

Deploying Machine Learning Application on Edge Devices, a Study of Challenges and Solutions

Students: Jenniffer Hierro

Mentor: Luis Robaina, MITRE

Instructor/Faculty: Seyedmasoud Sadjadi, Florida International University

PROBLEM

In recent years, the interest in edge intelligence has grown exponentially worldwide. Most of the current efforts on these fields are focused on developing better algorithms and architectures from the theoretical side, while not much effort is put on understanding the engineering challenges of bringing AI systems to real-life, especially on edge devices. The research concluded in this project serves to understand the challenges and possible solutions for the development, security and deployment of ML onto edge devices.

CURRENT SYSTEM

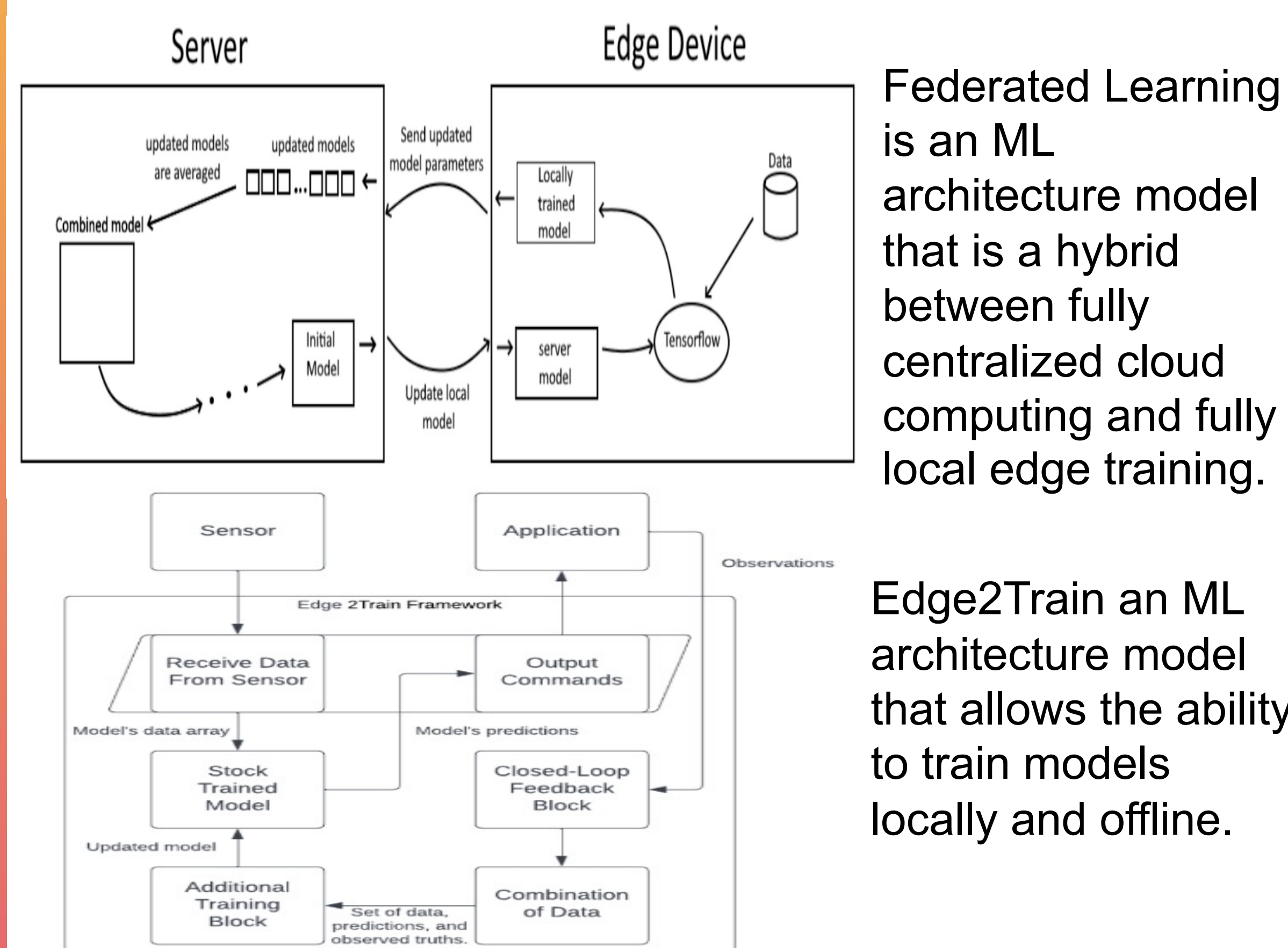
The current system is based of research on collaborative training architectures such as: Federated Learning and Edge2Train. Federated Learning shows where a server employs multiple devices and allocates training tasks for them, which works for more powerful devices. Edge2Train is a framework that allows for the training on less powerful edge devices locally and offline. These forms of training edge devices are still being researched as they do not operate successfully in all types of devices.

