tinyML-02: Efficient and High-performance solution for Plant Disease Detection

Sudharshan N
sudharshann106@gmail.com
Chennai, Tamil Nadu, India

Saran Sundar S
saransrimathi590@gmail.com
Chennai, Tamil Nadu, India

Raagav S R
raagav.ingenieur@gmail.com
Chennai, Tamil Nadu, India

Rohit S
rohitsarav14@gmail.com
Chennai, Tamil Nadu, India

ABSTRACT

This paper presents the development of an efficient, low-cost, and low-power TinyML solution for plant disease detection. The objective is to reliably identify and classify various plant diseases using a model optimized for deployment on resource-constrained devices. The paper concludes with insights into potential improvements with enterprise accounts and extended training, and the GitHub repository is provided for further exploration and implementation. The proposed solution opens avenues for scalable and high-performance TinyML applications in agriculture, contributing to early and efficient plant disease detection.

KEYWORDS

Plant Disease Detection, TinyML, ESP32-S3, CNN, EdgeAI, Machine Learning

1 INTRODUCTION

Plant diseases pose a significant threat to global agriculture, impacting food security, human health, and the world economy. Caused by various pathogens, these diseases can lead to crop losses, affecting livelihoods, nutrition, and trade. The consequences extend beyond the field, influencing economic stability, research expenditures, and environmental sustainability. Efforts to combat plant diseases are vital for preserving agricultural productivity and mitigating the far-reaching implications on humanity and the global economy. Innovative solutions, such as TinyML for early disease detection, hold promise in addressing these challenges and promoting sustainable agriculture. Our endeavor to advance plant disease detection has led us to the development of a state-of-the-art Quantized CNN. This novel approach harnesses the power of quantization, a technique that optimizes the computational efficiency of neural networks while preserving the essential features necessary for robust disease classification.

2 METHODOLOGY

The methodology encompasses defining objectives for an efficient and cost-effective plant disease detection system, selecting hardware (XIAO ESP32S3 Sense) and acquiring a diverse dataset. A custom TinyML model is trained and optimized on Edge Impulse for hardware constraints and accuracy. The model is quantized and integrated into the hardware prototype, including components like the OV2640 camera and Lithium-Ion battery. Performance evaluation involves testing for accuracy, latency, and resource utilization, with overall project cost transparency. Documentation and a GitHub repository facilitate sharing the streamlined process and outcomes, emphasizing reliability, efficiency, and cost-effectiveness in the development of a scalable TinyML solution for plant disease detection.

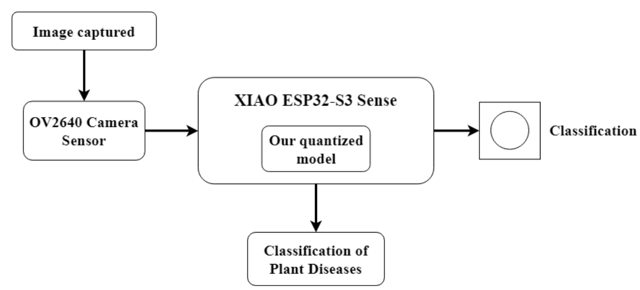

Figure 1: Flow chart representing the entire process

2.1 Experimental Setup

List of components used

- **Microcontroller Unit (MCU):** `XIAO ESP32-S3 Sense`with the ESP32-S3R8 32-bit Processor, 2.4GHz WiFi, low-power Bluetooth, Xtensa processor chip (240 MHz), 8MB PSRAM, and 8MB FLASH.
- **Camera Sensor:** Detachable OV2640 camera sensor for capturing images at flexible resolutions.
- **Power Source:** M-STAR 3.7v -18650 -4300 mAh Lithium-Ion Battery for providing power to the system.
- **3D Printed Case for XIAO-ESP32S3-Sense:** The 3D printed case for the XIAO ESP32-S3 Sense serves as a protective and functional housing, enhancing the overall robustness and aesthetic appeal of the hardware setup in our plant disease detection system.

