MINI PROJECT I

Analysis of Different types of Attacks on Industrial Automation & Control System

Write company name here

TEAM MEMBERS

Ankit Kumar	20bcs017
Aryan Tejiyan	20bcs021
Bazil Shaikh	20bcs029
Tanzeem	20bcs132

Content

- Introduction
- Objective
- Related Work
- Dataset Description
- ML Models
- DBN & Neural Network
- Implementation & Conclusion

INTRODUCTION

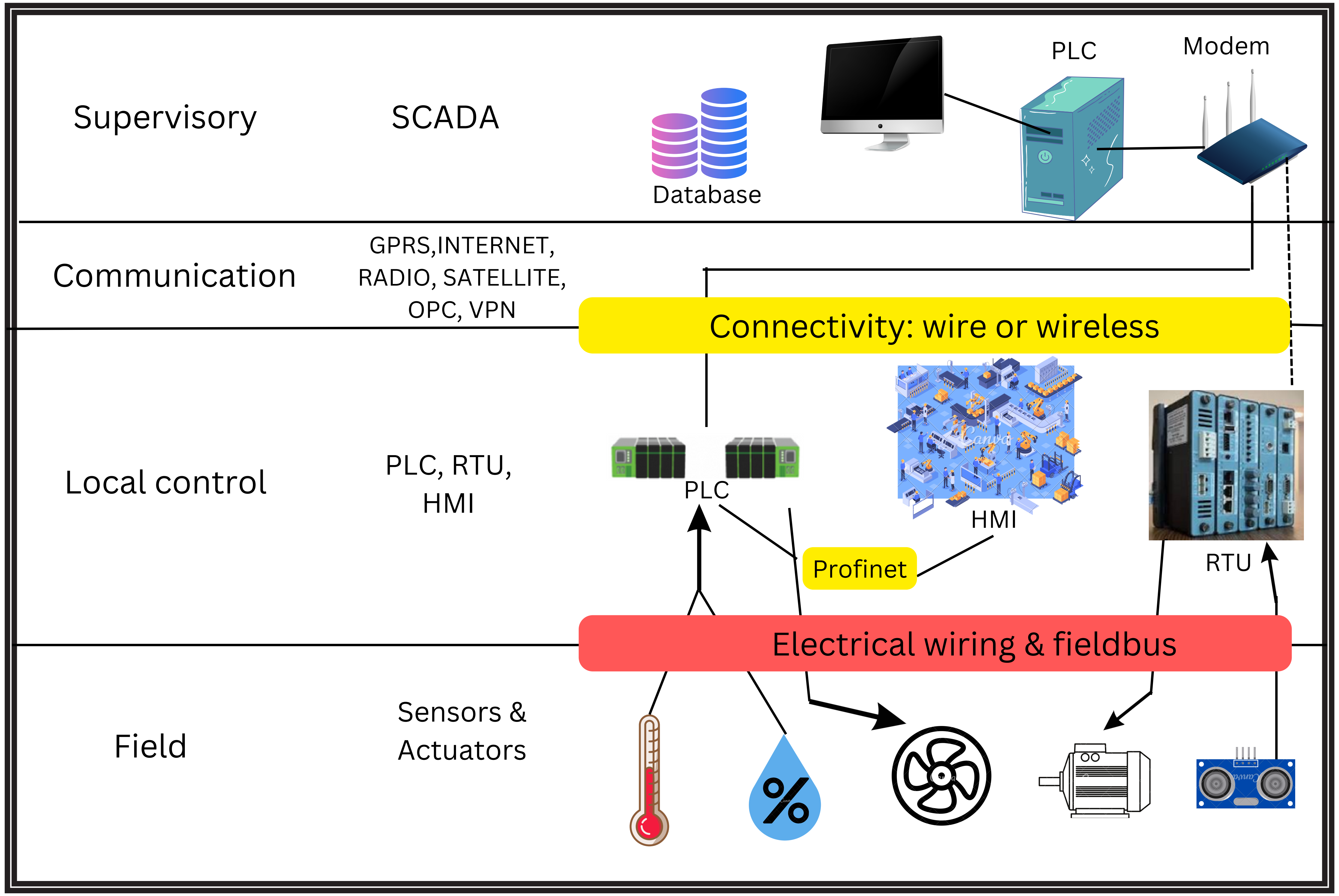
- IoT (Internet of Things) devices are increasingly being used to connect various physical objects to the Internet, allowing them to be remotely monitored, controlled, and managed.

- Industrial Automation Control Systems (IACS) are increasingly being targeted by cyber attackers seeking to disrupt or damage critical infrastructure.

- Supervisory control and data acquisition (SCADA) networks are widely used by industrial automation and control systems (IACS).

- SCADA (Supervisory Control and Data Acquisition) is a form of control system used to monitor and control industrial operations in Industrial Automation Control Systems (IACS).

GENERAL STRUCTURE



Layer structure	Technologies	Artitecture
Supervisory	SCADA	
Communication	GPRS,INTERNET, RADIO, SATELLITE, OPC, VPN	
Local control	PLC, RTU, HMI	
Field	Sensors & Actuators	

A general structure of SCADA-based IACS. This structure is composed of physical layer, cyber layer, and operation/corporate layer.

Objective

By recognising and minimising potential dangers, the goal of creating a neural network model for the study of assaults on industrial automation systems is to increase the cybersecurity of these crucial systems.

A neural network and DBN model can identify the assaults and detect unusual behaviour by analysing data from SCADA and ICS systems much more reliably than shallow machine learning methods. This allows for the detection of potential threats before they can do serious harm.

The neural network and DBN model can be used to identify deviations from the norm that can be signs of an attack after being trained on a large dataset of typical system behaviour.

