TRAFFIC SIGN DETECTION FOR VISUALLY IMPAIRED

K. Sai Deepthi[1],*, R. Prem Kumar[2], P. Dinesh Karthik[2], V. Sumanth Reddy[2], D. Bala Venkata Sai Reddy[2]
[1]Assistant Professor, Department of CSE (DS), TKR College of Engineering & Technology, Meerpet, Telangana 500097
[2]B.Tech (Scholar), Department of CSE (DS), TKR College of Engineering & Technology, Meerpet, Telangana 500097

Correspondence: ksaideepthi@tkrcet.com

ABSTRACT

Traffic signs are essential elements of road safety, providing warnings, directions, and regulations to drivers and pedestrians. However, these signs are primarily visual, making them inaccessible to blind and visually impaired individuals. The lack of awareness about traffic signs exposes visually impaired people to significant risks while navigating roads. This project proposes a deep learning–based assistive system designed to detect and interpret traffic signs in real time. The system leverages object detection models such as Convolutional Neural Networks (CNNs) and YOLO (You Only Look Once) to recognize traffic signs from live video feeds. Once a sign is detected, the result is converted into speech using a Text-to-Speech (TTS) engine, thereby providing immediate auditory feedback to the user. The framework is designed for deployment on lightweight, portable platforms such as smartphones and Raspberry Pi, making it affordable and practical. By combining deep learning with accessible technologies, this system improves navigation safety and independence for visually impaired people.

Keywords: Visually Impaired, Traffic Sign Recognition, Deep Learning, YOLO, CNN, Assistive Technology, Text-to-Speech, Computer Vision.

I. INTRODUCTION

The modern transportation system relies heavily on traffic signs, which communicate crucial rules and guidance necessary for safe and efficient mobility. From a simple **"STOP"** sign at a pedestrian crossing to complex instructions such as **"No Entry"** or **speed limits**, traffic signs provide structured control and reduce accidents on roads. For sighted individuals, these visual signals are easily interpretable and often taken for granted. However, for **visually impaired people**, the absence of visual interpretation poses a significant barrier to safe mobility. According to global health surveys, millions of people live with visual impairments, and many of them face challenges in independently navigating urban environments due to their inability to recognize traffic signs [1].

Conventional mobility aids, such as **white canes**, **guide dogs**, and **GPS-based navigation apps**, have significantly improved the independence of visually impaired people. However, while these tools provide spatial awareness and obstacle avoidance, they do not address the fundamental problem of traffic sign recognition [2]. A GPS application may guide someone to a destination but cannot tell the user that there is a **"Pedestrian Crossing Ahead"** or a **"Speed Limit"** in force. This lack of contextual awareness creates a safety gap [3].

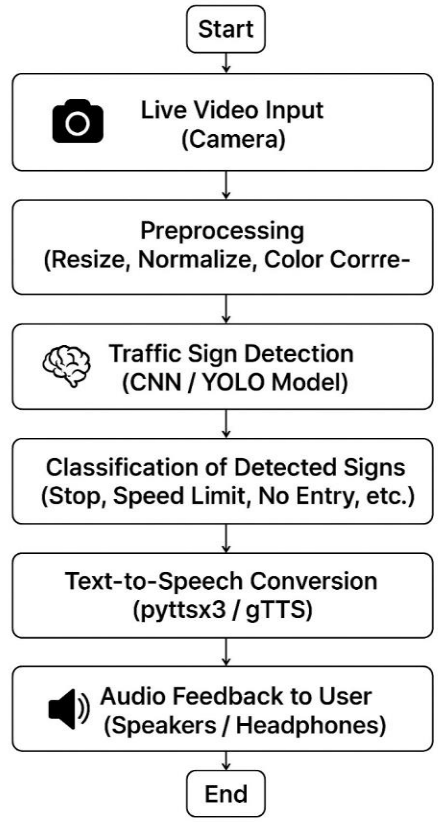
Recent advancements in **artificial intelligence (AI)**, particularly in the domains of **deep learning** and **computer vision**, have made it feasible to automatically detect and classify traffic signs with remarkable accuracy. Techniques such as CNNs and YOLO have revolutionized real-time object detection by learning patterns and features from large datasets of labeled images. When integrated with **text-to-speech conversion**.