REQUIREMENTS

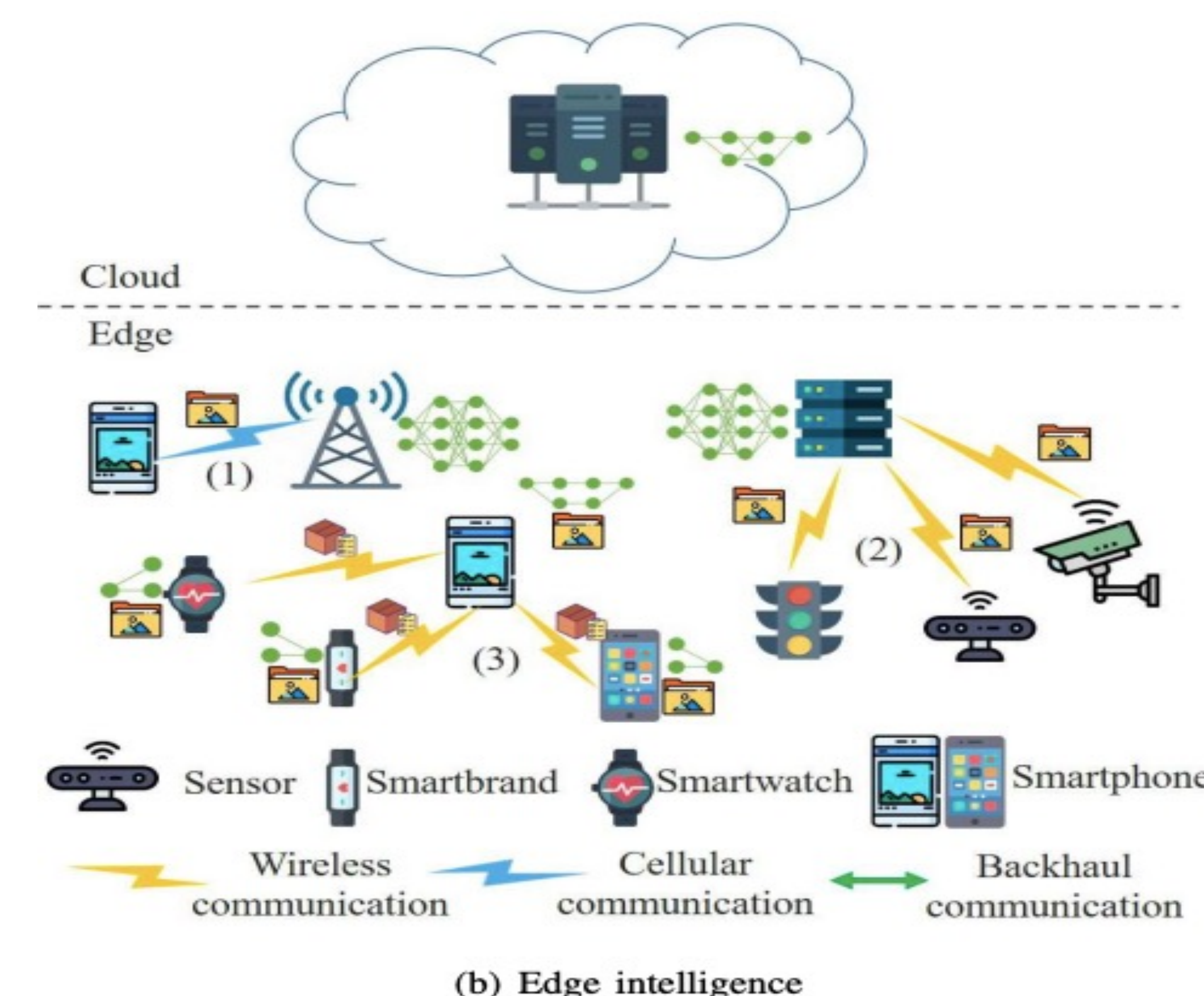
Deploying machine learning application on edge devices research consisted of:

Researching the development side of ML on edge devices.
Researching the deployment side of ML on edge devices.
Researching the security side of ML on edge devices.
Learning and utilizing LaTeX in VSCode to write the research.
Using and pushing onto GitHub.

SYSTEM DESIGN



OBJECT DESIGN



IMPLEMENTATION

LATEX



VERIFICATION

Verification consists of Capstone 1 & 2 students peer-reviewing the research written by each Capstone 2 student in their corresponding section. The research was all reviewed by the mentor as well.

REFERENCES

Object design image: <https://arxiv.org/pdf/2003.12172.pdf>

ACKNOWLEDGEMENT

The material presented in this poster is based upon the work supported by XXXXX (name of the project sponsor: federal, state or local government, or corporate partners, where applicable). We/I thank XXXX for his/her/their assistance and mentorship that I/we received throughout the senior design project.

SUMMARY

This research paper studied the challenges of deploying Machine Learning on edge devices. The team studied and wrote about three main areas: development, deployment, and security of ML on edge devices. The development team researched the different types of training architectures, data & model compression, and ways of running the model on cloud vs locally. The security team researched the types of threats that can occur on edge devices and what defense can be used for those security threats. The deployment team wrote about the deployment aspect and how monitoring after deployment of the model would be.

PERSONAL CONTRIBUTIONS

- Wrote and researched in the development section:
- Training architectures
 - Solo training
 - Collaborative training.
 - Data Compression
 - Model Compression
 - Quantization
 - Model Pruning
 - Attended meetings
 - Reviewed paper to ensure quality and correctness.