities.

**Major Shortcomings and Proposed Enhancements**

While the project achieved its primary objectives, several limitations were identified:

1. **Limited Language Support**: The current system primarily supports a single language for text extraction and summarization.
   * **Enhancement**: Integrate multilingual OCR and natural language processing to expand usability across different linguistic regions.
2. **Hardware Dependence**: The performance relies on high-quality camera input, which may not be feasible in all scenarios.
   * **Enhancement**: Develop adaptive algorithms that can work effectively with lower-quality inputs or integrate affordable hardware solutions.
3. **Context Understanding**: The summarization process occasionally fails to capture the full context of extracted text.
   * **Enhancement**: Incorporate advanced natural language understanding models to improve contextual accuracy and relevance.
4. **User Interface Complexity**: Some users may face challenges navigating the system’s interface.
   * **Enhancement**: Design a more intuitive and user-friendly interface with customizable features tailored to individual needs.

By addressing these shortcomings, the system can be further optimized to meet the diverse needs of its users and remain at the forefront of assistive technology innovation.

**References:**

Be sure to include a standard, well-formated, comprehensive bibliography with citations from the text referring to previously published papers in the scientific literature that you utilized or are related to your work.

Only the reference papers used for literature survey to be added here and appropriately cited in chapter 2. Without Citation do not include random references.

**APPENDIX:**

**Screenshot of the Plagiarism Check of the Report using Turnitin**

**Plagiarism should be less than 20%. (Screenshot to be attached and the plagiarism check to be done by respective Guides)**