RELATED WORK

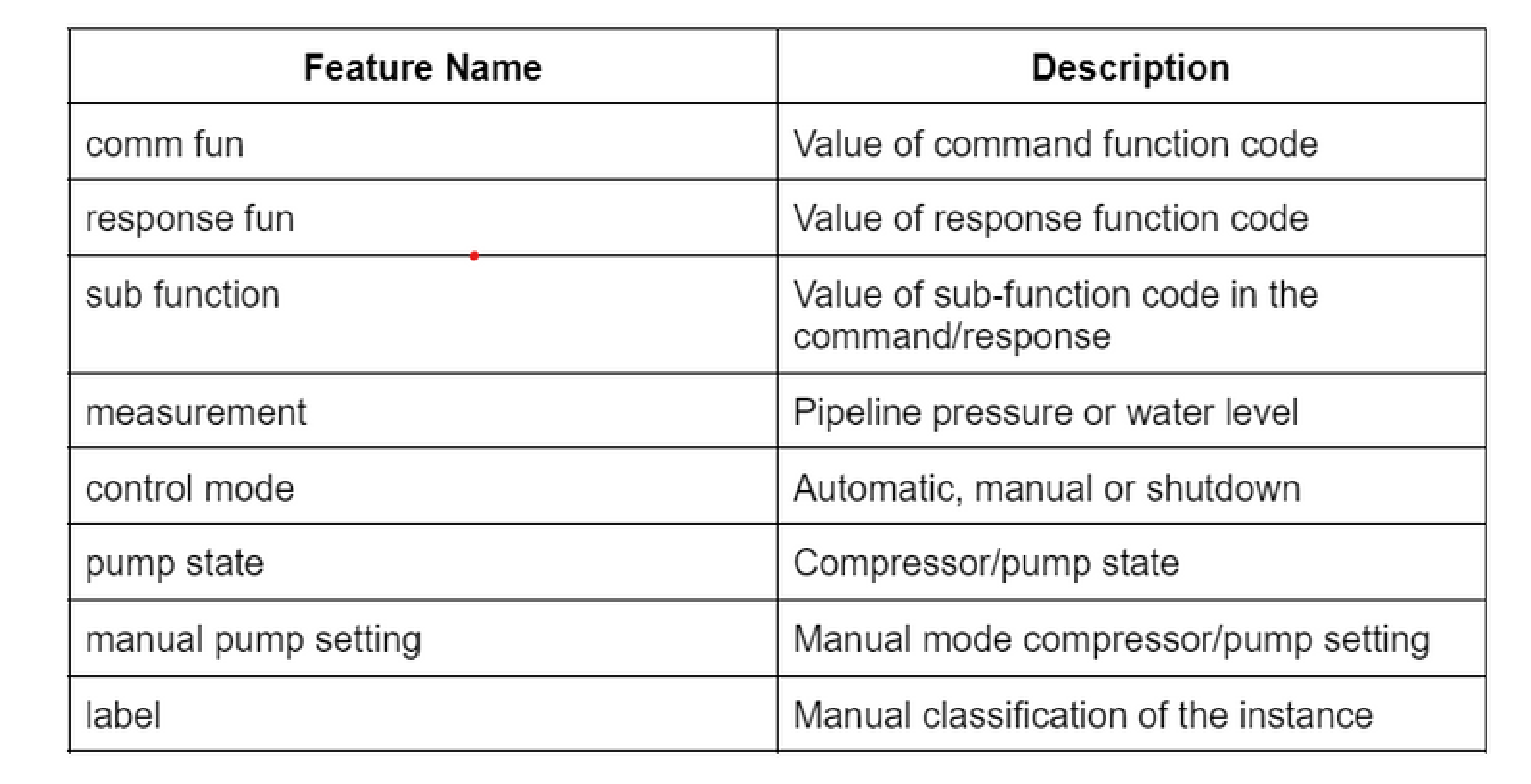
Author	Algorithm Used	Findings
Shahriari, M. et al	Multi-layer feedforward neural network	Effectively detect and classify different types of faults that may occur in an industrial robot.
Tao, Y. et al	Neural network-based soft sensor	robust and can handle variations in wastewater composition and operating conditions.
M. Masud et al	1. feature extraction module(CNN & LSTM) 2. classification module(MLP).	Deep learning-based IDSs have shown promise for detecting cyber attacks in SCADA networks.
M. Rashid et al	decision trees, support vector machines, artificial neural networks, deep learning models, and ensemble methods	Hybrid approaches that combine signature-based and anomaly-based methods can improve the detection accuracy and reduce the false positive rate.

DESCRIPTION OF ATTACKS

1. **NMRI attack:** Through the previously known information of network servers and devices, the attackers can inject random invalid information to the packet. But they cannot achieve the information of the underlying process being monitored and controlled.
2. **CMRI attack:** Here, the attacks have full information of SCADA network and devices, so that they can mask the actual state of the physical process and cause bad influence of the feedback control process.
3. **MSCI attack:** Through the actuators operation, the state of the physical system (e.g., ON/OFF) can be controlled. MSCI attack can change the state of the register, which may cause the operation of actuators incorrectly.
4. **MPCI attack:** In this type of attack, the attack can alter the set points parameters in field devices, e.g., PLC. After changing these parameters (e.g., PID controller’s parameters), the controller performance will be influenced or even does not work at all.
5. **MFCI attack:** If the attacks have the knowledge of the built-in protocol functions provided by the manufacturers for diagnostic purposes, they may abuse the functions.
6. **DoS attack:** DoS attack targets communications links and system programs to stop part of SCADA network. Sometimes, the attackers can change a packet and send it to the filed devices, which may result in crash for the operating system.
7. **Recon attack:** This attack collects SCADA system information, maps the network structure, as well as identifies device characteristics.

