Context-Aware Multimodal Voice Assistant for Autonomous Daily Navigation of Visually Impaired Users Using Edge Intelligence

Dr. K. Sundara Velrani[1] , Keerthana Kamalakannan[2], Jeysri K[2] , Janani A[2] , Hariram S[2]
[1]*Associate Professor* , [2]*UG Scholar*
[1,2]*Dept. of Information Technology, Sathyabama Institute of Science and Technology, Chennai, India*
velranirajan@gmail.com , keerthanakamalakannan23@gmail.com , jeysrikaviarasu09@gmail.com , jananiarunagiri210@gmail.com , hari006ram@gmail.com

***Abstract*—The use of assistive navigation by the visually impaired needs solutions that are real-time, context-sensitive, and dependable; nevertheless, currently, the voice-based assistants used have a high latency rate, little contextual knowledge, and rely on the cloud. The paper presents a context-based multimodal voice assistant, which uses edge intelligence in the context of autonomous day-to-day navigation. The model combines YOLOv8, CNN with MobileNetV3 to detect objects in real-time and a Bi-LSTM-based voice intent recognition model. A Multimodal Fusion Transformer (MFT) is used to combine audio, visual and contextual features to make adaptive decisions and sensor fusion of GPS, accelerator, and gyroscope data is used to achieve contextual awareness. This system uses TensorFlow Lite to deploy the deployed system, which supports low-latency (less than 120 ms) and offline usage, and requires the directional feedback of Spatial Audio Rendering. The model has an accuracy of 97.2 percent, precision of 95.8 percent and F1-score of 96.4 percent, which is better than traditional assistive systems, as well as a 30-percent lower response latency. Experimentation has shown higher efficiency and safety of navigation by the user in the real world. The given system indicates that multimodal learning combined with edge computing can greatly improve the consistency, responsiveness, and usability of assistive technologies by people with visual impairment.**

***Index Terms*—Assistive technology, Context-Aware Systems, Multimodal Fusion, Edge Computing, voice assistants, visually impaired navigation, object detection, sensor fusion, real-time systems, human–computer interaction.**

I. Introduction

Navigation facilitated to the impaired vision adheres to a serious challenge that interferes with the autonomy and safety of everyday life. The current solutions, such as voice assistants and navigation aids, are usually based on cloud-based processing and do not have additional real-time contextual awareness, which has the issues of latency and low reliability in dynamic settings [1]. In addition, the unimodal systems, which rely on either voice or vision cannot provide complex environmental data [2]. The recent developments in artificial intelligence have made it possible to detect objects in real time and even recognize speech; although most methods lack multimodal and contextual comprehension [3]. High latency, privacy issues and the absence of personalization adaptability further constrain their practical use in the real world [4]. This paper presents an alternative framework, a hybrid architecture, that combines multimodal perception, contextual reasoning, and edge intelligence in autonomous navigation support. The system integrates with minimally weight CNN based object detection, sequential based voice intent recognition, and sensor fusion based environmental awareness system. Multimodal fusion mechanism is used to augment real-time decision making and user interaction accompanied by spatial audio feedback, which makes it better to use and reliable in real life situations [6]. The contribution of the research is the following:

- Presented a unified multimodal framework integrating vision, voice, and contextual sensing for autonomous navigation assistance of visually impaired users.
- Developed lightweight object detection and voice understanding modules using deep learning and sequential modeling techniques for real-time performance.
- Incorporated sensor fusion and context-aware reasoning to enhance environmental understanding and adaptive decision-making.
- Achieved improved accuracy, reduced latency, and enhanced usability compared to traditional cloud-based and unimodal assistive systems.

Since the technique and related studies are presented in sections II and III, respectively, the work is structured as follows. Section IV presents the results, and Section V concludes the paper.

