INDUSTRIAL PLACEMENT WORK DIARY

Academic Year: 2022/2023

20 February – 11 June 2023

Student's Name: Hanlin Cai

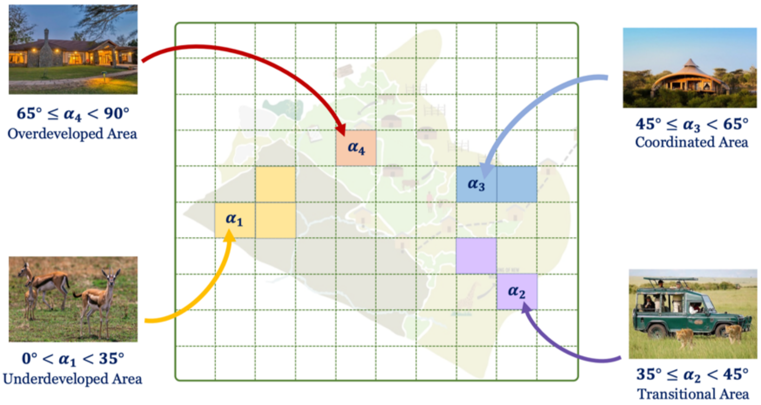
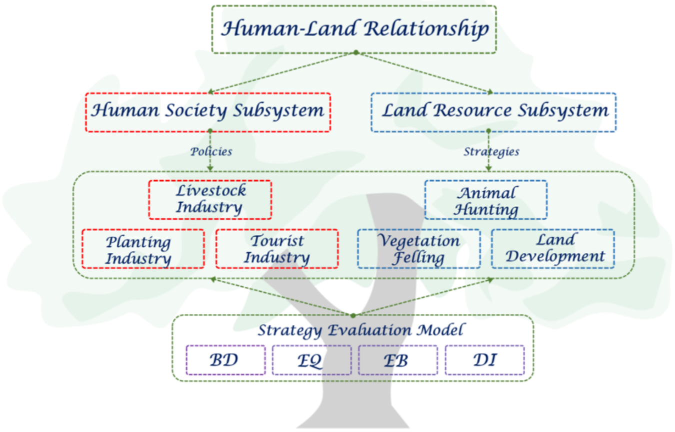
Student ID (MU/FZU): 20122161 / 832002117

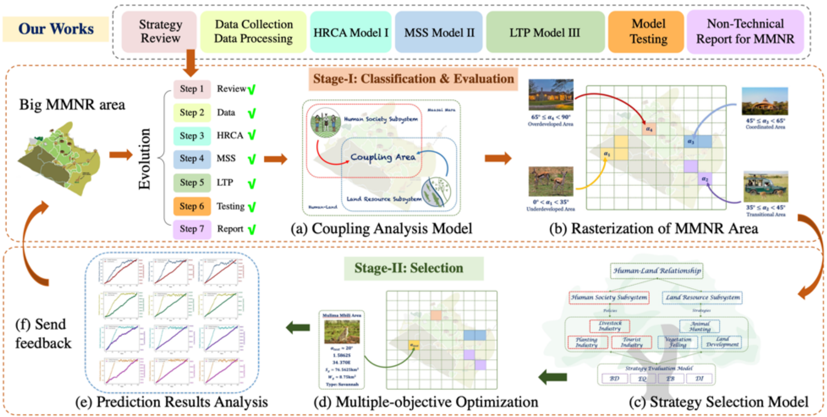
Module Code / Name: EE382 (EE388FZ)

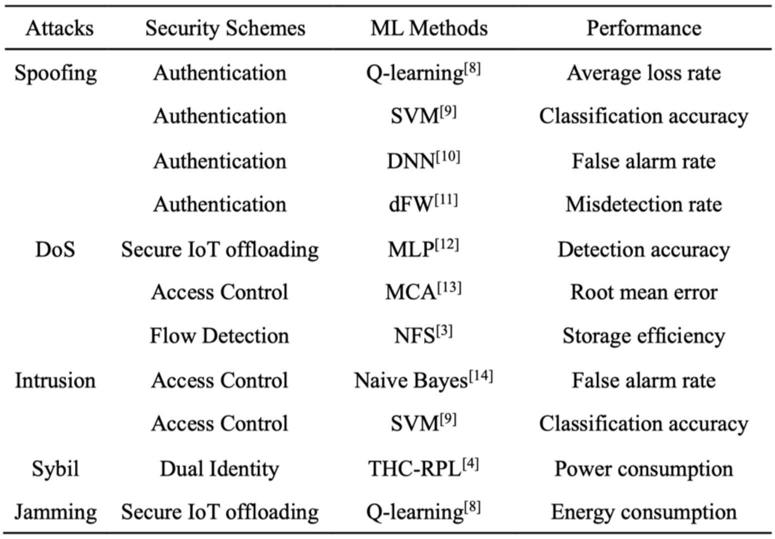
Programme: Robotics and Intelligent Devices