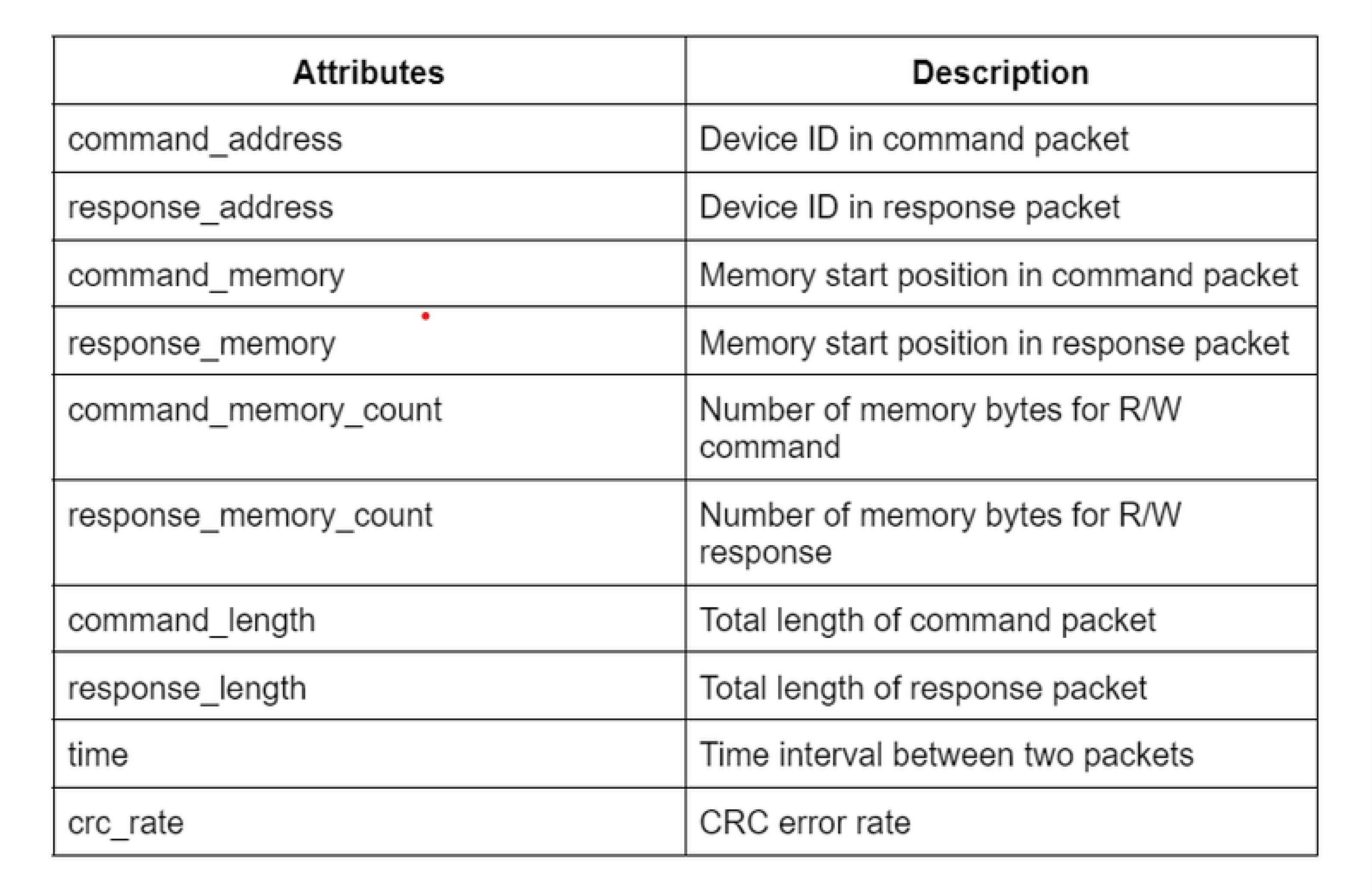
DATASET DESCRIPTION

- The dataset that is used in this project is the Gas pipeline dataset & Water storage tank system that we get from the google dataset website
- The data sets presented in this project include network traffic, process control and process measurement features from normal operations and attacks against the SCADA system.
- Two categories of features are present in the data sets: network traffic features and payload content features.
- Network traffic features describe the communications patterns in SCADA systems. Payload content features describe the current state of the SCADA system; they are useful for detecting attacks that cause devices (e.g., PLCs) to behave abnormally.



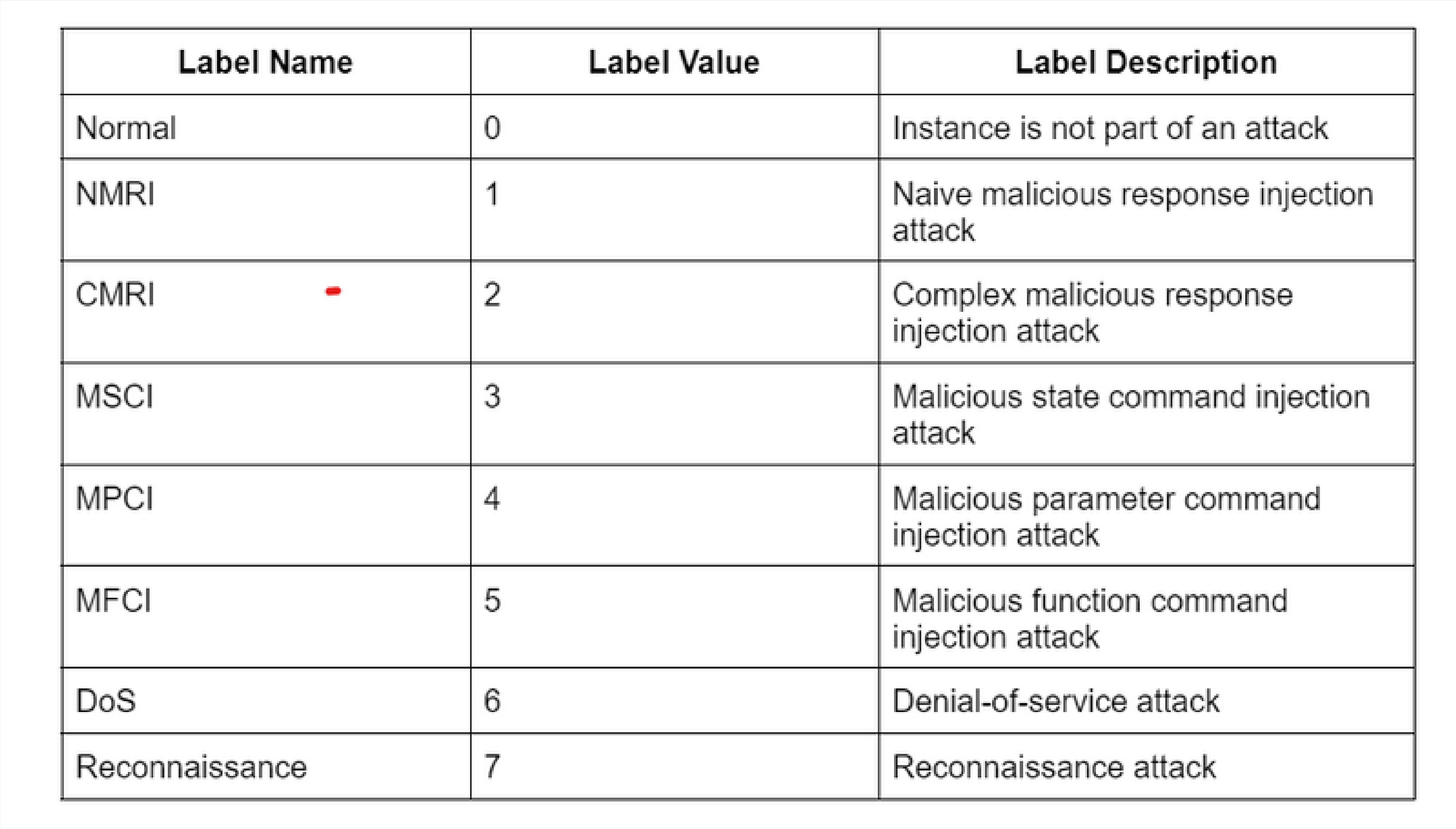
Feature Name	Description
comm fun	Value of command function code
response fun	Value of response function code
sub function	Value of sub-function code in the command/response
measurement	Pipeline pressure or water level
control mode	Automatic, manual or shutdown
pump state	Compressor/pump state
manual pump setting	Manual mode compressor/pump setting
label	Manual classification of the instance

Payload Feature



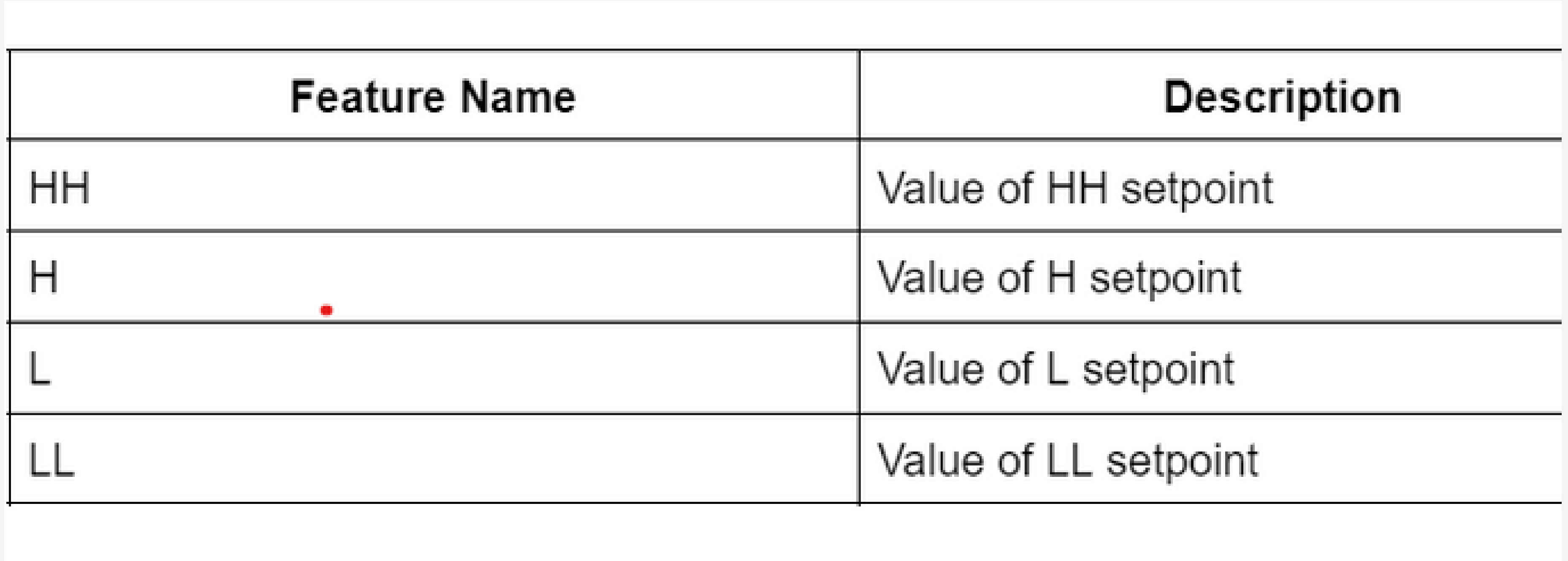
Attributes	Description
command_address	Device ID in command packet
response_address	Device ID in response packet
command_memory	Memory start position in command packet
response_memory	Memory start position in response packet
command_memory_count	Number of memory bytes for R/W command
response_memory_count	Number of memory bytes for R/W response
command_length	Total length of command packet
response_length	Total length of response packet
time	Time interval between two packets
crc_rate	CRC error rate