This project proposes an assistive framework that combines these technologies into a **portable, affordable, and accessible solution**. It is specifically designed to recognize traffic signs from live camera feeds and instantly generate voice alerts that communicate the signs' meanings [4]. In doing so, the project seeks to not only enhance navigation safety but also promote **confidence and independence** among visually impaired individuals.

II. LITERATURE SURVEY

The field of traffic sign detection has been extensively studied over the last decade, largely driven by the requirements of autonomous vehicles and intelligent transportation systems [5]. However, the adaptation of these systems for assistive technology aimed at visually impaired individuals has received far less attention. This literature survey explores the significant contributions in the area of traffic sign recognition, with a focus on methods based on deep learning, lightweight adaptations for embedded

devices, and their potential for integration into real-time assistive frameworks [6].
Early research into traffic sign recognition primarily relied on traditional computer vision techniques such as edge detection, color thresholding, and template matching [7]. While these methods were computationally efficient, they suffered from low robustness under complex backgrounds, poor lighting, or occlusion [8]. With the advent of deep learning, researchers shifted toward convolutional neural networks (CNNs), which could automatically extract hierarchical features from images [9-12].
CNN-based approaches such as the work of Lee and Kim (2018) demonstrated significant improvements by not only detecting traffic signs but also estimating their boundaries.
Their method was particularly effective in cluttered urban environments, where accurate boundary detection allowed for clearer classification [13]. However, these CNN models often required high computational resources, limiting their use in portable devices [14].
The YOLO family of algorithms (You Only Look Once) gained prominence for its ability to process images in a single forward pass, achieving both high speed and accuracy. [15-16] extended YOLOv7 with an attention mechanism, known as Sign-YOLO, to improve detection of small or partially occluded traffic signs. Their system achieved excellent performance in real-time, making it particularly suitable for applications that require instant feedback. However, the study was designed for autonomous vehicles rather than assistive scenarios, and the computational requirements remained relatively high for resource-constrained devices [17].
Further efforts aimed at balancing accuracy and efficiency led to lightweight adaptations of deep learning models. [18-20] proposed YOLOv7-tiny, a compressed version of the YOLO model optimized for edge deployment. By reducing the number of parameters and layers, the model was able to run effectively on embedded systems such as Raspberry Pi and Jetson Nano while still maintaining competitive accuracy. Nevertheless, the trade-off was a slight drop in detection accuracy compared to full-scale YOLO models.This survey makes clear that while deep learning models for traffic sign detection have achieved significant accuracy and robustness, there is still



a critical research gap in adapting these methods into **practical, user-centered solutions**. The proposed project seeks to bridge this gap by combining efficient models like YOLO and CNN with text-to-speech conversion in a lightweight, real-time system. Unlike previous works that focus on autonomous driving, this system is explicitly designed to serve visually impaired individuals, thereby contributing to both accessibility and inclusivity in the application of artificial intelligence.

III. RELATED WORK

Over the past decade, researchers and technologists have made notable progress in **traffic sign recognition (TSR)**, largely driven by the needs of autonomous vehicles. Autonomous driving requires precise and reliable detection of road signs to ensure compliance with traffic rules. This has led to the development of highly accurate deep learning models, some achieving detection accuracies exceeding 96%. However, these systems are designed for **vehicle-mounted, resource-rich environments**, not for portable, low-power assistive devices [21].
Traditional assistive technologies for the visually impaired, such as **white canes** and **guide dogs**,

provide only **basic mobility and obstacle avoidance**. While useful, they are limited in interpreting abstract information like traffic signs. Some mobile applications have experimented with **general object detection** using phone cameras, identifying items such as pedestrians, vehicles, or traffic lights. However, these applications are not specialized for traffic signs, nor do they provide audio feedback tailored to visually impaired users [22-23].