Company & Address: Fujian Huading Intelligent Manufacturing Technology Co. LTD
Building 44, Zone C, Fuzhou Software Park, Software Avenue
Fuzhou City, Fujian Province, China

THE INDUSTRIAL PLACEMENT WORK DIARY MUST BE UPDATED WEEKLY BY THE END OF EACH WEEK IN MOODLE AS A SINGLE DOCUMENT.

Week	Day / Date	Activity / Portfolio
1	Mon 20/2/2023	Supervised by Prof. Zhezhuang Xu, I participate in the Mathematical Contest in Modeling (**MCM** 2023), which is a highly acclaimed contests for international undergraduates. During this competition, I cooperate with Yufei Wu and Wenxuan Luo. Our team once won the First Prize in China Undergraduate Mathematical Contest in Modeling. Besides, the MCM contest lasts for five days. And in the first day, we decided to choose the Topic B (Reimagine Maasai Mara), which is a **discrete problem**. Based on the problem background, we have comprehensively reviewed the related research and resource. Finally, I finished to write the part of **Introduction** and **Literature Review** of our competition paper. For details, our final paper can be accessed here: https://caihanlin.com/mypaper/modeling/202302COMAP.pdf
	Tue 21/2/2023	The MCM Problem B required us to resolve **four objective,** hence we have made a suitable time schedule in advance. On Tuesday, we build our model I (**Human-land Relationship** Coupling Analysis) to address the objective of human-land interaction analysis and area classification. The following Figure 1 illustrates the classification of the Maasai Mara area.  $65° \leq \alpha_4 < 90°$ Overdeveloped Area α_4 $45° \leq \alpha_3 < 65°$ Coordinated Area α_3 α_1 α_2 $0° < \alpha_1 < 35°$ Underdeveloped Area $35° \leq \alpha_2 < 45°$ Transitional Area Finally, we suggest six specific strategies for practical implementation.
	Wed 22/2/2023	On Day 3, we firstly propose a four-layer strategy evaluation model based on the **AHP method** to evaluate and rank the policies. The following Figure 2 shows the structure of the strategy evaluation model.  Human-Land Relationship Human Society Subsystem Land Resource Subsystem Policies Strategies Livestock Industry Animal Hunting Planting Industry Tourist Industry Vegetation Felling Land Development Strategy Evaluation Model BD EQ EB DI Second, we present the **multi-objective optimization** model to quantify the economic and ecological impact of the optimal combination of the strategies and policies. Finally, we utilized our proposed model to test a representative grid in the MMNR area. The simulation results verify the effectiveness and rationality of our model.