Network Traffic Feature



Label Name	Label Value	Label Description
Normal	0	Instance is not part of an attack
NMRI	1	Naive malicious response injection attack
CMRI	2	Complex malicious response injection attack
MSCI	3	Malicious state command injection attack
MPCI	4	Malicious parameter command injection attack
MFCI	5	Malicious function command injection attack
DoS	6	Denial-of-service attack
Reconnaissance	7	Reconnaissance attack

Instance Classification values



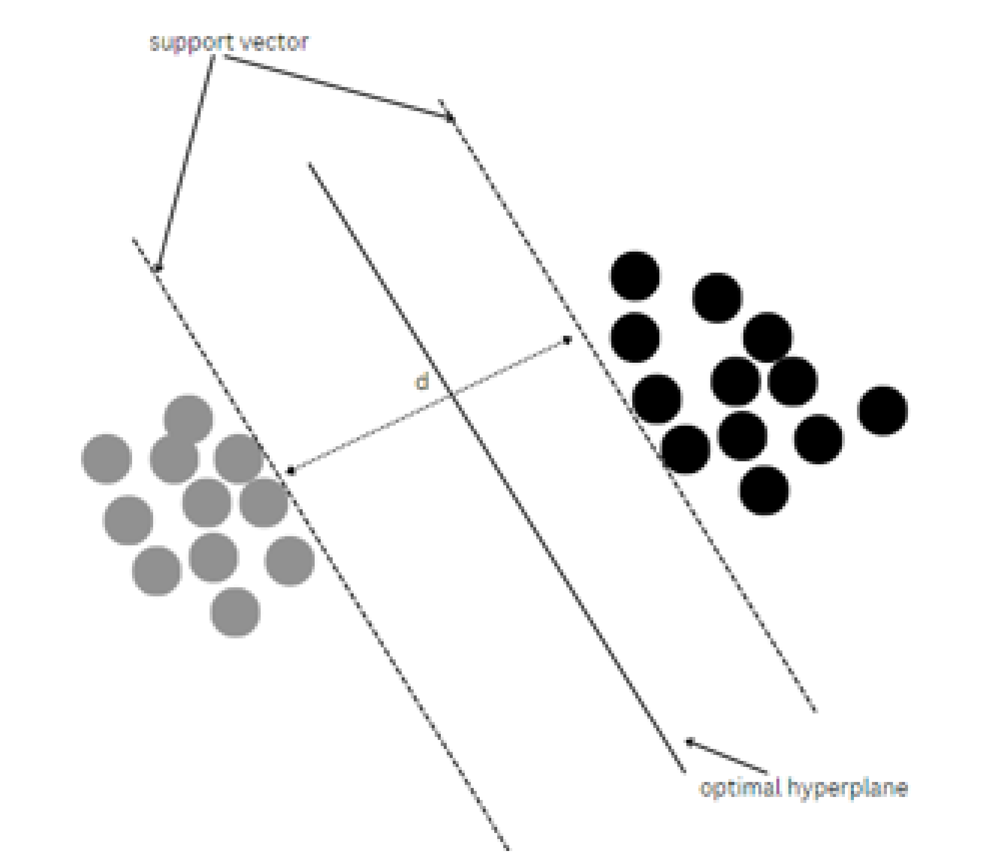
Feature Name	Description
HH	Value of HH setpoint
H	Value of H setpoint
L	Value of L setpoint
LL	Value of LL setpoint

Above Table shows the four attributes that are specific to the water storage tank system: HH, H, L and LL.

ML MODELS

SVM(Support vector machine)

- It is used for supervised machine learning problem where we try to find a hyperplane that best separates the two classes.
- “Support Vector Machine” (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges.



Decision tree

- Decision tree builds classification or regression models in the form of a tree structure.
- It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed.
- The decision tree algorithm makes decisions by recursively partitioning the input data based on the feature that provides the most information gain.