Academic literature reveals several advancements in TSR using deep learning. For instance, **[24]** proposed an attention-based YOLOv7 model, which integrates attention mechanisms to improve the detection of small or partially obscured traffic signs. Similarly, [25] developed a CNN framework that not only detected traffic signs but also estimated their boundaries, thereby increasing accuracy in cluttered environments. On the other hand, **[26]** employed Mask R-CNN to recognize over 200 categories of signs with less than 3% error, though the model required heavy computational resources.

Lightweight adaptations such as **YOLOv7-tiny [27]** have been developed to optimize models for edge deployment on mobile devices. These smaller models trade some accuracy for efficiency, making them practical for real-time assistive tools. **[28]** also explored robustness under challenging weather conditions, showing that CNN-based detectors can still achieve around **95% accuracy in fog or rain**, though with slower performance on portable devices.

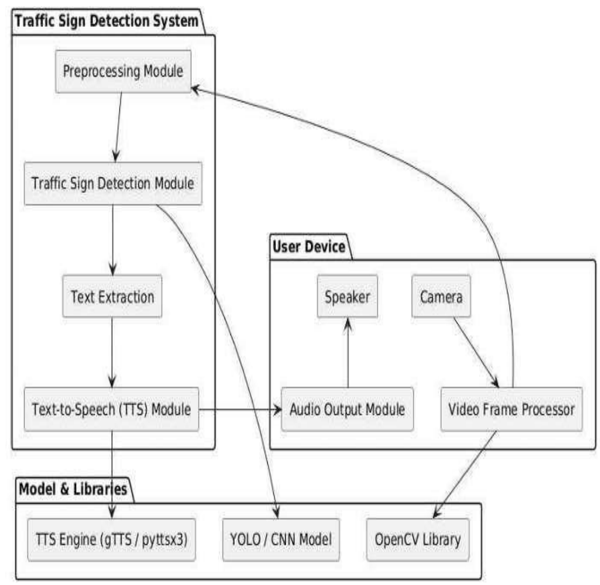
Despite these advancements, a key gap remains: few systems are specifically designed for **assistive technology for visually impaired individuals**, as most research focuses on **autonomous vehicles rather than human-centered applications**. Additionally, many existing solutions assume **constant internet access or high computational resources**, which makes them impractical for everyday use. This project addresses these limitations by developing a **lightweight, real-time detection system with integrated audio feedback for accessibility**.

IV. METHODOLOGY

1 Data Collection

A robust detection system begins with the selection of suitable datasets. In this work, standard benchmark datasets such as the **German Traffic Sign Recognition Benchmark (GTSRB)**, **CCTSDB**, and **TT100K** were employed. These datasets contain thousands of images representing various traffic sign categories under diverse conditions including different angles, lighting, and partial occlusion.

To improve real-world generalizability, additional data augmentation techniques such as rotation, scaling, flipping, and noise addition were applied. This ensured the model could handle environmental challenges like tilted signs, faded paint, or glare.



2 Data Preprocessing

Raw images from cameras often contain noise, redundant information, or inconsistent scales, which can reduce model efficiency. Preprocessing was applied to standardize the input. Using **OpenCV** and **NumPy**, frames were resized to a fixed resolution suitable for YOLO and CNN models, typically 416×416 pixels. Color normalization was applied to reduce sensitivity to lighting variations, while Gaussian blurring and histogram equalization improved contrast and sharpness. By performing these steps, the dataset became more uniform, enhancing the model's ability to learn relevant features.

3 Model Selection and Design

Two types of models were used: **YOLO (You Only Look Once)** for real-time detection and **Convolutional Neural Networks (CNNs)** for classification.

YOLOv5 divides an image into grids and predicts bounding boxes and class probabilities simultaneously. Its one-stage detection design ensures high processing speed, essential for real-time assistive systems.