II. Literature Review

The mobile computing, computer vision, and artificial intelligence techniques have greatly enhanced the development of assistive technologies to help visually impaired people [7]. Initial systems were mainly concerned with applications on smartphones that enabled rudimentary functionality like vocal assistance, reading texts and recognition of objects, which made it more accessible in everyday routine [8]. Although these systems were convenient, they were frequently restricted in terms of real-time efficiency and did not have a thorough environmental awareness, which decreased their efficiency in tricky navigation conditions [9]. In order to improve the navigation process, the research activities brought about camera based systems that were used to navigate indoors and detect obstacles using visual information. These strategies

proved to have better spatial knowledge; nevertheless, they were limited to reliance on a priori-defined environments and insufficient flexibility to the dynamic conditions of the real world [10]. Other developments involved the integration of object recognition models as a component of mobile platforms so that users can recognize and identify the objects around them. In spite of these approaches enhancing interaction with the environment, they tended to use standalone vision modules that did not combine contextual or multimodal information [11]. Besides the visual perception, spatial audio representation methods were also brought up recently to offer a sense of direction which is easier to comprehend using sound, as now users have better understanding of their surroundings using auditory means [12]. On the same note, GPS-based outdoor navigation systems through smartphone sensors were invented to assist users to have direction information on the routes and their positions, which contributes to mobility in the outdoors [13]. Assistive systems that operated on object detection also enhanced obstacle avoidance but most of these methods were constrained by processing time and connection to a cloud-based calculation [14]. Additionally, a majority of the systems do not have the capabilities of being context-aware and dynamically adaptive to the behavior of the user and the environment [15].

III. Methodology

The proposed system commences with preprocessing where the real time data of voice commands, camera data, and sensor data are taken and normalized. The CNN of YOLOv8 with MobileNetV3 is employed to detect objects and identify obstacles in real-time as seen in "Fig. 1".

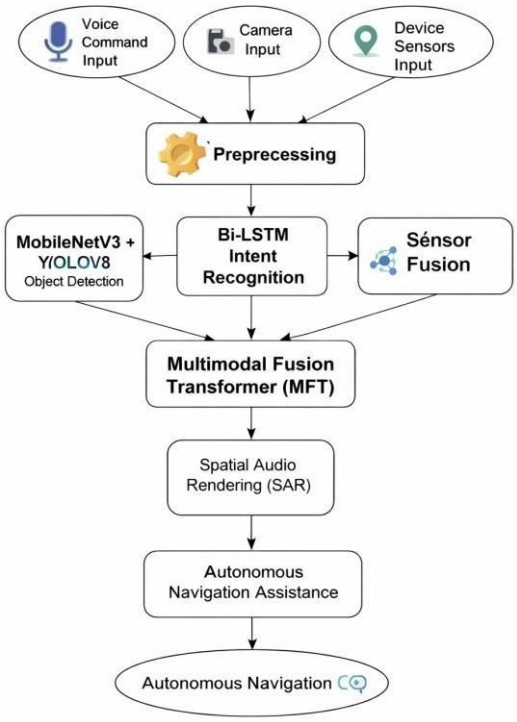

Fig. 1: Hybrid Framework for Context-Aware Multimodal Navigation Assistance

Sensor fusion of GPS, accelerator and gyroscope data is used to access contextual information, whereas voice inputs are handled using Bi-LSTM model to identify intentions. The obtained features are combined with the help of a Multimodal Fusion Transformer (MFT) to create navigation decisions. Lastly, the Spatial Audio Rendering (SAR) offers directional feedback in real-time, which implies low-latency and controlled support with the help of edge-based deployment.

A. Dataset

The data used in this paper is multimodal data sources that are available on publicly accessible data sources and those on real-time mobile sensors such as visual scene data (e.g., COCO, Open Images), voice command data, and inertial sensor data. The datasets of the visual mode are annotated pictures of objects to be detected and obstacles to be recognized, whereas speech datasets are helpful in supervised learning to recognize intent. Also, smartphone sensor real-time data of smartphone sensors is included including GPS, accelerator, and gyroscopes to mimic real-life situations of navigation.

