AI4D Research Lab

TinyML Challenge 2023

Scalable and High-Performance TinyML Solutions for Wildlife Monitoring

AI4D Research Lab

www.ai4dlab.ac.tz

27th November, 2023

Outline

- Objective
- Problem Statement
- Methodology
- Results and discussion
- Limitation and Future works
- Conclusion

Objective

AI4D LAB
Anglophone Africa

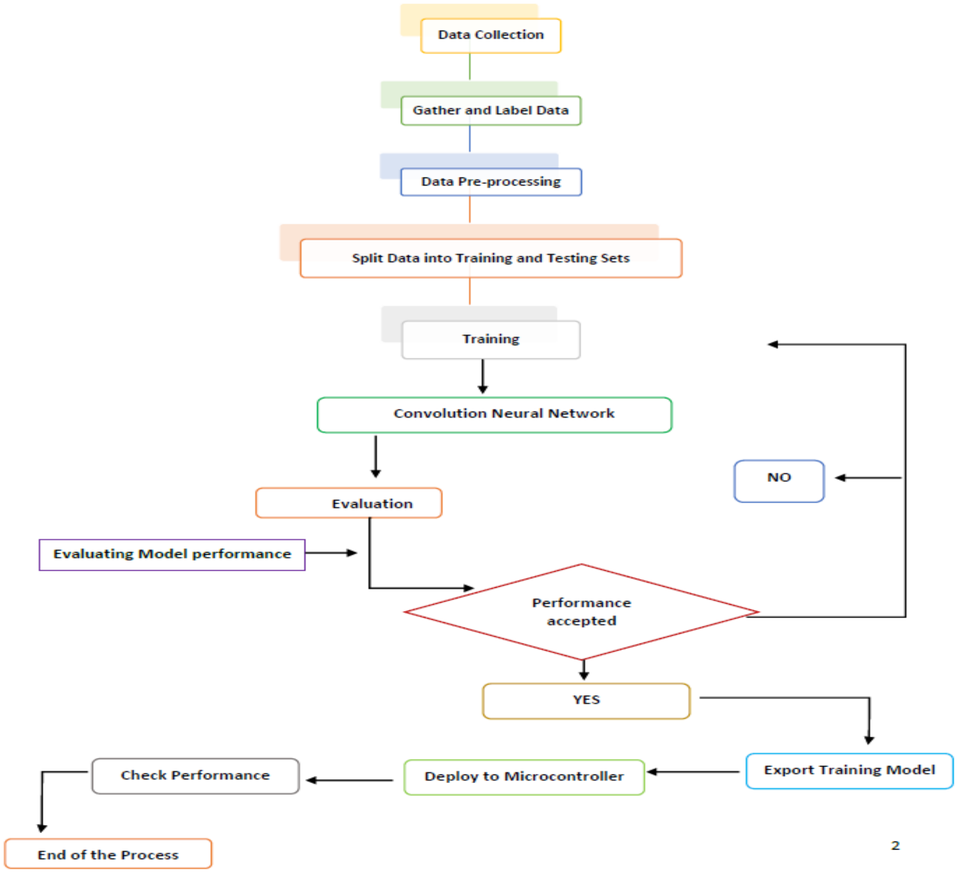
To develop a wild life monitoring tool using TinyML models that will be capable of detecting and classifying wild animals in the limited resources areas to enhance wildlife conservation.

Problem statement

AI4D LAB
Anglophone Africa

- Monitoring wild animals constitutes ecological preservation and wildlife management
- Machine learning (ML) approach can easily achieve with high accuracy with minimal time and cost
- However, resides in the development of unobtrusive, energy-efficient, and precise monitoring solutions capable of functioning in remote and arduous environments.
- The developed tool leveraged TinyML, to tackle the obstacles of proficiently implementing ML models on low-resource hardware microcontrollers.

Methodology



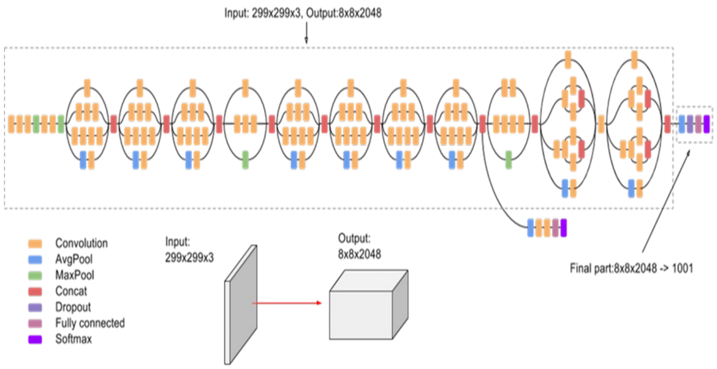
- Traditional model (Using Python)
- Edge Impulse Model

Model 1: Total sample **70552**
Classes **10** Training **56446** Testing **14106**

Model 2: Classes **10** Total sample **2500**

- **Data Source**
- **Data Pre-processing**
- **Model Training**
- **Model Evaluation**
- **Model Deployment**