	Thu ***23/2/2023***	On Day 4, considering to resolve the Objective III, we design a long-term trend prediction model to project and assess the long-term ecological and economic situation in the MMNR area based on the optimal management strategies. **First**, we quantify the reflection of trend using the change of animal numbers and resident economic incomes. **Second**, we propose the specific expression of the Logistic Equation, maximum environmental capacity and resident economic incomes. **Third**, based on mathematical definitions, we design a Python program for our trend prediction models. **Finally**, we obtain and present twelve 100-year prediction results referring to twelve different sets of parameter configurations.
	Fri ***24/2/2023***	On the last day, we firstly design a two-page **non-technical report** for the Kenyan Tourism and Wildlife Committee (Objective IV). After finishing the first version of main body, I begin to write the conclusion and summary. Also, I draw a ‘fantastic’ figure to show the overview of our works, as follows:  Our Works Strategy Review Data Collection Data Processing HRCA Model I MSS Model II LTP Model III Model Testing Non-Technical Report for MMNR Stage-I: Classification & Evaluation Big MMNR area Evolution Step 1 Review ✓ Step 2 Data ✓ Step 3 HRCA ✓ Step 4 MSS ✓ Step 5 LTP ✓ Step 6 Testing ✓ Step 7 Report ✓ Human Society Subsystem Coupling Area Land Resource Subsystem Maasai Mara Human-Land 65° ≤ α₄ < 90° Overdeveloped Area 45° ≤ α₃ < 65° Coordinated Area 0° < α₁ < 35° Underdeveloped Area 35° ≤ α₂ < 45° Transitional Area Stage-II: Selection Human-Land Relationship Human Society Subsystem Land Resource Subsystem Policies Strategies Livestock Industry Animal Hunting Planting Industry Tourist Industry Vegetation Felling Land Development Strategy Evaluation Model BD TQ EB DI Mulima Mbili Area Type: Savannah (f) Send feedback (a) Coupling Analysis Model (b) Rasterization of MMNR Area (c) Strategy Selection Model (d) Multiple-objective Optimization (e) Prediction Results Analysis Ultimately, I review and polish the whole paper under the guidance of Prof. Xu, and we upload the final manuscript to the COMAP committee.
Summary: In brief, I take part in the 2023 Mathematical Contest in Modeling (MCM) in the first week. Advised by Prof. Xu, we present a paper called Reshape the Crowning Glory of Maasai Mara. Since the MCM contest only lasts for **five days**, it is impossible for our paper to be comprehensive and perfect. Therefore, I plan to further improve and refine our paper in the following weeks. Finally, our final paper can be accessed here: https://caihanlin.com/mypaper/modeling/202302COMAP.pdf		

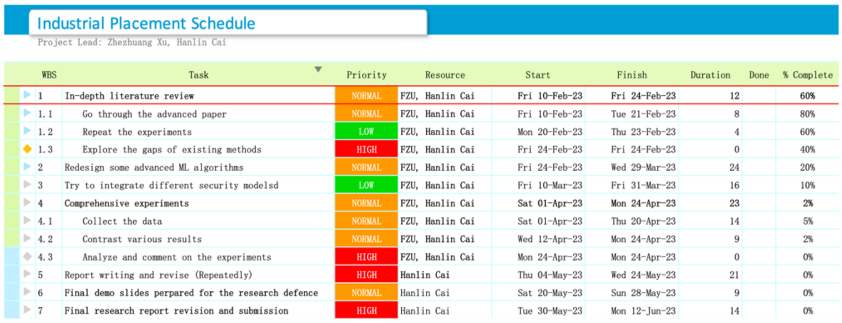
<table>
<tr><th>Week</th><th>Day / Date</th><th>Activity / Portfolio</th></tr>
<tr><td rowspan="3">2</td><td>Mon
27/2/2023</td><td>In this week, we are required to finish the Intern Proposal of Industrial Placement (IP). Therefore, firstly I have a face-to-face talk with my supervisor. And based on the suggestions given by Prof. Xu, I choose the research topic of Exploring Multiple IoT Security Attacks with Machine Learning Based Schemes. Then, I take the most of time to read the related paper. Also, I begin to write the <b>Literature Review</b> of my IP proposal. For details, my IP proposal can be accessed here:
https://caihanlin.com/mypaper/IP/Proposal.pdf</td></tr>
<tr><td>Tue
28/2/2023</td><td>In Day 2, I continue to review the existing research works and write the part of Related Literature. Based on the reference, I obtain that many papers have proposed various ML-based methods to prevent specific IoT attacks and improve IoT security. For instance, the Q-learning schemes proposed by Xiao, et al. performed well in the face of both spoofing and jamming attacks. While the SVM schemes proposed by Ozay, et al. could effectively identify and defend against intrusion and spoofing attacks. And the following Table shows the <b>Summary</b> of my reviewing works.