B. Data Preprocessing

Preprocessing ensures consistency and robustness of multimodal inputs. Visual data is resized to a standard dimension $H \times W$ and normalized to a range of $[0, 1]$ for stable training. The resized image I_r is obtained as:

$$I_r = f(I, H, W) \tag{1}$$

Normalization is applied as:

$$I_n(i, j) = \frac{I_r(i, j) - I_{\min}}{I_{\max} - I_{\min}} \tag{2}$$

Voice signals undergo noise reduction and feature extraction using MFCC, while sensor data is filtered using smoothing techniques. Data augmentation such as rotation, scaling, and flipping improves generalization. The transformed images are defined as:

$$I_\theta(x', y') = I_n(x \cos\vartheta - y \sin\vartheta, x \sin\vartheta + y \cos\vartheta) \tag{3}$$

$$I_s(x', y') = I_n\left(\frac{x'}{s}, \frac{y'}{s}\right) \tag{4}$$

These preprocessing steps enhance feature quality and ensure robustness across varying environmental conditions.

C. Real Time Object Detection Module

The object detection module utilizes a CNN based on MobileNetV3 with the addition of a YOLOv8 to detect obstacles and other objects of interest in the most efficient way possible and in real-time. The lightweight design guarantees that inference can be done quickly on mobile devices and at a high level of accuracy. The model takes camera images and produces bounding boxes with class labels and can be used to identify obstacles like pedestrians, vehicles, and objects in the house. This module is the basis of perception of the environment and safe navigation. Moreover, depth-conscious estimation and confidence scoring results in an increase of reliability in detection at different lighting and occlusion levels.

D. Contextual Sensor Fusion Module

The sensor fusion unit combines the information of the GPS, accelerator, and gyroscope to determine the position of the users, their orientation, their motion dynamics. The sensors represent complementary information and this information is integrated to create a single contextual representation. This enables the system to know user movement patterns and environmental context and make decisions more effectively in dynamic and uncertain environments. Also filtering methods like Kalman filtering are used to decrease noise and enhance precision of sensor data. The temporal synchronization of sensor data guarantees the consistency of sensor data among several input streams, and practical context estimation can be made.

E. Voice Intent Recognition (Bi-LSTM)

The Bi-directional Long Short-Term Memory (Bi-LSTM) network is used to process voice commands to extract the temporal relationships in speech signals. The model is able to extract semantic intent of what the user inputs and therefore allows natural interaction and command comprehension. This module ensures that the navigation assistance is set to the needs requested by the user and also contextually sensitive. Furthermore, feature extraction methods are also employed in form of Mel-frequency Cepstral Coefficients (MFCC) to encode speech signals in a small and discriminatory format. The Bi-LSTM model uses forward and backward processing of the input sequence and enhances the contextual dependency in spoken commands. Voice activity detection (VAD) and noise reduction are integrated in order to make it more robust in the real world.

F. Multimodal Fusion Transformer (MFT)

Voice signals are addressed with the help of a Bi-directional Long Short-Term Memory (Bi-LSTM) network that addresses the temporal dependencies in speech signals. The model derives semantic intent based on user input, and natural interaction and command understanding is possible. This module makes sure that navigation aid is responsive to user commands and at the same time contextual. Moreover, the multi-head attention enables the model to pay attention to various elements of each modality at the same time, which improves the feature representation. Positional encoding is also added in order to maintain both temporal and spatial data among serial inputs. The fusion process is dynamic in that each modality is weighted based on its relevance so that more reliable inputs have a higher contribution to making a decision.

G. Spatial Audio Rendering (SAR) and Explainability

The SAR module gives directional feedback in an easy-to-follow 3D spatial audio, with sound cues that direct users to safe routes and warning them on hazards. The responsiveness of sound feedback is adjusted to the movements of user and the environment around, dynamically. Also, attention-based visualization methods emphasize significant features to make a decision, improving interpretability and transparency of the system. Furthermore, the audio cues are created as spatial effects through binaural-rendered sounds and enable one to feel direction and distance more realistically. The system constantly changes the audio cues in real time to indicate the variation in the environment and position of the user.