Methodology



Using Python

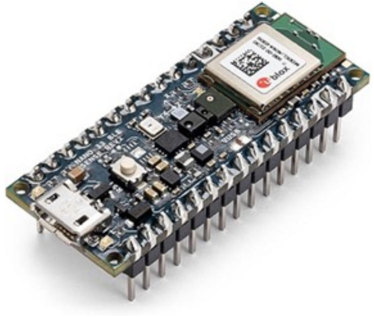
- Convolutional Neural Network (CNN)
- To achieve high performance with a relatively modest computation cost Inception v3 was used
- Hyperparameters were fine-tuned, and transfer learning techniques were employed to ensure efficient model training

Edge Impulse Studio

- MobileNetV1 and MobileNetV2 models
- Supports automatic hyperparameter tuning and data augmentation, streamlining the training process
- Optimize model to run efficiently on microcontrollers or edge devices, taking into account the limited computational resources

Model deployment

Arduino Nano 33 BLE Sense



Results and Discussion

Results

From CNN Inception
- Model performance

Classes	Precision	Recall	F1-score	Support
Buffalo	0.84	0.23	0.36	647
Elephant	0.97	0.36	0.52	718
Rhinoceros	0.87	0.53	0.66	1485
Zebra	0.97	0.79	0.87	1124
Giraffe	0.86	0.56	0.68	599
Impala	0.93	0.91	0.92	5903
Eland	0.13	0.97	0.24	230
Hyaenabrown	0.60	0.52	0.55	196
Lion	0.43	0.82	0.56	814
Leopard	0.95	0.92	0.94	2390
Accuracy	**-**	**-**	**0.78**	**14106**
Macro avg	**0.76**	**0.66**	**0.63**	**14106**
Weighted avg	**0.88**	**0.78**	**0.80**	**14106**

Results and Discussion

AI4D LAB
Anglophone Africa

From Edge Impulse studio

Using MobileNetV1

Model — Model version: Unoptimized (float32)

Last training performance (validation set)

ACCURACY 87.5%

LOSS 0.40

Confusion matrix (validation set)

	BUFFAL(	DEER	ELEPHAI	GIRAFFE	HORSE	LION	MONKEY	RHINO	TIGER	ZEBRA
BUFFALC	83.3%	0%	8.3%	0%	0%	4.2%	0%	4.2%	0%	0%
DEER	0%	90.7%	0%	3.7%	1.9%	0%	0%	1.9%	0%	1.9%
ELEPHAN	4.8%	0%	73.8%	0%	7.1%	2.4%	2.4%	9.5%	0%	0%
GIRAFFE	0%	2.4%	0%	88.1%	4.8%	0%	2.4%	0%	0%	2.4%
HORSE	0%	2.1%	2.1%	0%	89.4%	0%	0%	2.1%	4.3%	0%
LION	0%	2.4%	0%	0%	0%	95.1%	2.4%	0%	0%	0%
MONKEY	0%	3.2%	0%	0%	6.5%	3.2%	87.1%	0%	0%	0%
RHINO	8.3%	0%	8.3%	0%	2.8%	0%	0%	80.6%	0%	0%
TIGER	0%	0%	0%	0%	0%	2.1%	2.1%	0%	95.8%	0%
ZEBRA	0%	2.3%	2.3%	0%	2.3%	2.3%	0%	0%	4.7%	86.0%
F1 SCORI	0.82	0.91	0.78	0.91	0.85	0.92	0.87	0.81	0.94	0.90

Enable EON™ Compiler *Same accuracy, up to 50% less memory. Learn more*

Quantized (int8) — Selected ✓

	IMAGE	TRANSFER LEARNING	TOTAL
LATENCY	11 ms.	1,215 ms.	**1,226 ms.**
RAM	4.0K	124.8K	**124.8K**
FLASH	-	304.7K	-
ACCURACY			**84.21%**

Unoptimized (float32) — Select

	IMAGE	TRANSFER LEARNING	TOTAL
LATENCY	11 ms.	9,241 ms.	**9,252 ms.**
RAM	4.0K	300.1K	**300.1K**
FLASH	-	862.3K	-
ACCURACY			**83.21%**

Expected On-device performance metrics after deployment

Results and Discussion

AI4D LAB
Anglophone Africa

From Edge Impulse Studio

Using MobileNetV2

Model Model version: Quantized (int8)

Last training performance (validation set)

ACCURACY 96.3%

LOSS 0.14

Confusion matrix (validation set)

	BUFFAL	DEER	ELEPHA	GIRAFFE	HORSE	LION	MONKEY	RHINO	TIGER	ZEBRA
BUFFALO	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%
DEER	0%	98.1%	0%	0%	0%	0%	1.9%	0%	0%	0%
ELEPHAN	0%	0%	97.6%	0%	2.4%	0%	0%	0%	0%	0%
GIRAFFE	0%	2.4%	0%	97.6%	0%	0%	0%	0%	0%	0%
HORSE	0%	2.1%	0%	0%	93.6%	2.1%	0%	2.1%	0%	0%
LION	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%
MONKEY	0%	0%	3.2%	0%	0%	0%	96.8%	0%	0%	0%
RHINO	5.6%	0%	5.6%	0%	0%	0%	0%	88.9%	0%	0%
TIGER	0%	0%	0%	0%	0%	2.1%	0%	0%	97.9%	0%
ZEBRA	0%	0%	2.3%	0%	0%	2.3%	0%	0%	2.3%	93.0%
F1 SCOR	0.96	0.97	0.94	0.99	0.96	0.96	0.97	0.93	0.98	0.96