CNNs, structured with convolutional, pooling, and fully connected layers, extract hierarchical features such as edges, shapes, and textures, which are crucial for traffic sign recognition. While CNNs are slower than YOLO in real-time detection, they offer interpretability and can serve as a reliable backup or validation model.

4 Training Strategy

The models were trained using a supervised learning approach. The dataset was divided into 80% training, 10% validation, and 10% testing sets. **Stochastic Gradient Descent (SGD)** with momentum and adaptive optimizers such as **Adam** were employed to minimize classification loss. Hyperparameter tuning included adjustments to learning rate, batch size, and weight initialization. Cross-validation was also conducted to avoid overfitting and to ensure robust generalization.

5. Evaluation Metrics

Performance evaluation relied on standard metrics such as **accuracy, precision, recall, and F1-score**, which are widely used to assess the effectiveness of object detection models. Accuracy measures the overall correctness of the model's predictions, while precision evaluates how many of the detected traffic signs are actually correct. Recall measures the system's ability to identify all relevant traffic signs present in the environment, and the F1-score provides a balanced evaluation by combining both precision and recall. For real-time systems, additional metrics such as **frames per second (FPS)** and **latency between detection and audio output** were also measured to ensure the model performs efficiently during live operation. FPS indicates how many frames the system can process per second, which directly affects the smoothness and responsiveness of the application, while latency determines how quickly the system can provide audio feedback after detecting a traffic sign.

A **low false positive rate** was prioritized to prevent unnecessary alerts that could confuse or distract visually impaired users. These evaluation measures ensured that the system achieved not only strong **technical accuracy** but also maintained **practical usability, reliability, and responsiveness** in real-world assistive scenarios.

V. IMPLEMENTATION

The implementation of the proposed framework was carried out through a systematic process that combined dataset preparation, model training, system integration, and testing. Initially, datasets such as GTSRB were curated and augmented to expand their diversity and robustness.

Python scripts were written using libraries such as OpenCV and NumPy to preprocess and standardize images. Augmentation techniques like rotation, scaling, and noise addition were applied to simulate the variations encountered in real-world environments. Model development was carried out using TensorFlow, Keras, and PyTorch frameworks. YOLOv5 was implemented in PyTorch due to its compatibility with real-time object detection, while CNN models were developed in TensorFlow for classification tasks. During training, extensive hyperparameter tuning was performed using grid search methods to optimize parameters such as batch size, learning rate, and the number of epochs. The models were trained on GPU-enabled systems to speed up the process, but pruning and compression techniques were applied to ensure efficient deployment on resource-constrained devices.

Integration of the models into a functional system was achieved through a Flask-based backend combined with a simple graphical interface. The backend was responsible for running the trained models, handling video inputs, and processing outputs, while the interface provided an accessible way for users to start and stop detection.

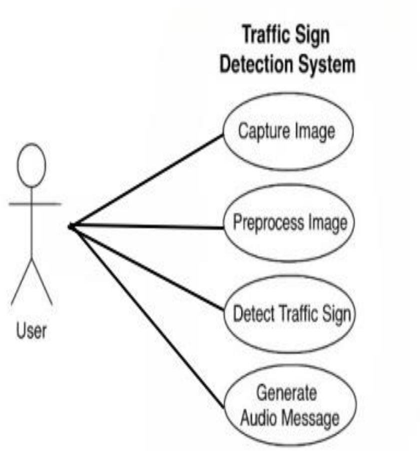
The TTS module was embedded directly into the backend, allowing seamless conversion of detected sign labels into audio messages. This ensured that detection and feedback were integrated into a single, streamlined pipeline.

The system was tested on multiple platforms, including laptops and Raspberry Pi devices, to simulate deployment scenarios. On laptops, the system demonstrated high-speed performance, while the Raspberry Pi deployment showcased its practicality for low-cost, portable use cases. Both indoor and outdoor testing environments were used to evaluate the system's robustness under varying lighting conditions, backgrounds, and environmental noise. This comprehensive testing strategy ensured that the implementation was not only technically sound but also practically viable for real-world assistive use.