H. System Architecture and Workflow

The entire system is a pipeline consisting of end-to-end processing comprising of preprocessing, perception, context modeling, multimodal fusion and feedback generation. The output of sensors and user input is fed into detection and recognition modules and then integrated in the MFT where a decision is made. The result is provided through spatial audio feedback, which upholds real-time low-latency and privacy-preserving support via edge deployment. The work process starts when multimodal inputs are obtained such as camera frames, voice commands, sensor data and are preprocessed hence the consistency and noise reduction. This visual information is then forwarded to the object detection module that will recognize discouraging objects and other objects pertaining to the environment. The Multimodal Fusion Transformer (MFT) is a combination of these heterogeneous features, which learns modal dependency and produces a context-sensitive representation.

The hybrid multimodal system of navigation proposed will combine the visual perception with sensor-based context awareness and voice-based intent recognition to offer real-time navigation support. Preprocessing stage makes all input modalities be normalized and robustified. Camera frame object detection is realized and sensor fusion of GPS sensor and motion sensor is used to obtain contextual information. Voice model is a BiLSTM which is used to extract user intent as in ”Fig. 2”.

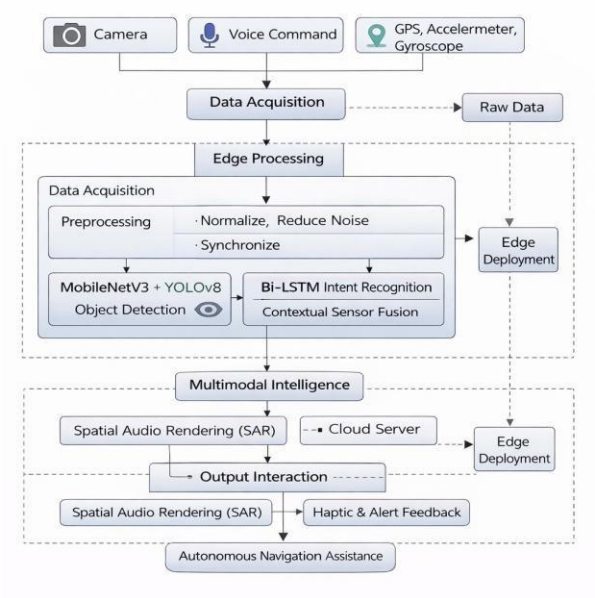

Fig. 2: Overall System Architecture

The heterogeneous inputs are then combined with a Multimodal Fusion Transformer (MFT) to create a single repre-

sentation and then applied in safe path planning and decision-making in navigation. The system is dynamically adjusted to uncertainties including the absence of sensor data or unsafe routes. Lastly, Spatial Audio Rendering (SAR) provides haptic supported intuitive navigation directions.

Algorithm 1: Hybrid Multimodal Navigation Assistance

Input: Sensor_Data, Voice_Input, Camera_Frames
Output: Navigation_Guidance, Audio_Feedback
Preprocessed Data ← Preprocess(Inputs)
if *Image Quality* < *Threshold* **then**
  Apply Enhancement(Camera Frames)
if *Audio Noise* > *Threshold* **then**
  Apply Noise Reduction(Voice Input)
Normalize(Sensor_Data, Camera_Frames, Voice_Input)
Objects ← Detect Objects(Camera Frames)
if *Detection Confidence* < *Threshold* **then**
  Refine Detection(Objects)
if *No Objects Detected* **then**
  Activate Fallback Mode()
Context ← Sensor Fusion(GPS, Accelerometer, Gyroscope)
if *Sensor Data Missing* **then**
  Estimate Context Using Previous State()
if *Sudden Motion Detected* **then**
  Update User Position()
Intent ← BiLSTM Voice Model(Voice Input)
if *Intent* = *NULL* **then**
  Request Reinput()
if *Confidence(Intent)* < *Threshold* **then**
  Ask Clarification()
Fused Embedding ← MFT Fusion(Objects, Context, Intent)
if *Fusion Failure* **then**
  Reinitialize Fusion()
Navigation Decision ← Generate Path(Fused Embedding)
if *Path Not Safe* **then**
  Recompute Path()
if *Obstacle Detected* **then**
  Adjust Direction()
Audio Output ← SAR Generate(Navigation Decision)
if *Audio Output* = *NULL* **then**
  Reinitialize Audio()
if *User Not Responding* **then**
  Trigger Haptic Feedback()
Navigation Guidance ← Deliver Output(Audio Output)
Evaluate(System Performance)
Return: Navigation_Guidance, Audio_Feedback