ML MODELS

Naive Bayes

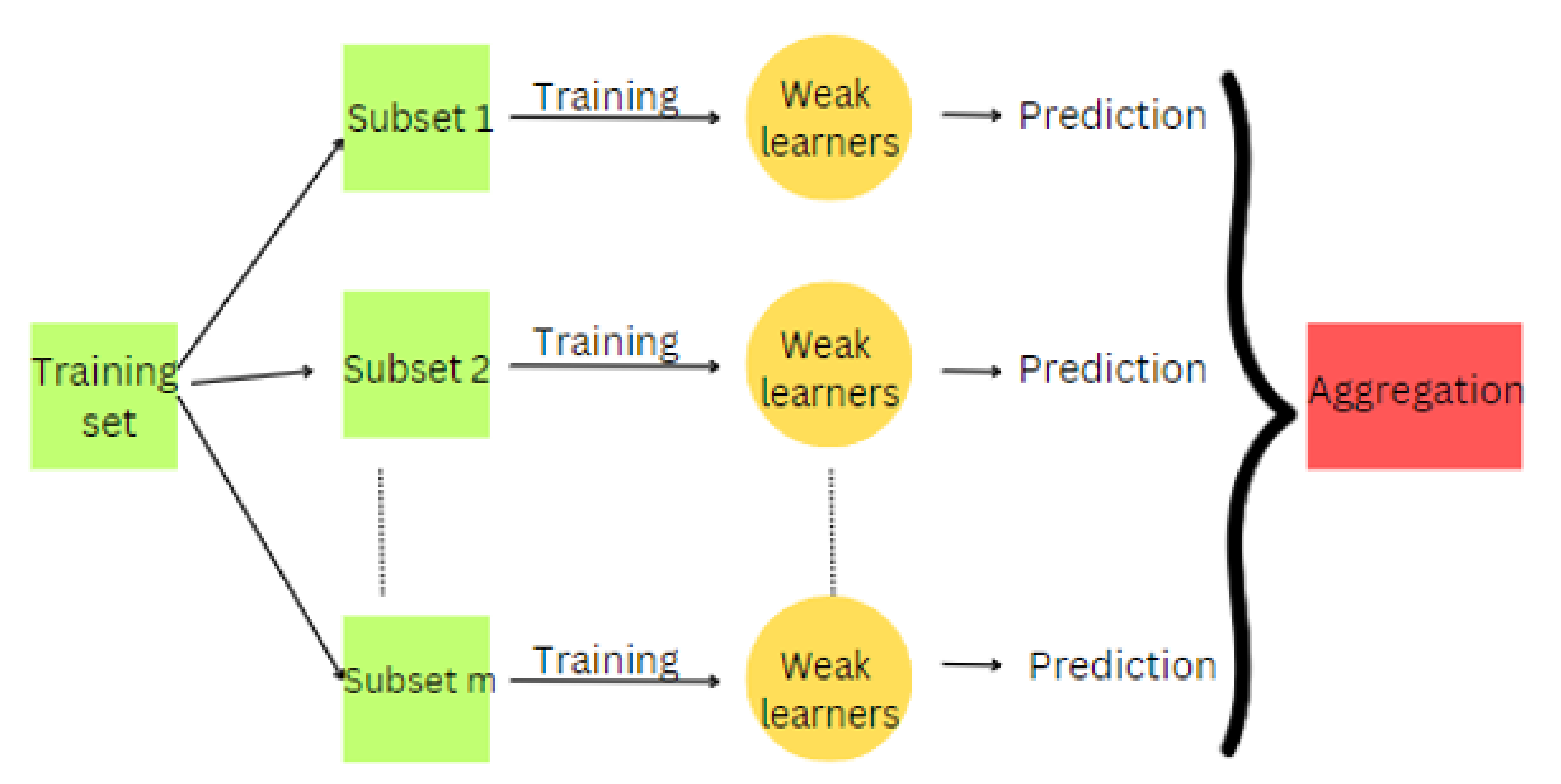
- Naive Bayes classification is a probabilistic algorithm that uses Bayes' theorem to calculate the probability of a data point belonging to a particular class.
- It assumes that the features of the data point are independent of each other, hence the term "naive".
- The Naive Bayes classification algorithm uses the following mathematical formula to predict the class of a new data point:
- **P(class | features) = (P(class) * P(features | class)) / P(features)**

ML MODELS

Ensemble Methods: Bagging and Boosting

- Approach in Ensemble learning:

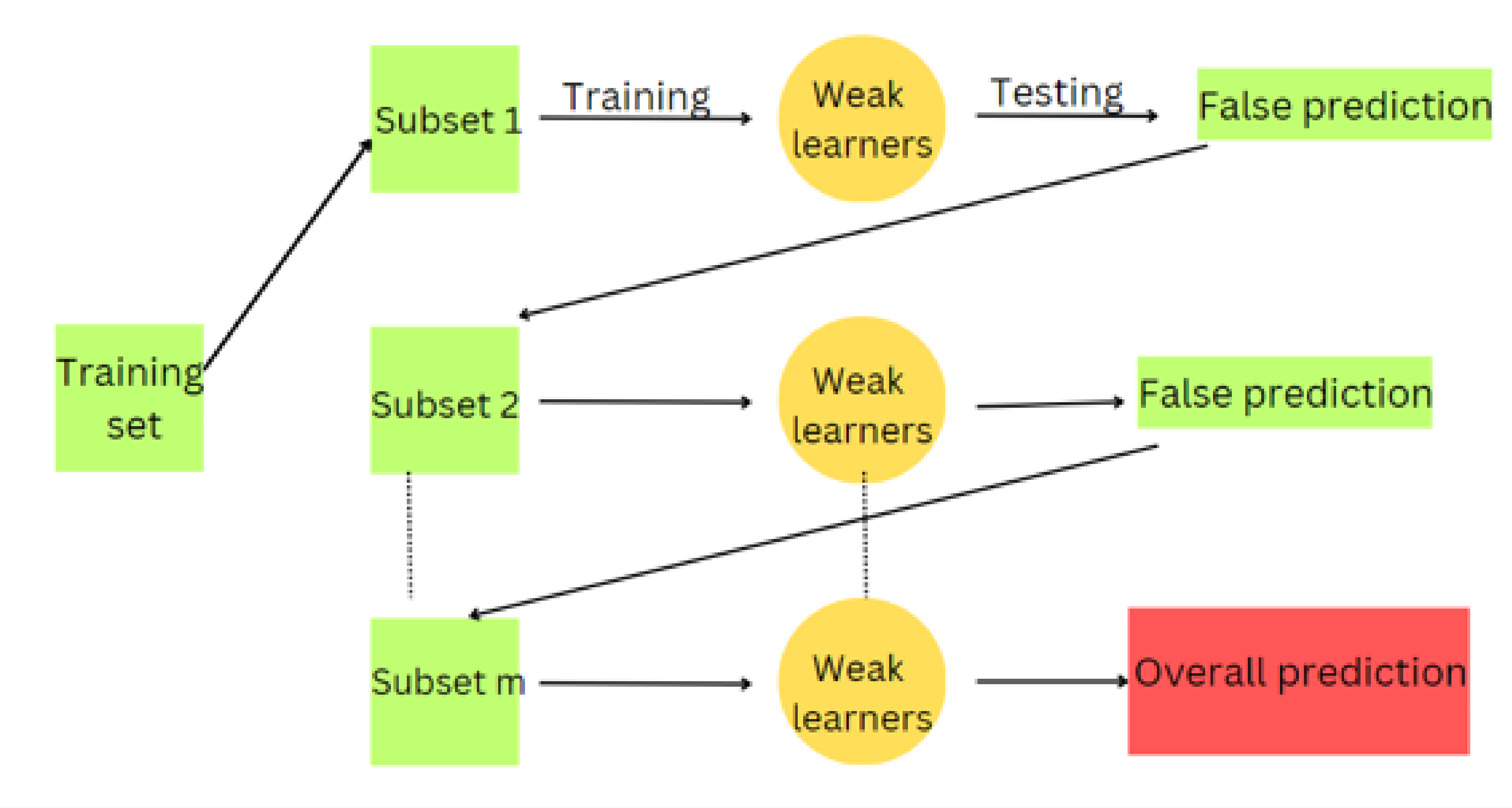
  Bagging: used to reduce the variance of weak learners