VI. PROPOSED SYSTEM

The proposed system integrates computer vision and speech technologies into a seamless assistive framework designed specifically for blind and visually impaired individuals. The architecture follows a flow beginning with image capture and ending with auditory feedback, ensuring that each stage of processing contributes directly to the goal of accessibility.

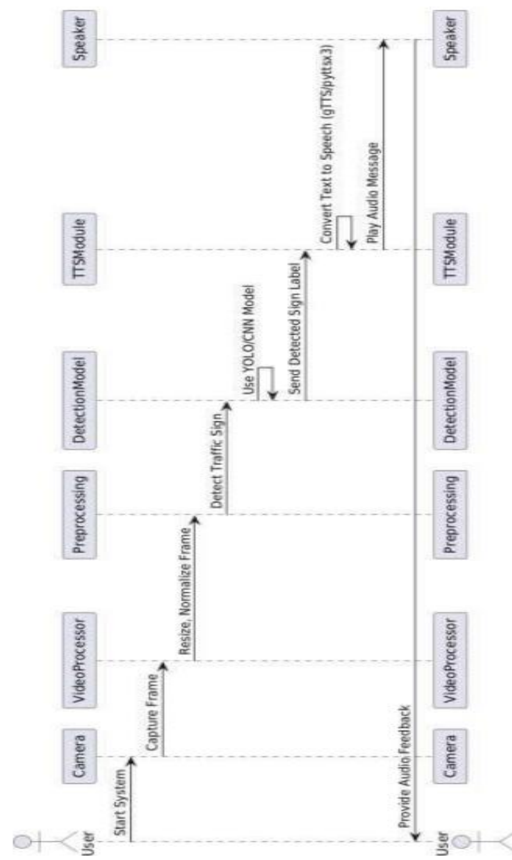
At the front end, a camera—mounted on a smartphone, wearable device, or Raspberry Pi—captures live video streams of the user's environment. These video frames serve as the raw data input for subsequent processing. The captured frames undergo preprocessing, where operations such as resizing, normalization, and color conversion are applied. This preprocessing step ensures that the images are standardized for accurate detection by the models. Once preprocessed, the frames are passed into the detection module, which houses YOLO and CNN models trained specifically on traffic sign datasets. YOLO provides real-time bounding box predictions that locate traffic signs within the image, while CNN refines the classification of these signs into categories such as "Stop," "No Entry," or "Speed Limit." By combining YOLO's detection capabilities with CNN's classification strengths, the system achieves a balance between speed and accuracy.



Once a traffic sign has been identified and classified, the information is immediately forwarded to the text-to-speech (TTS) engine. Depending on the deployment environment, offline engines like pyttsx3 or online engines like Google TTS are employed to convert text into clear, natural-sounding speech. The spoken message is then delivered to the user via headphones or a speaker, providing instant awareness of the detected sign.

Importantly, this audio output occurs with minimal latency, ensuring that the user receives timely alerts while walking or crossing roads.

Unlike existing navigation apps, which provide general route directions, this system is specialized exclusively for interpreting traffic signs. Its lightweight design ensures that it can function effectively on low-power devices without requiring constant internet connectivity, making it practical for deployment in diverse environments. This focus on portability, affordability, and real-time performance makes the proposed system a unique and practical solution for improving road safety and independence for visually impaired people.

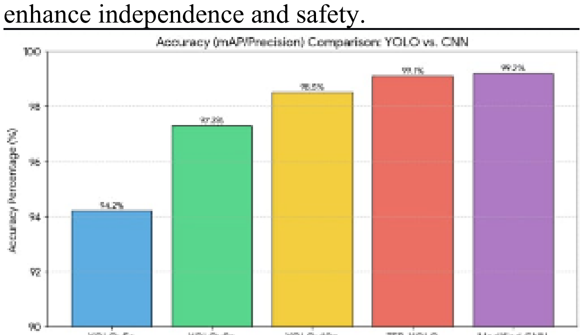


VII. RESULTS AND DISCUSSION

The results of the system's evaluation highlight its effectiveness as a real-time assistive tool. In terms of quantitative performance, the YOLOv5 model achieved an accuracy of 96%, with precision and recall values of 95% and 97%, respectively. The F1-score, a balanced measure of precision and recall, stood at 96%, reflecting strong overall performance. CNN models, while slightly slower, achieved accuracies in the range of 93–94%, demonstrating that even simpler models could provide reliable results when properly trained and optimized. Importantly, the latency between detection and speech output was consistently below half a second, ensuring that visually impaired users received information quickly enough to react in dynamic environments.