<table>
<tr><th>Attacks</th><th>Security Schemes</th><th>ML Methods</th><th>Performance</th></tr>
<tr><td>Spoofing</td><td>Authentication</td><td>Q-learning[8]</td><td>Average loss rate</td></tr>
<tr><td></td><td>Authentication</td><td>SVM[9]</td><td>Classification accuracy</td></tr>
<tr><td></td><td>Authentication</td><td>DNN[10]</td><td>False alarm rate</td></tr>
<tr><td></td><td>Authentication</td><td>dFW[11]</td><td>Misdetection rate</td></tr>
<tr><td>DoS</td><td>Secure IoT offloading</td><td>MLP[12]</td><td>Detection accuracy</td></tr>
<tr><td></td><td>Access Control</td><td>MCA[13]</td><td>Root mean error</td></tr>
<tr><td></td><td>Flow Detection</td><td>NFS[3]</td><td>Storage efficiency</td></tr>
<tr><td>Intrusion</td><td>Access Control</td><td>Naive Bayes[14]</td><td>False alarm rate</td></tr>
<tr><td></td><td>Access Control</td><td>SVM[9]</td><td>Classification accuracy</td></tr>
<tr><td>Sybil</td><td>Dual Identity</td><td>THC-RPL[4]</td><td>Power consumption</td></tr>
<tr><td>Jamming</td><td>Secure IoT offloading</td><td>Q-learning[8]</td><td>Energy consumption</td></tr>
</table>
Note that my literature exploring is heavily based on the Review Paper published by Xiao, et al in 2020.</td></tr>
<tr><td>Wed
1/3/2022</td><td>As for Day 3, today I have already finished the part of Related Literature, and begin to analyze and compare the methodologies and experimental results between different research. Besides, I write the part of <b>Gap in Existing Knowledge</b> to show the possible improvements we can conduct in the near future.

Through in-deep literature review, I find that most of the solutions proposed by existing studies can address specific security attacks but cannot to define more patterns for detecting <b>dynamic multiple attacks</b>. Therefore, it is feasible for us to design a hybrid defense scheme to resolve this challenge. And it may lead to the potential publication opportunities and positive contributions.</td></tr>
</table>

Thu

2/3/2023

In Day 4, after presenting the gap in existing literature, today I begin to organize the workflow and develop the research schedule. To advance my research scientifically and effectively, I developed a specific research schedule as follows:



Industrial Placement Schedule

Project Lead: Zhezhuang Xu, Hanlin Cai

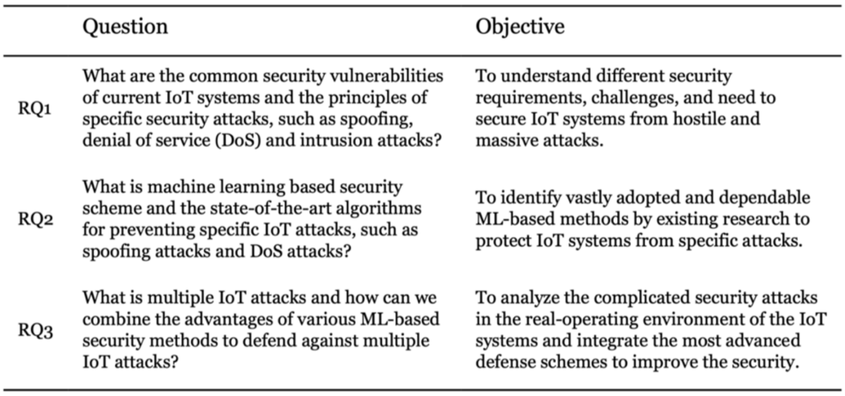
WBS	Task	Priority	Resource	Start	Finish	Duration	Done	% Complete
1	In-depth literature review	NORMAL	FZU, Hanlin Cai	Fri 10-Feb-23	Fri 24-Feb-23	12		60%
1.1	Go through the advanced paper	NORMAL	FZU, Hanlin Cai	Fri 10-Feb-23	Tue 21-Feb-23	8		80%
1.2	Repeat the experiments	LOW	FZU, Hanlin Cai	Mon 20-Feb-23	Thu 23-Feb-23	4		60%
1.3	Explore the gaps of existing methods	HIGH	FZU, Hanlin Cai	Fri 24-Feb-23	Fri 24-Feb-23	0		40%
2	Redesign some advanced ML algorithms	NORMAL	FZU, Hanlin Cai	Fri 24-Feb-23	Wed 29-Mar-23	24		20%
3	Try to integrate different security modelsd	LOW	FZU, Hanlin Cai	Fri 10-Mar-23	Fri 31-Mar-23	16		10%
4	Comprehensive experiments	NORMAL	FZU, Hanlin Cai	Sat 01-Apr-23	Mon 24-Apr-23	23		2%
4.1	Collect the data	NORMAL	FZU, Hanlin Cai	Sat 01-Apr-23	Thu 20-Apr-23	14		5%
4.2	Contrast various results	NORMAL	FZU, Hanlin Cai	Wed 12-Apr-23	Mon 24-Apr-23	9		2%
4.3	Analyze and comment on the experiments	HIGH	FZU, Hanlin Cai	Mon 24-Apr-23	Mon 24-Apr-23	0		0%
5	Report writing and revise (Repeatedly)	HIGH	Hanlin Cai	Thu 04-May-23	Wed 24-May-23	21		0%
6	Final demo slides perpared for the research defence	NORMAL	Hanlin Cai	Sat 20-May-23	Sun 28-May-23	9		0%
7	Final research report revision and submission	HIGH	Hanlin Cai	Tue 30-May-23	Mon 12-Jun-23	14		0%