3 COST ESTIMATIONS

The following table (table 1) provides a comprehensive breakdown of the estimated budget for the development and implementation of the product. It outlines the detailed allocation of financial resources across various components and expenses associated with the project

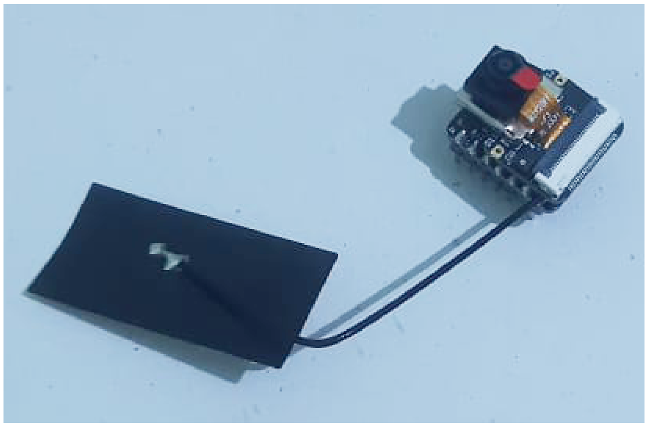

Figure 2: ESP32 XIAO Sense board S3

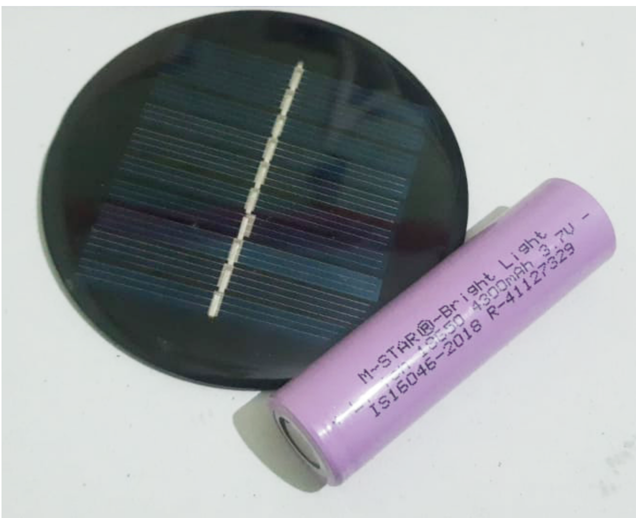

Figure 3: M-STAR Battery with Solar Panel

Table 1: Cost Estimation

Components	Quantity	Price (INR)
XIAO ESP32-S3 Sense	1	1499
BATTERY-Li-ion 18650 4300mah	2	155 each
3D Printed Case	1	20

4 DATASET

We've utilized the dataset given in the data source section of the competition https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-datasetTo address imbalances inherent in the dataset, we strategically removed heavily undersampled classes, ensuring a more equitable representation of various plant diseases. This preprocessing step is essential for preventing model bias towards more frequently occurring classes and improving the overall robustness of the machine learning model. To further enhance the model's generalization and mitigate overfitting, we leveraged advanced techniques available in Edge Impulse. The platform's Data Augmentation feature was employed to introduce variations into the training data, thereby exposing the model to a more diverse range of scenarios. Additionally, class re-weighting was applied to account for the imbalances in the remaining classes, a crucial step to prevent the model from favoring the majority class during training. For the final model, we made a deliberate decision to exclusively utilize the testing data from the dataset. This ensures a more objective evaluation of the model's performance on unseen instances, reflecting its real-world applicability. In contrast, the initial model underwent training using both testing and training data, allowing us to explore the model's potential in different scenarios. This strategic approach to dataset utilization and preprocessing underscores our commitment to developing a robust and unbiased plant disease detection model. The integration of cutting-edge techniques from Edge Impulse enriches the dataset preprocessing pipeline, contributing to the model's accuracy and reliability in identifying and classifying plant diseases.