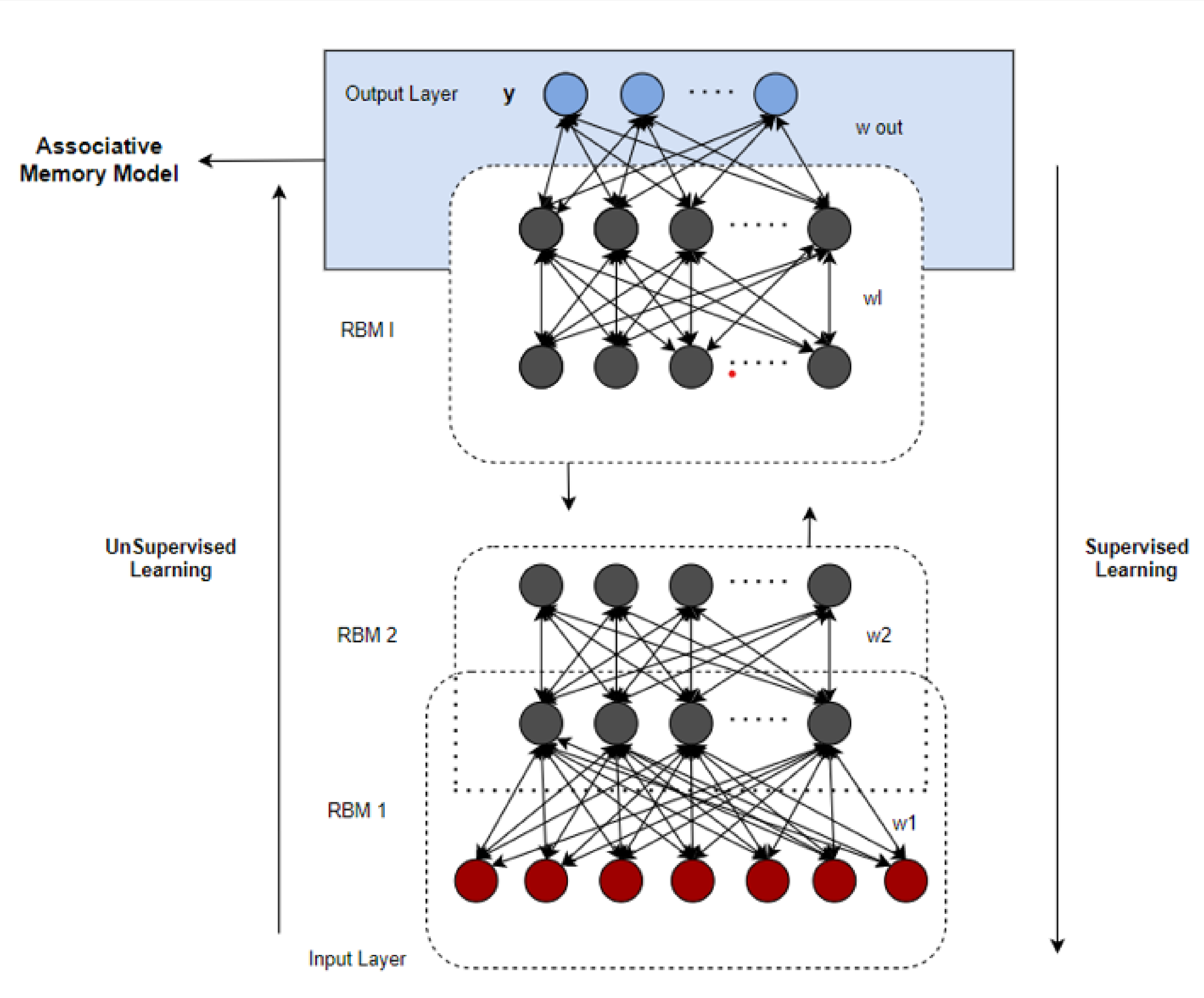
Ensemble Methods: Bagging and Boosting

- Approach in Ensemble learning:

  Boosting: used to reduce the bias of weak learners



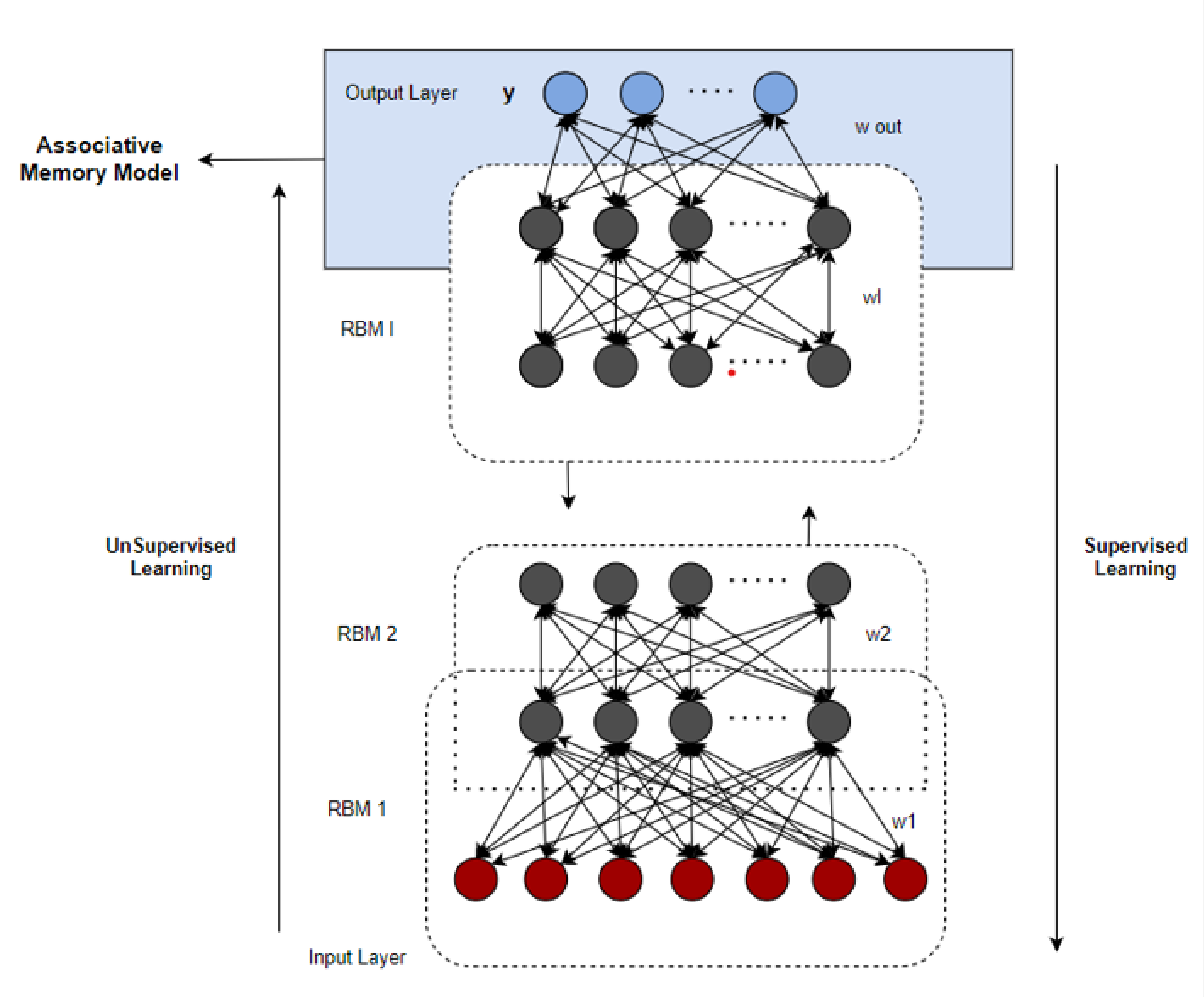
Deep Belief Networks



- Deep Belief Networks (DBNs) are a type of deep learning model that consists of multiple layers of Restricted Boltzmann Machines (RBMs).