To be honest, I understand the research cannot always proceed smoothly as planned, so I will modify the schedule according to the actual situation.

Fri

3/3/2023

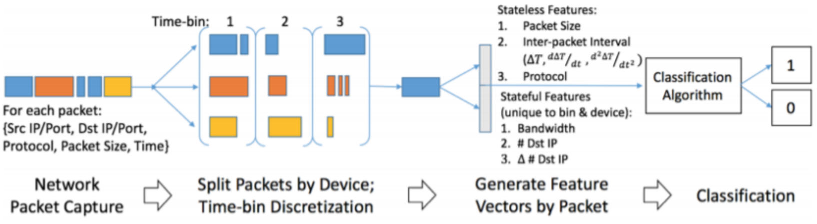
In Day 5, after analysing the research background and existing gaps, in order to address the research problem systematically, I have divided it into the specific research questions listed in the following table.



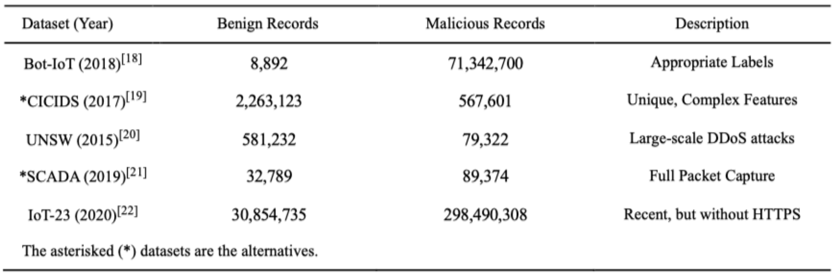
	Question	Objective
RQ1	What are the common security vulnerabilities of current IoT systems and the principles of specific security attacks, such as spoofing, denial of service (DoS) and intrusion attacks?	To understand different security requirements, challenges, and need to secure IoT systems from hostile and massive attacks.
RQ2	What is machine learning based security scheme and the state-of-the-art algorithms for preventing specific IoT attacks, such as spoofing attacks and DoS attacks?	To identify vastly adopted and dependable ML-based methods by existing research to protect IoT systems from specific attacks.
RQ3	What is multiple IoT attacks and how can we combine the advantages of various ML-based security methods to defend against multiple IoT attacks?	To analyze the complicated security attacks in the real-operating environment of the IoT systems and integrate the most advanced defense schemes to improve the security.

Besides, today I have an online meeting with my supervisor and seniors. Based on the work already done, they give me some useful advice and encourage me to carry out my plan.

Summary: In a word, this week I take most of my time to write the IP Proposal and read related research paper. An in-deep literature review has been conducted to analyse and compare the strength and weakness of existing works. Based on the review, I organize the research workflow and develop an experimental schedule. Finally, a research questions table is presented to illustrate the main challenge and objective in my future research. Again, my IP proposal can be accessed here: https://caihanlin.com/mypaper/IP/Proposal.pdf