Figure 4: 3D model for casing

5 ML MODEL

In our plant disease detection project, we adopted a custom model architecture tailored to meet the constraints imposed by Edge Impulse for non-enterprise entities. The model design comprises three layers of a convolutional neural network (CNN), strategically constructed to balance performance and resource efficiency. The first layer features a block of 32 filters with a kernel size of 3, providing the initial capability to capture intricate patterns in the input data.

Following this, the subsequent two layers incorporate blocks of 16 filters each, each with a kernel size of 3. This configuration is selected with careful consideration of the computational limitations imposed by the Edge Impulse platform, ensuring the model remains efficient and deployable on resource-constrained devices.

These design choices prioritize simplicity and effectiveness, allowing the model to efficiently extract relevant features from the plant disease dataset without exceeding the computational constraints. While the model architecture may be compact, its effectiveness is showcased through rigorous training and optimization, as evidenced by the model's accuracy in identifying and classifying plant diseases in the final implementation. There are some cases where the model fails to predict the classes correctly. This can be solved by adding couple of layers and training on complete dataset.

The Whole model was trained on Edge Impulse as we had trouble deploying our custom model in the Edge device. The First model we trained was very good in terms of accuracy and size(We'll attach the .ipynb file, a quantized model with the results in a miscellaneous section of the GitHub link we'll provide).

Table 2: Unoptimized float32 model

Properties	Image	Classifier	Total
Latency	15 ms	7217 ms	7232 ms
Ram	4.0 K	1.4 mb	1.4 mb
Flash	-	1.4 mb	-

Table 3: Quantised Int8 Model

Properties	Image	Classifier	Total
Latency	15 ms	3227 ms	3242 ms
Ram	4.0 K	363.3 K	363.3 K
Flash	-	377.1 K	-

6 ASSEMBLY

7 POWER CONSUMPTION

Webcam Web application:

1. Type-C Average power consumption: 5V/138mA
2. Type-C Photo moment: 5V/341mA
3. Battery Average power consumption: 3.8V/154mA
4. Battery Photo moment: 3.8V/304mA
5. Charging battery current: 100mA

Average Consumption during Inference is 1.74 watts, with 2 lithium-ion batteries (4300 mah) this system will last for 16.5 hours.

8 RESULTS

We have summarized our results as follows:

Model 1: Based on MobileNetV2, demonstrated remarkable performance, achieving approximately 99.9% categorical accuracy on both the training and test sets after 31 epochs of training. Despite its exceptional accuracy, we encountered challenges in quantizing this model for deployment on the ESP32-S3, facing several deployment errors that hindered its integration into our targeted hardware platform.

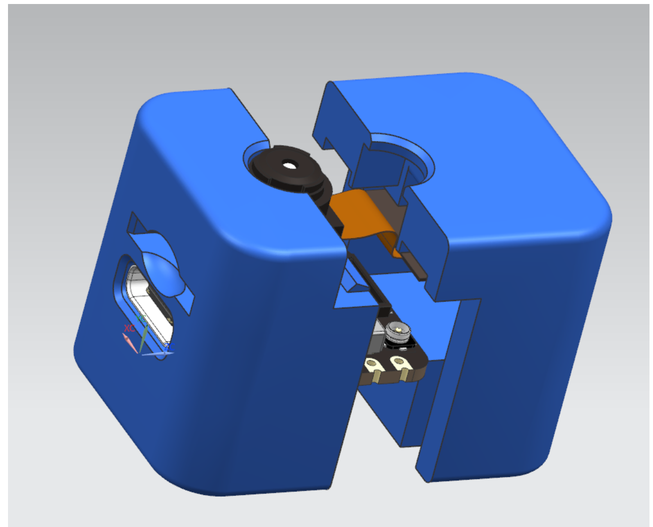

Figure 5: 3D Model of casing, ESP32, Camera assembled

Figure 6: 3D Printed Case for XIAO-ESP32S3-Sense

Model 2: a custom-designed architecture detailed in the ML MODEL section, was employed for the final implementation. Trained for 7 epochs with a grayscale approach using Edge Impulse, it yielded a respectable accuracy of 74.00%Ṅotably, Edge Impulse's constraints for non-enterprise accounts, limiting training to 20 minutes and 4 GB of data, influenced the training duration and model complexity. The quantized and unquantized properties of this model are thoroughly outlined in Table 2 and Table 3, reflecting a balance between accuracy and resource efficiency.