- DBNs are able to learn complex hierarchical representations of data and have been successfully applied in various applications such as image and speech recognition, natural language processing and fault detection.

Deep Belief Networks



DBN: Algorithm

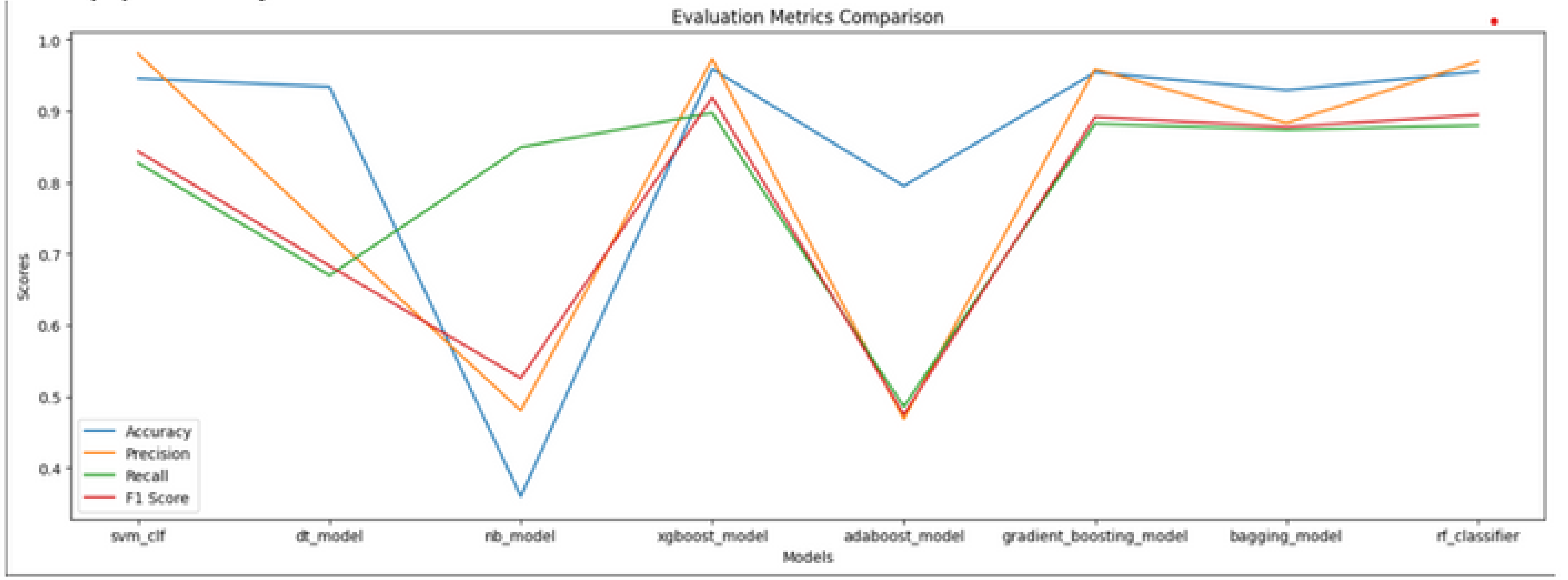
1. Input:
   a. Training dataset: X_train,Y_train
   b. Number of hidden layers e
   c. Number of nodes in each hidden layer H
   d. Learning rate α
   e. number of epochs E
2. Steps:
   a. Pre-train the DBN model by unsupervised learning using Contrastive Divergence Algorithm
   b. for l ← 1 to L do
   c. Train the RBM in layer l using X_train and the output of layer l-1 as input, with H nodes and α learning rate, for E epochs.
   d. Compute the output of layer l by passing the input through the trained RBM
   e. end for
   f. Concatenate the output of all hidden layers as the input of the softmax classifier.
   g. Train the softmax classifier using X_train and y_train as input, with softmax activation function and categorical cross-entropy loss function, for E epochs.