<table>
<tr><th>Week</th><th>Day / Date</th><th>Activity / Portfolio</th></tr>
<tr><td rowspan="4">3</td><td>Mon
06/3/2023</td><td>In this week, I am going to further improve the Intern Proposal. I will focus on the part of research methodology, which includes testbed establishment, training datasets collection, learning model training and experimental results analysis. **In Day 1**, I have an online meeting with my supervisor and we detailly talk about the implementation of the experiments. Finally, I make a Week Plan to advance my works. For details, my latest IP proposal can be accessed here:
https://caihanlin.com/mypaper/IP/Proposal.pdf</td></tr>
<tr><td>Tue
07/3/2023</td><td>**In Day 2**, I firstly read some related paper and think about the merit and demerit of each methodology. According to an in-deep review paper by Xiao, et al, I select two different testbeds (FACT and Stratosphere Lab) and compare their similarity and difference. As shown in the following figure, the DoS detection pipeline is a suitable test framework for evaluating the accuracy and precision of various DoS classification algorithms, which would be useful to modify and improve the performance of classifiers.


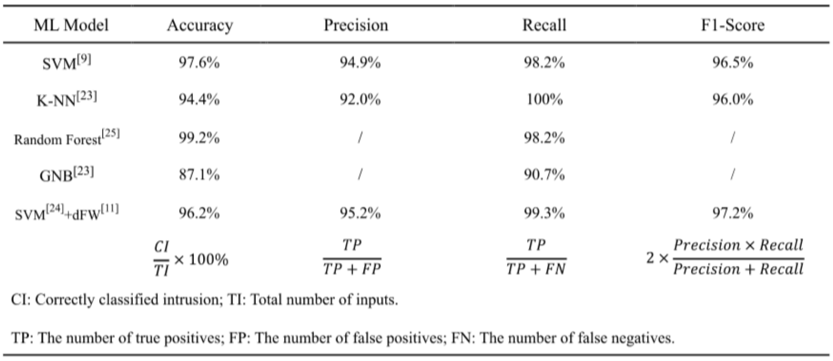
Time-bin: 1 2 3
For each packet:
{Src IP/Port, Dst IP/Port,
Protocol, Packet Size, Time}
Stateless Features:
1. Packet Size
2. Inter-packet Interval
($\Delta T, ^{d\Delta T}/_{dt}, ^{d^2\Delta T}/_{dt^2}$)
3. Protocol
Stateful Features
(unique to bin & device):
1. Bandwidth
2. # Dst IP
3. Δ # Dst IP
Classification Algorithm
1
0
Network Packet Capture
Split Packets by Device; Time-bin Discretization
Generate Feature Vectors by Packet
Classification

</td></tr>
<tr><td>Wed
08/3/2023</td><td>**As for Day 3**, to begin with, I explore some Github and Gitee project to find some suitable tools for launching IoT attacks. I find that the Ostinato, EXPLIoT and Kali Linux are suggested to generate benign and malicious activities simultaneously. Also, Argus Tool can be utilized for feature extraction and forensic analysis. Finally, I use the Ostinato and Argus to simulate an attack scenario and extract the traffic data.</td></tr>
<tr><td>Thu
09/3/2023</td><td>**In Day 4**, today I am required to investigate some existing IoT attack datasets and analyze the respective features. Since IoT-related companies rarely release the primary data for the reasons of confidentiality laws and user privacy Restrictions. I select the following five secondary datasets (Bot-IoT, IoT-23, etc.) through the research paper.

<table>
<tr><th>Dataset (Year)</th><th>Benign Records</th><th>Malicious Records</th><th>Description</th></tr>
<tr><td>Bot-IoT (2018)[18]</td><td>8,892</td><td>71,342,700</td><td>Appropriate Labels</td></tr>
<tr><td>*CICIDS (2017)[19]</td><td>2,263,123</td><td>567,601</td><td>Unique, Complex Features</td></tr>
<tr><td>UNSW (2015)[20]</td><td>581,232</td><td>79,322</td><td>Large-scale DDoS attacks</td></tr>
<tr><td>*SCADA (2019)[21]</td><td>32,789</td><td>89,374</td><td>Full Packet Capture</td></tr>
<tr><td>IoT-23 (2020)[22]</td><td>30,854,735</td><td>298,490,308</td><td>Recent, but without HTTPS</td></tr>
</table>
The asterisked (*) datasets are the alternatives.
Once the traffic activities datasets are collected and recorded, further work is to extract the traffic feature which can be used to distinguish malicious activities from benign or normal activities. And we will conduct this procedure in our experiments.</td></tr>
</table>

Fri

10/3/2023

In Day 5, after selecting suitable testbed template and training datasets, I am trying to do some **preliminary tests**. At first, I intend to utilize and integrate classical ML algorithms, such as Random Forest, Support Vector Machine and K-Nearest Neighbours. At present, I have carried out a preliminary test for these ML models and obtained some elementary results, as shown in the following Table.