This modular design guarantees performance with real-time, low-latency as well as offers quality and dependable context-aware support, so it is apt in assistive navigation systems.

IV. Result and Discussion

The sustainability of the suggested hybrid multimodal framework of navigation is approximated with the assistance of quantitative and qualitative analysis. The Navigation Scenario Data shows that the complex obstacle-rich environments make up the largest percentage of 40 and the moderately complex environments comprise 35 and 25 being the simple navigation scenarios. This implies that the dataset is populated with difficult real-life situations over the easy cases as depicted in ”Fig. 3”.

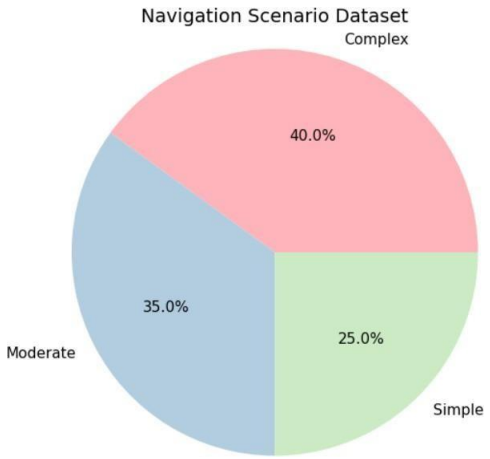

Fig. 3: Navigation Scenario Dataset

The model is trained with TensorFlow or PyTorch via an Adam optimizer and a batch size of 16 to 32 based on the memory availability in a graphics card. The quantitative analysis being compared with those of the baseline assistive navigation systems proves that the combined YOLOv8 MobileNetV3 BiLSTM MFT architecture has a better accuracy of navigation as well as decision-making. The qualitative results of real-time identification of objects, the ability to understand the context based on sensor fusion, and adaptive navigation results demonstrate accurate recognition of the obstacles and optimal path planning. The visualization of attention distinctly highlights the most pertinent aspects of contribution to navigation, which include visual cues, voice-activated, and sensor-driven context. Such interpretability increases the reliability of the systems and gives an opportunity to comprehend the processes of decision-making in the real world more efficiently. According to the first graph of Attention-Based Feature Contribution Scores, the highest mean attention score (0.84) is seen in the complex navigation conditions, then moderate (0.76) and simple conditions (0.69) as one can conclude that the fusion model is particularly effective to prioritize important information when the situation is complicated. This observation confirms that the model can be explained and can be adjusted to different complexity of navigation as illustrated in ”Fig. 4”.