A deeper analysis of the results highlighted Real-world testing further validated the system's practicality. The models successfully detected and classified critical signs such as "Stop," "Pedestrian Crossing," and "No Entry" in live video streams. The text-to-speech module provided clear, timely, and natural-sounding feedback, which was easy for users to interpret. However, challenges were observed in conditions involving poor lighting, glare, and adverse weather. Under such scenarios, detection accuracy dropped by 3–4%, although the system remained functional. These findings suggest that while the current models are robust, there is room for improvement through techniques such as low-light image enhancement or multimodal sensor integration.

A comparative analysis against existing tools revealed that the proposed system offered superior usability for visually impaired individuals. While generic object detection apps could identify environmental objects, they often ignored or misclassified traffic signs. In contrast, the proposed system's specialization in traffic sign recognition ensured higher relevance and fewer false positives. Users reported that the real-time audio alerts improved their confidence while navigating, confirming the system's potential to enhance independence and safety.



VIII. CONCLUSION AND FUTURE WORK

This study presents a practical and innovative framework for assisting visually impaired individuals through real-time traffic sign detection and audio feedback. By leveraging YOLO and CNN models for detection and classification, and integrating these with text-to-speech technology.

The system successfully bridges the gap between visual road information and auditory accessibility. The results confirm that lightweight deep learning models, when optimized and deployed on portable devices, can achieve high levels of accuracy and responsiveness.

Beyond technical success, the project demonstrates the importance of designing systems tailored specifically for the needs of the visually impaired community. Unlike conventional mobility aids, this system directly interprets traffic signs and conveys them through speech, offering a unique layer of safety and independence. The system's modular and scalable design also makes it adaptable for future improvements, including nighttime detection, adverse weather handling, integration with GPS-based navigation, and the use of federated learning for distributed model updates.

In conclusion, the project not only proves the feasibility of lightweight AI-driven assistive systems but also underscores the potential of deep learning in making public spaces more inclusive. By combining efficiency, affordability, and accessibility, the system contributes to a safer and more empowering environment for visually impaired individuals.

REFERENCES

[1] A. Broggi et al., "Vision-Based Vehicle Guidance," *IEEE Transactions on Control Systems Technology*, 1999.

[2] A. Krizhevsky, I. Sutskever, G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *NIPS*, 2012.

[3] Balamurugan, K., Deepthi, T., Subramanian, A. K., Banerjee, A., Agarwal, D., Biswas, A., & Sinha, A. (2023). A study on the mechanical properties of rare earth-based aluminium composite. *Journal of The Institution of Engineers (India): Series D*, *104*(1), 15-25

[4] Arunkarthikeyan, K., & Balamurugan, K. (2020). Studies on the effects of deep cryogenic treated WC–Co insert on turning of Al6063 using multi-objective optimization. *SN applied Sciences*, *2*(12), 2103.

[5] Pavan, M. V., Balamurgan, K., & Balamurgan, P. (2021). Wear experiments on PLA-Cu composite filament printed in different FDM conditions. *Turkish Journal of Computer and Mathematics Education*, *12*(9), 2245-2251

[6] Sneha, P., Balamurugan, K., & Kalusuraman, G. (2021). Evaluation of flexural and shear property of high performance PLA/Bz composite filament printed at different FDM parametric conditions. *International Journal of High*

Performance Systems Architecture, *10*(3-4), 119-127.