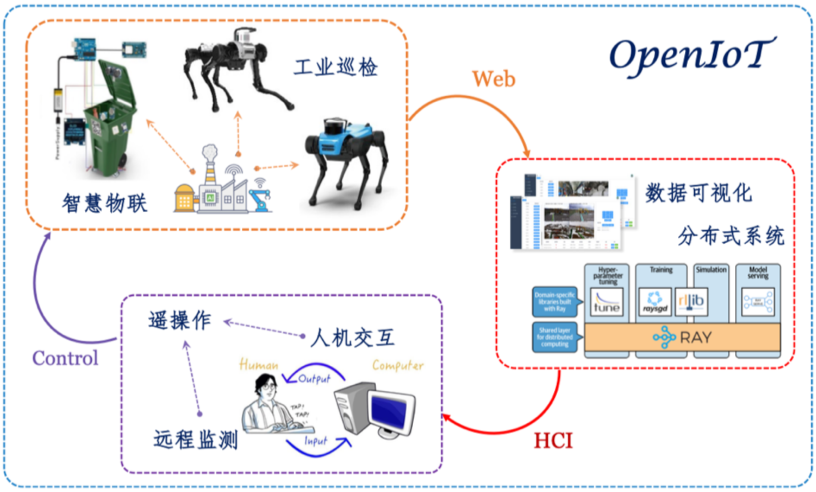
ML Model	Accuracy	Precision	Recall	F1-Score
SVM[9]	97.6%	94.9%	98.2%	96.5%
K-NN[23]	94.4%	92.0%	100%	96.0%
Random Forest[25]	99.2%	/	98.2%	/
GNB[23]	87.1%	/	90.7%	/
SVM[24]+dFW[11]	96.2%	95.2%	99.3%	97.2%
	$\frac{CI}{TI} \times 100\%$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FN}$	$2 \times \frac{Precision \times Recall}{Precision + Recall}$

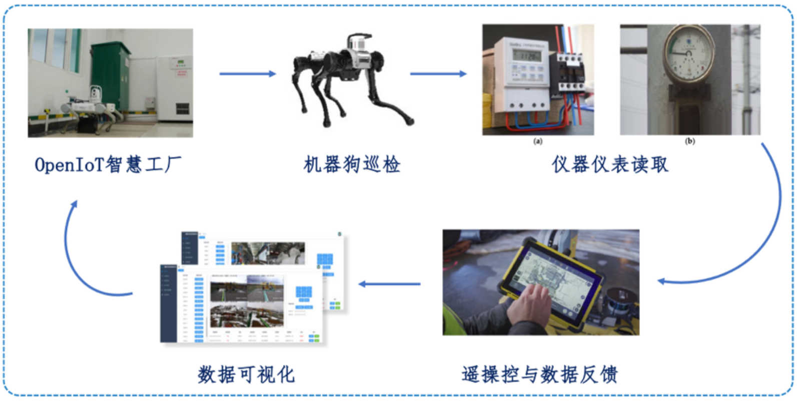
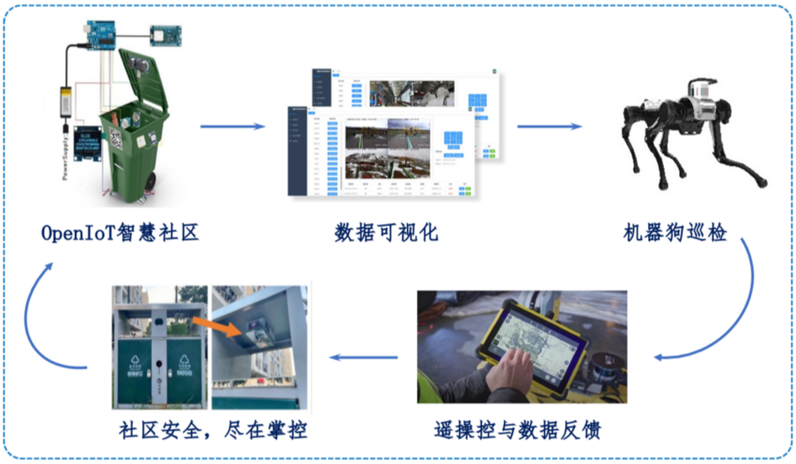
CI: Correctly classified intrusion; TI: Total number of inputs.