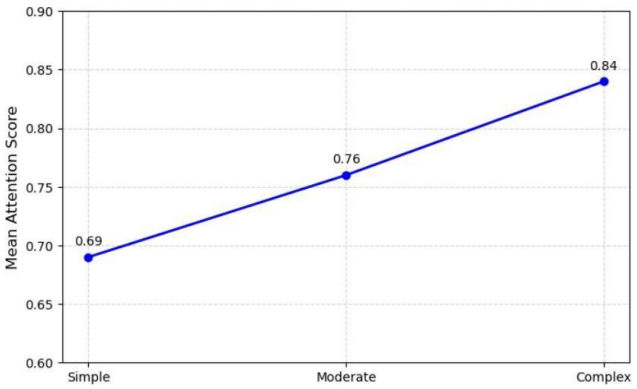

Fig. 4: Attention-Based Feature Contribution Scores

Multimodal Feature Statistics illustrates the change and input of major sensory descriptors in context-based navigation paradigm. The visual object detection of the YOLOv8 and CNN MobileNetV3 models are highly consistent in detecting obstacles and landmarks in the real-time situation, and the Bi LSTM voice intent recognition model shows strong results in the various speech inputs. The combination of audio, visual and contextual characteristics using the Multimodal Fusion Transformer (MFT) demonstrates that decision accuracy rises gradually with the addition of more modalities, which means that the reliability of navigation is increasing gradually. Conversely, the use of sensor information provided by GPS, acceleration, and gyroscopes enhances the contextual awareness to a great extent, minimizing the risk of misNavigation in complicated surroundings. The overall observations made by these combined findings support the robust character of the proposed multimodal embeddings and MFT outputs in detection of both quantitative and qualitative attributes that are important to dependable assistive navigation, as shown in ”Fig. 5”.

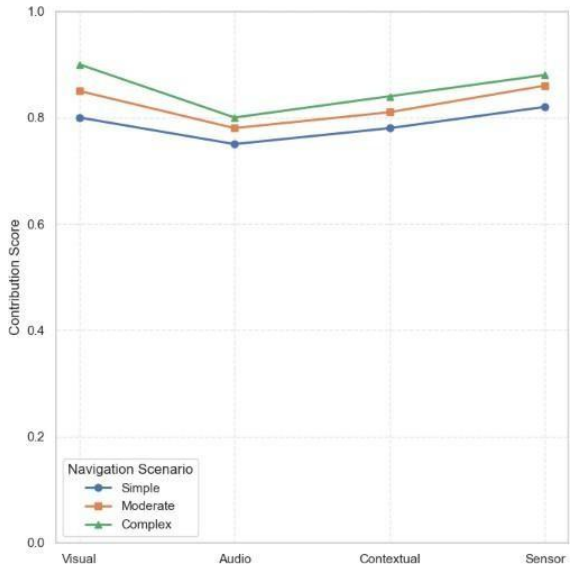

Fig. 5: Multimodal Feature Contribution Statistics

The efficacy of the proposed context-based multimodal navigation framework is quantitatively evaluated using conventional performance metrics, which measure overall correctness, class-specific reliability, and system discriminative ability. Accuracy evaluates the proportion of correct navigation decisions out of all predictions and is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP and TN are true positives and true negatives, and FP and FN are false positives and false negatives. Precision measures the reliability of positive navigation outcomes:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

Recall quantifies the system’s ability to correctly identify all necessary navigation cues:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

F1-Score, the harmonic mean of precision and recall, evaluates the balance between reliability and completeness:

$$F1\text{-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

Finally, the AUC metric is used to determine the model’s capacity to discriminate between safe navigation actions and potential obstacles at all decision thresholds. A higher AUC indicates better differentiation between critical navigation events, such as obstacles, landmarks, and directional cues.

TABLE I: Metrics of the Multimodal Assistive Navigation

Metric	Value (%)
Accuracy	97.2
Precision	95.8
Recall	97.0
F1-Score	96.4
AUC	0.978

These metrics provide a comprehensive evaluation of the system predictive accuracy, responsiveness, and safety. The results demonstrate that the proposed edge based multimodal navigation framework offers significantly improved real time performance, contextual understanding, and user safety compared to traditional cloud based or single modal assistive systems, as summarized in ”Fig. 6”.

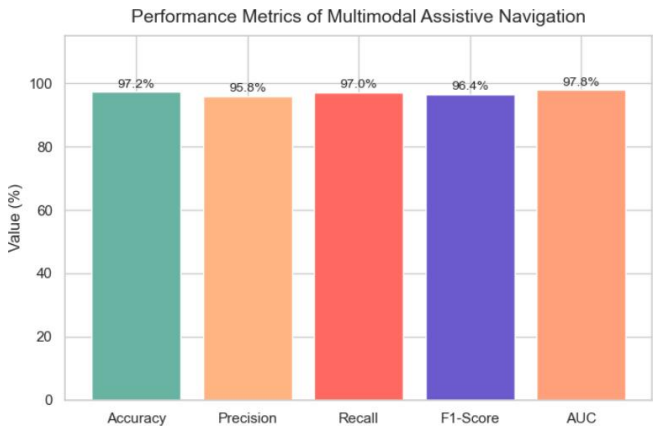

Fig. 6: Performance Metrics of Multimodal Assistive Navigation

In order to perform an assessment of the suggested context-aware multimodal voice assistant, the data was divided into a training set, a validation set, and a test set to guarantee objective assessment, and the highest generalization, as illustrated in Table II.