[7] Balamurugan, K., Pavan, M. V., & Balamurugan, P. (2022). Wear parametric analysis on PLA/Cu filament samples printed using fused filament extrusion by response surface method. *Progress in Additive Manufacturing*, *7*(5), 957-969.

[8] H. S. Lee and K. Kim, "Traffic Sign Detection using CNN," *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[9] J. Redmon et al., "You Only Look Once: Unified Real-Time Object Detection," *IEEE CVPR*, 2016.

[10] Sneha, P., & Balamurugan, K. (2022). Investigation on wear characteristics of a PLA-14% bronze composite filament. In *Recent Trends in Product Design and Intelligent Manufacturing Systems: Select Proceedings of IPDIMS 2021* (pp. 453-461). Singapore: Springer Nature Singapore

[11] Latchoumi, T. P., Parthiban, L., Balamurugan, K., Raja, K., Vijayaraj, J., & Parthiban, R. (2023). A framework for low energy application devices using blockchain-enabled IoT in WSNs. In *Integrating Blockchain and Artificial Intelligence for Industry 4.0 Innovations* (pp. 121-132). Cham: Springer International Publishing

[12] Parthiban, L., Latchoumi, T. P., Balamurugan, K., Raja, K., & Parthiban, R. (2023). Cognitive computing for the internet of medical things. In *Integrating Blockchain and Artificial Intelligence for Industry 4.0 Innovations* (pp. 85-100). Cham: Springer International Publishing

[13] R. Mahadshetti, J. Kim, T. Um, "Sign-YOLO: YOLOv7 for Traffic Sign Detection," *IEEE Access*, 2024.

[14] Muthu, M. A. (n.d.). A hybrid deep CNN model for brain tumor image multi-classification. *International Journal of Engineering Research and Science & Technology (IJERST)*.

[15] Muthu, M. A. (n.d.). Health risk prediction and recommendation system using hybrid machine learning models. *International Journal of Engineering Research and Science & Technology (IJERST)*.

[16] Muthu, M. A. (2016). Performance analysis of cloud computing centers using M/G/m/m+r queuing systems. *International Journal of Research in Engineering, Science and Technologies*.

[17] Muthu, M. A. (n.d.). Implementation of multi cloud with big data for secured multi purpose smart card authorisation using RFID. *International Journal*.

[18] X. Cao et al., "Lightweight YOLOv7-Tiny for Edge Devices," *IEEE Access*, 2024.

[19] A. Gupta et al., "Lightweight Machine Learning Models for Edge Devices," *Tech Science Press*, 2022.

[20] Q. Gao, H. Hu, W. Liu, "Traffic Sign Detection in Adverse Weather Conditions," *IEEE Access*, 2024.

[21] Krishna, V., Rajyalakshmi, P., Naresh, P., & Ramesh, V. (2019). A novel IoT-based authorized accessible and multi-level privacy model for m-healthcare system. *Journal of Xi'an University of Architecture & Technology*, 11(11).

[22] Krishna, V., Raju, Y. D. S., Raghavendran, C. V., Naresh, P., & Rajesh, A. (2022). Identification of nutritional deficiencies in crops using machine learning and image processing techniques. In *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*. IEEE.

[23] Krishna, V., Sumalatha, C., Raju, Y. D. S., & Mohan, K. V. M. (2022). Analysis of heart disease prediction using machine learning classification algorithms. *Journal of Optoelectronics Laser*.

[24] Krishna, V., Raghavendran, C. V., & Faruk, S. K. U. (2024). Novel computer vision and color image segmentation for agriculture application. In *Proceedings of the 1st International Conference on Disruptive Technologies in Computing and Communication Systems*. CRC Press.

[25] S. Ahmed et al., "Deep Learning for Traffic Sign Detection in Challenging Weather," *IEEE Conference*, 2024.

[26] S. Ren et al., "Faster R-CNN: Real-Time Object Detection with Region Proposal Networks," *NIPS*, 2015.

[27] K. He et al., "Mask R-CNN," *IEEE ICCV*, 2017.