Enable EON™ Compiler *Same accuracy, up to 50% less memory. Learn more*

Quantized (int8)

Selected ✓

	IMAGE	TRANSFER LEARNING	TOTAL
LATENCY	11 ms.	3,758 ms.	**3,769 ms.**
RAM	4.0K	254.0K	**254.0K**
FLASH	-	307.1K	-
ACCURACY			**89.28%**

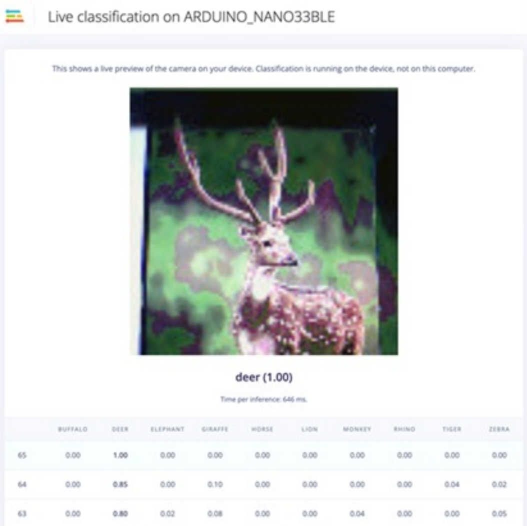
Unoptimized (float32)

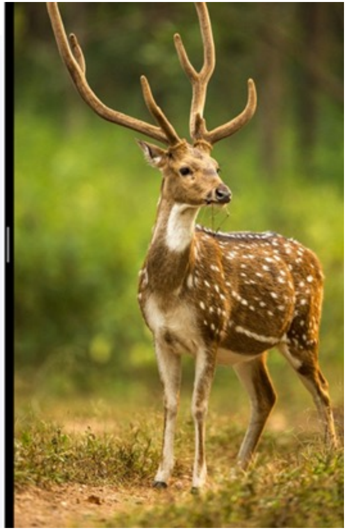
Select

	IMAGE	TRANSFER LEARNING	TOTAL
LATENCY	11 ms.	12,069 ms.	**12,080 ms.**
RAM	4.0K	816.3K	**816.3K**
FLASH	-	870.9K	-
ACCURACY			**92.27%**

Expected On-device performance metrics after deployment

Live Classification on Web API





Live Classification on Terminal

```
❯ edge-impulse-run-impulse --debug
Edge Impulse impulse runner v1.22.0
[SER] Connecting to /dev/tty.usbmodem144201
[SER] Serial is connected, trying to read config...
[SER] Retrieved configuration
[SER] Device is running AT command version 1.8.0

Want to see a feed of the camera and live classification in your browser? Go to
http://192.168.1.64:4915

[SER] Started inferencing, press CTRL+C to stop...
Predictions (DSP: 14 ms., Classification: 646 ms., Anomaly: 0 ms.):
    buffalo: 0.01172
    deer: 0.02344
    elephant: 0.27344
    giraffe: 0.18359
    horse: 0.03906
    lion: 0.08203
    monkey: 0.05469
    rhino: 0.23828
    tiger: 0.00000
    zebra: 0.09375
Predictions (DSP: 14 ms., Classification: 646 ms., Anomaly: 0 ms.):
    buffalo: 0.01562
    deer: 0.01953
```

$ Edge-impulse-run-impulse --debug

Discussion

This work provides promising results on both models. Moreover, the Edge Impulse models provide outstanding results from both MobileNetV1 and MobileNetV2 as compared to a work conducted by **Richard Gotthard and Marcus Broström** in which a results of f1-score on MobileNetV2 was 0.67 %

Limitation and Future Works

AI4D LAB
Anglophone Africa

Limitation

- Resources such as a GPU and a powerful computer capable of running the model in a short period of time
- Due of time constraints, we were unable to install the device in an actual wildlife park.

Future works

- To advance the capabilities of TinyML models by exploring Real-Time Operating Systems (RTOS) for resource-constrained embedded devices.
- To develop solutions (Embedded devices) that are not only low-cost and low-power but also scalable and adaptable, ensuring seamless integration into diverse ecosystems.

Conclusion

AI4D LAB
Anglophone Africa

- We found that MobileNetV1 is the best model for wildlife detection and classification, however we can't compare it to the Python model *(Inception model)* because they utilize different datasets.

- This work honors researchers and conservationists who persistently study animals and promote sustainable practices. TinyML integration is at the forefront of a technological revolution in conservation, promoting sustainable practices and strengthening global biodiversity protection efforts. As we implement these solutions, we expect technology to help protect our natural world.

Thank You!