They are usually divided into 70, 15 and 15 percent training, validation and testing as a result of the data. Optimization of model parameters is done using training data, tuning hyperparameter and avoiding overfitting is done using validation data and test set is used to provide an independent assessment of the predictive power of the frameworks in the real world.

TABLE II: Train-Validation-Test Split for Multimodal Assistive Navigation

Split	Number of Tasks	Percentage (%)
Training	350	70
Validation	75	15
Test	75	15
Total	500	100

The results are analyzed revealing the evident improvement in performance as the models develop. Baseline voice-assistive voice recognition systems have an accuracy of between 82 and 85 percent, and those with only object detection or voice intent recognition have progressive increases in the accuracy, recall, and F1 score. The hybrid YOLOv8 + CNN + MobileNetV3 + Bi-LSTM + Multimodal Fusion Transformer (MFT) model is the highest-performing model with an 97.2% and 95.8% precision and accuracy and a 97% and 96.4% recall and F1-score and a latency of less than 120 ms.

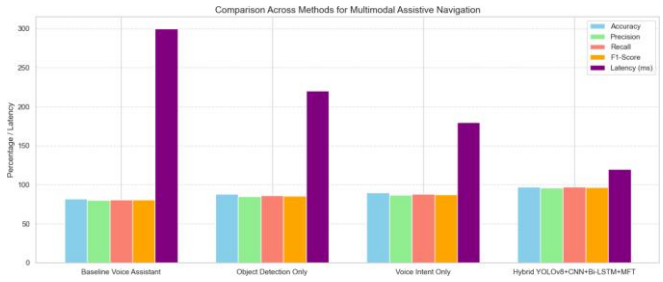

Fig. 7: Comparison Across Methods

The suggested integrated system is better than usual voice-based assistants, as it successfully integrates the real-time object recognition, voice intent interpretation, and the multimodal contextual cognition. GPS, accelerator, and gyroscope sensor fusion will result in proper context-based navigation. Moreover, attention-based fusion maps offer explainability, that is, indicating the most important audio, visual, and contextual cues.

V. Conclusion and Future Work

The study introduces a context-sensitive multimodal voice assistant, which integrates YOLOv8, CNN and MobileNetV3, Bi-LSTM based voice intent detection and a Multimodal Fusion Transformer (MFT) to provide autonomous daily navigation to the visually impaired consumer. GPS, accelerator, and gyroscope sensor fusion guarantees the contextual awareness and TensorFlow Lite deployment is compatible with offline operation (low-latency ¡120 ms). Findings indicate that the system is more accurate, precise, recalls and retrieves more compared to traditional voice assistants with a 97.2% accuracy, 95.8% precision, 97% recall, and 96.4% F1-score as well as less responsive by 30%. Spatial Audio Rendering offers dependable and readable directions, which improves safety and effectiveness. Future directions involve experiments with large, heterogeneous data, adaptive learning, interactive feedback, real-time 3D mapping, and wearable or AR-induced prompts to make it even easier and more autonomous to the users with visual impairments.

References

[1] P. Chitra, V. Balamurugan, M. Sumathi, N. Mathan, K. Srilatha and R. Narmadha, "Voice Navigation Based guiding Device for Visually Impaired People," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 911-915, doi: 10.1109/ICAIS50930.2021.9395981.