It's essential to acknowledge the potential for accuracy improvement by leveraging enterprise accounts and extending training epochs, factors constrained by Edge Impulse's non-enterprise limitations. Moreover, the XIAO ESP32-S3 Sense, with its compact dimensions of 21 x 17.5mm, offers a practical and versatile solution for deployment. Its capabilities, including Wi-Fi and Bluetooth functionality for message transmission and an SD card facility for data storage, further enhance its suitability for integration into drone applications. This integration aligns with the broader aim of fostering seamless and efficient communication within agricultural ecosystems for timely plant disease detection and management.

9 LINKS

9.1 GitHub

https://github.com/sudharshan2001/tinyML-02-Plant-Disease-Detection

10 APPENDIX

The "`figure of confusion matrix`" provides a detailed analysis of the predictive model. It provides a detailed breakdown of the model's predictions, comparing them to the actual outcomes.*(refer to the fig on the next page for a detailed view)*

REFERENCES

[1] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. 2011. Algorithms for Hyper-Parameter Optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems* (Granada, Spain) *(NIPS'11)*. Curran Associates Inc., Red Hook, NY, USA, 2546–2554.

[2] Robert David, Jared Duke, Advait Jain, Vijay Janapa Reddi, Nat Jeffries, Jian Li, Nick Kreeger, Ian Nappier, Meghna Natraj, Shlomi Regev, Rocky Rhodes, Tiezhen Wang, and Pete Warden. 2021. TensorFlow Lite Micro: Embedded Machine Learning on TinyML Systems. arXiv:2010.08678

[3] Amir Gholami, Sehoon Kim, Zhen Dong, Zhewei Yao, Michael W. Mahoney, and Kurt Keutzer. 2021. A Survey of Quantization Methods for Efficient Neural Network Inference. arXiv:2103.13630

[4] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861

[5] Xijie Huang, Zechun Liu, Shih-Yang Liu, and Kwang-Ting Cheng. 2023. Efficient Quantization-aware Training with Adaptive Coreset Selection. arXiv:2306.07215

[6] Shawn Hymel, Colby Banbury, Daniel Situnayake, Alex Elium, Carl Ward, Mat Kelcey, Mathijs Baaijens, Mateusz Majchrzycki, Jenny Plunkett, David Tischler, Alessandro Grande, Louis Moreau, Dmitry Maslov, Artie Beavis, Jan Jongboom, and Vijay Janapa Reddi. 2023. Edge Impulse: An MLOps Platform for Tiny Machine Learning. arXiv:2212.03332

[7] Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. 2017. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. arXiv:1712.05877

[8] Andrey Kuzmin, Markus Nagel, Mart van Baalen, Arash Behboodi, and Tijmen Blankevoort. 2023. Pruning vs Quantization: Which is Better? arXiv:2307.02973

[9] Ji Lin, Wei-Ming Chen, Yujun Lin, John Cohn, Chuang Gan, and Song Han. 2020. MCUNet: Tiny Deep Learning on IoT Devices. arXiv:2007.10319

[10] Simon Narduzzi, Engin Türetken, Jean-Philippe Thiran, and L. Andrea Dunbar. 2022. Adaptation of MobileNetV2 for Face Detection on Ultra-Low Power Platform. arXiv:2208.11011

[11] Shvetank Prakash, Matthew Stewart, Colby Banbury, Mark Mazumder, Pete Warden, Brian Plancher, and Vijay Janapa Reddi. 2023. Is TinyML Sustainable? Assessing the Environmental Impacts of Machine Learning on Microcontrollers. arXiv:2301.11899

[12] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2019. MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv:1801.04381