TP: The number of true positives; FP: The number of false positives; FN: The number of false negatives.

The Random Forest model got the best accuracy of 99.2%, and the combination of SVM and dFW models achieved the highest F1-Score of 97.2%, which represented the good balance between precision and recall.

Summary: In a nutshell, this week I take most of time to further improve my IP Proposal and design the part of **methodology and experiment.** Also, literature review has been conducted to analyse and compare the strength and weakness of related research. Referring to existing works, I select suitable IoT attack testbed and training dataset. Then, I do some preliminary tests and analyse the corresponding results. Finally, I have a face-to-face meeting with my tutor Mr. Jerry Qiu in the weekend and he give me some constructive advice. Again, my latest IP proposal can be accessed here: https://caihanlin.com/mypaper/IP/Proposal.pdf

Week	Day / Date	Activity / Portfolio
3	Mon 13/3/2023	In this week, I am required to help my supervisor to draft the grant application. Therefore, **in Day 1**, we have a face-to-face meeting to discuss the detailed tasks. My supervisor Prof. Xu gives me some related research paper and application forms. And I take the most of my time to read the materials and organize the related data. Moreover, because today is the lab open day, I show the sophomores around our laboratory.
	Tue 14/3/2023	**In Day 2**, I take most of my time to read the China Industrial Internet Security Situation Report (2021), which is the state-of-the-art report for the industrial internet of things. And the report shows that in 2020, the 360 Security Capability Center intercepted a total of 782 million virus samples and 76.871 billion virus infections, a decrease of 10.46% compared to the same period in 2019 in terms of virus infection numbers.
	Wed 15/3/2023	**As for Day 3**, I begin to write the part of research objective (Grant), based on our research topics of the IIoT patrol system, I focus on investigating the current situation and development mode of industrial device inspections. Also, today I draw a flowchart to illustrate our proposed systems, as shown in the following figure.  工业巡检 智慧物联 Web OpenIoT 数据可视化 分布式系统 Hyper-parameter tuning Training Simulation Model serving tune rllib RAY Control 遥操作 人机交互 Human Computer Output Input 远程监测 HCI
	Thu 16/3/2023	**In Day 4**, today I finish the part of the research objective, and start to write the part of research contents. And in this project, we focus on the advanced IoT inspection system, and the main research contents can be listed as follows: (1) Independently build a new IoT inspection system; (2) IoT Security Attack detection and prevention based on my system; (3) Using Kitex and Hertz distributed technology to upgrade the system. Additionally, in order to ensure the use, deployment and customization of the system by relevant enterprises and community customers, we will also design a set of user-friendly software prototypes and entry guides. Also, we will write detailed and thoughtful user manuals and technical documents, so as to reduce the difficulty coefficient of the secondary development of the system.

	Fri ***17/3/2023***	**In Day 5**, after finishing the part of the research contents, today I am required to visualize our application path. Our IoT inspection system is considered to utilized in the field of new promoting industrial region and smart living communities. Therefore, I draw the corresponding figures to illustrate our technical route as follows: ➢ Route 1: solution for smart IoT factories (Chinese version)  OpenIoT智慧工厂 机器狗巡检 (a) (b) 仪器仪表读取 数据可视化 遥操控与数据反馈 ➢ Route 2: solution for smart living communities (Chinese version)  OpenIoT智慧社区 数据可视化 机器狗巡检 社区安全，尽在掌控 遥操控与数据反馈
Summary: In a word, this week I focus on writing the grant application under the guidance of my supervisor and tutor. White report reading and related paper review have been conducted to analyse and compare the competitive products in the market. Besides, this week I take many of my time to visualize the application path of our project. Finally, I have a face-to-face meeting with my tutor Mr. Haoran Zhang in the weekend and he show me his upcoming plan and give me some useful advice.		