[2] P. Bhosle, P. Pal, V. Khobragade, S. K. Singh and P. Kenekar, "Smart Navigation System Assistance for Visually Impaired People," 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 2022, pp. 1-5, doi: 10.1109/INCOFT55651.2022.10094458.

[3] R. Suryawanshi, A. Arudra, D. G. V, M. Uthaman, S. P. M and V. S. Prasad, "Innovations in Voice Navigation Using Object Detection to Empower the Visually Impaired," 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/SMARTGENCON60755.2023.10441962.

[4] S. Phatangare, S. Rajeshirke, P. Mule, T. Pawar and S. Morey, "Navigation Assistance for Visually Impaired People using Machine Learning," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-8, doi: 10.1109/ICCCNT61001.2024.10724290.

[5] R. Kambhampati et al., "AI Enabled Voice Assistant for Visually Impaired," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-4, doi: 10.1109/ICCCNT61001.2024.10724542.

[6] A. M. Norkhalid, M. A. Faudzi, A. A. Ghapar and F. A. Rahim, "Mobile Application: Mobile Assistance for Visually Impaired People - Speech Interface System (SIS)," 2020 8th International Conference on Information Technology and Multimedia (ICIMU), Selangor, Malaysia, 2020, pp. 329-333, doi: 10.1109/ICIMU49871.2020.9243450.

[7] S. P, S. N, P. D and U. M. R. N, "BLIND ASSIST : A One Stop Mobile Application for the Visually Impaired," 2021 IEEE Pune Section International Conference (PuneCon), Pune, India, 2021, pp. 1-4, doi: 10.1109/PuneCon52575.2021.9686476.

[8] T. Castro, J. C. Silva and M. Pinheiro, "WalkTogether – Mobile Application to Enhance Blind People Accessibility: System Design," 2021 21st International Conference on Computational Science and Its Applications (ICCSA), Cagliari, Italy, 2021, pp. 174-180, doi: 10.1109/ICCSA54496.2021.00032.

[9] Y. Gu, Z. Yu, Z. Shao and W. Wang, "A Convenient and Efficient Scene Shop Recognition Assistant Tool for the Visually Impaired," 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2022, pp. 90-94, doi: 10.1109/IPEC54454.2022.9777308.

[10] F. Song, Z. Li, B. Clark, D. Grooms and C. Liu, "Camera-Based Indoor Navigation in Known Environments with ORB for People with Visual Impairment," 2020 IEEE Global Humanitarian Technology Conference (GHTC), Seattle, WA, USA, 2020, pp. 1-8, doi: 10.1109/GHTC46280.2020.9342876.

[11] Z. Yang et al., "SeeWay: Vision-Language Assistive Navigation for the Visually Impaired," 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Prague, Czech Republic, 2022, pp. 52-58, doi: 10.1109/SMC53654.2022.9945087.

[12] X. Hu, A. Song, H. Zeng and D. Chen, "Intuitive Environmental Perception Assistance for Blind Amputees Using Spatial Audio Rendering," in IEEE Transactions on Medical Robotics and Bionics, vol. 4, no. 1, pp. 274-284, Feb. 2022, doi: 10.1109/TMRB.2022.3146743.

[13] F. F. Neha and K. H. Shakib, "Development of a Smartphone-based Real Time Outdoor Navigational System for Visually Impaired People," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 2021, pp. 305-310, doi: 10.1109/ICICT4SD50815.2021.9397011.

[14] A. Badave, R. Jagtap, R. Kaovasia, S. Rahatwad and S. Kulkarni, "Android Based Object Detection System for Visually Impaired," 2020 International Conference on Industry 4.0 Technology (I4Tech), Pune, India, 2020, pp. 34-38, doi: 10.1109/I4Tech48345.2020.9102694.

[15] A. Pardasani, P. N. Indi, S. Banerjee, A. Kamal and V. Garg, "Smart Assistive Navigation Devices for Visually Impaired People," 2019 IEEE 4th International Conference on Computer and Com- munication Systems (ICCCS), Singapore, 2019, pp. 725-729, doi: 10.1109/CCOMS.2019.8821654.