[13] Mingzhu Shen, Feng Liang, Ruihao Gong, Yuhang Li, Chuming Li, Chen Lin, Fengwei Yu, Junjie Yan, and Wanli Ouyang. 2021. Once Quantization-Aware Training: High Performance Extremely Low-bit Architecture Search. arXiv:2010.04354

[14] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556

[15] Suorong Yang, Weikang Xiao, Mengchen Zhang, Suhan Guo, Jian Zhao, and Furao Shen. 2023. Image Data Augmentation for Deep Learning: A Survey. arXiv:2204.08610

ACCURACY
74.0%

LOSS
1.12

Confusion matrix (validation set)

	F1-SCORE	PRECISION	RECALL
APPLE___APPLE_SCAB	0.72	[illegible]	0.78
APPLE___BLACK_ROT	0.75	0.79	0.72
APPLE___CEDAR_APPLE_RUST	[illegible]	[illegible]	0.79
APPLE___HEALTHY	[illegible]	[illegible]	0.81
BLUEBERRY___HEALTHY	0.81	0.75	0.88
CHERRY_(INCLUDING_SOUR)___POWDERY_MILDEW	0.79	0.76	0.82
CHERRY_(INCLUDING_SOUR)___HEALTHY	0.83	0.91	0.77
CORN_(MAIZE)___CERCOSPORA_LEAF_SPOT GRAY_LEAI	0.77	0.85	0.71
CORN_(MAIZE)___COMMON_RUST_	0.98	0.97	0.98
CORN_(MAIZE)___NORTHERN_LEAF_BLIGHT	0.77	0.72	0.83
CORN_(MAIZE)___HEALTHY	0.95	0.91	1.00
GRAPE___BLACK_ROT	0.76	[illegible]	0.90
GRAPE___ESCA_(BLACK_MEASLES)	0.82	0.95	0.71
GRAPE___LEAF_BLIGHT_(ISARIOPSIS_LEAF_SPOT)	0.91	1.00	0.84
GRAPE___HEALTHY	0.85	0.82	0.88
ORANGE___HAUNGLONGBING_(CITRUS_GREENING)	0.86	1.00	0.76
PEACH___BACTERIAL_SPOT	0.76	0.94	[illegible]
PEACH___HEALTHY	0.79	0.72	0.89
PEPPER,_BELL___BACTERIAL_SPOT	[illegible]	[illegible]	[illegible]
PEPPER,_BELL___HEALTHY	0.52	[illegible]	0.41
POTATO___EARLY_BLIGHT	0.90	0.89	0.91
POTATO___LATE_BLIGHT	[illegible]	[illegible]	[illegible]
POTATO___HEALTHY	[illegible]	0.96	0.35
RASPBERRY___HEALTHY	0.77	0.72	0.83
SOYBEAN___HEALTHY	0.81	0.84	0.78
SQUASH___POWDERY_MILDEW	0.87	0.90	0.85
STRAWBERRY___LEAF_SCORCH	0.77	0.84	0.72
STRAWBERRY___HEALTHY	0.90	0.87	0.94
TOMATO___BACTERIAL_SPOT	0.74	[illegible]	0.84
TOMATO___EARLY_BLIGHT	0.47	0.41	0.56
TOMATO___LATE_BLIGHT	0.36	0.33	0.40
TOMATO___LEAF_MOLD	[illegible]	0.74	0.49
TOMATO___SEPTORIA_LEAF_SPOT	[illegible]	[illegible]	[illegible]
TOMATO___SPIDER_MITES TWO-SPOTTED_SPIDER_MIT	0.79	0.81	0.77
TOMATO___TARGET_SPOT	[illegible]	0.55	[illegible]
TOMATO___TOMATO_YELLOW_LEAF_CURL_VIRUS	0.86	0.90	0.83
TOMATO___TOMATO_MOSAIC_VIRUS	0.88	0.83	0.93
TOMATO___HEALTHY	[illegible]	0.78	[illegible]

Figure 7 : Precision and Recall score for each classes