Output: Trained DBN Model

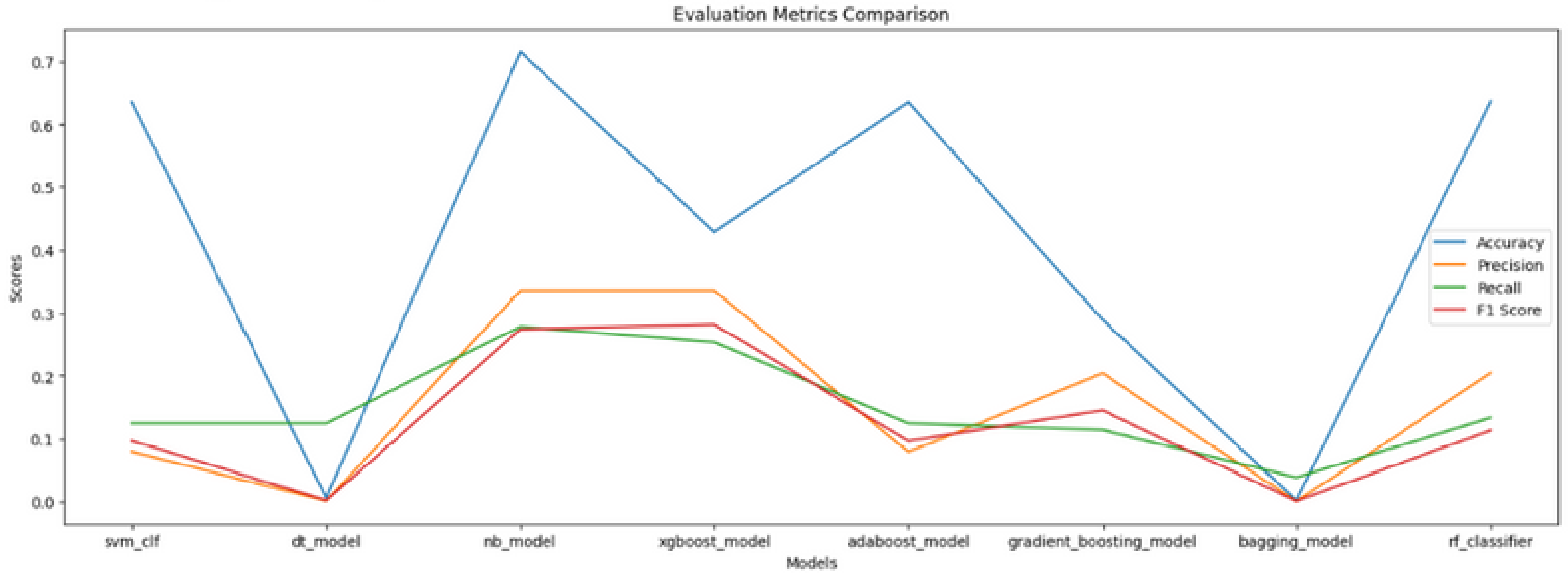
Results

Evaluation Metrics

Gas pipeline system



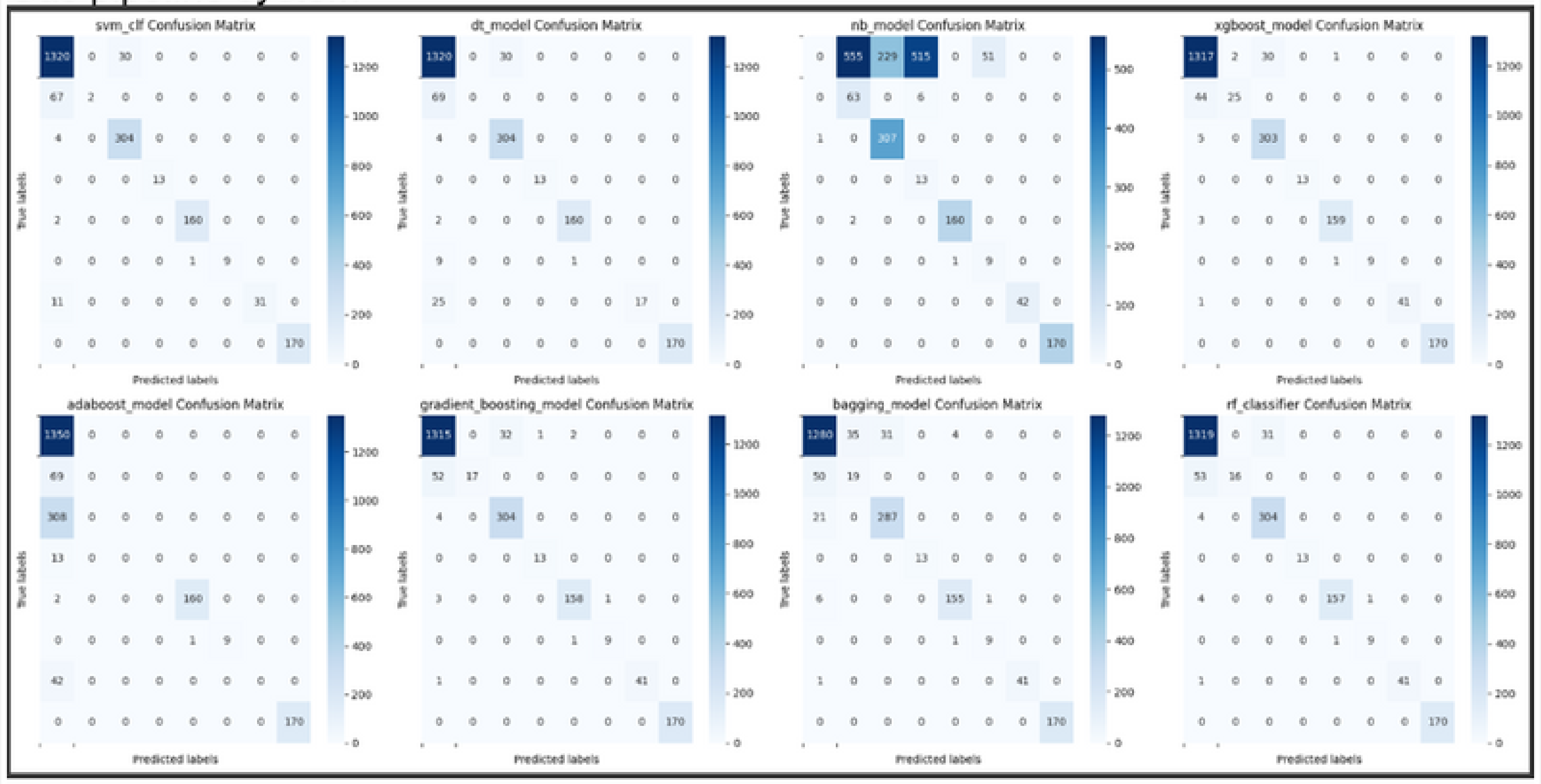
Water storage tank system



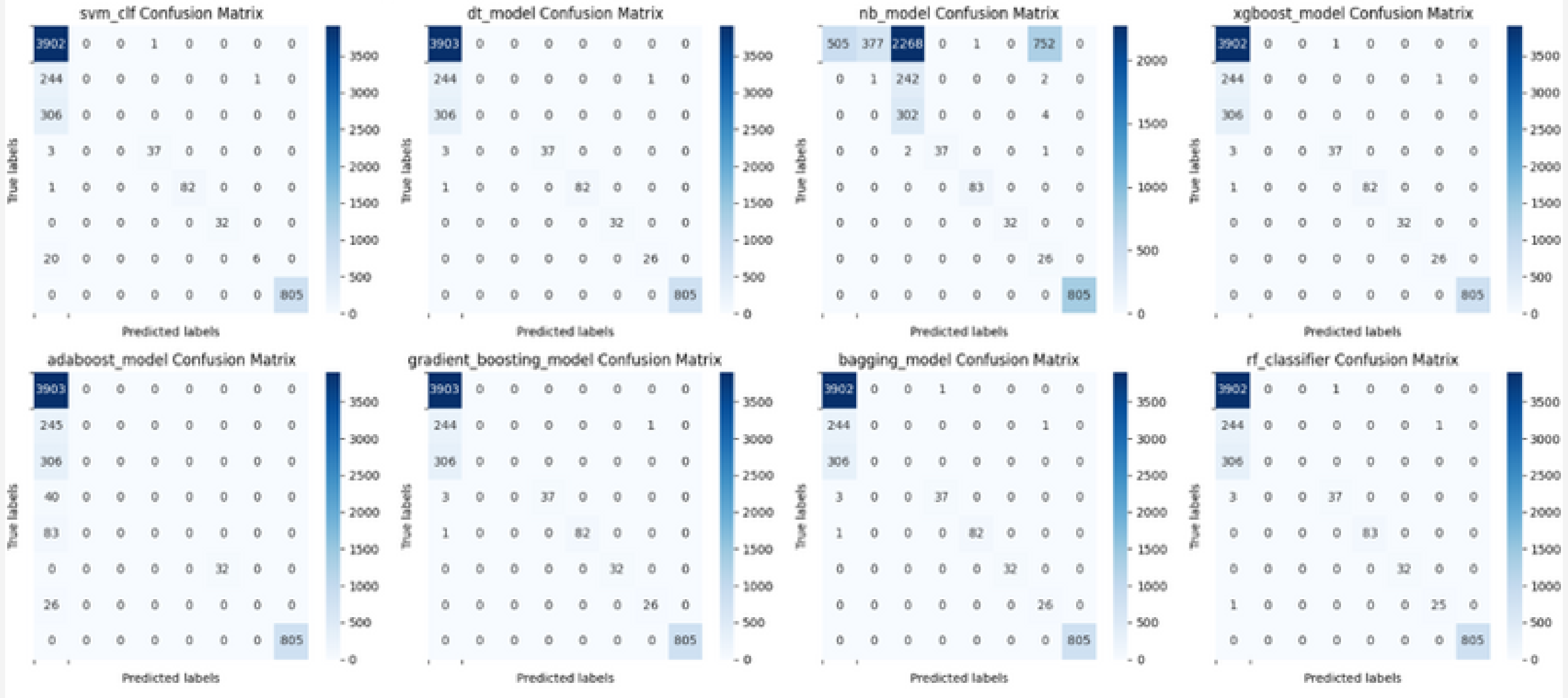
Results

Confusion Matrix

Gas pipeline system



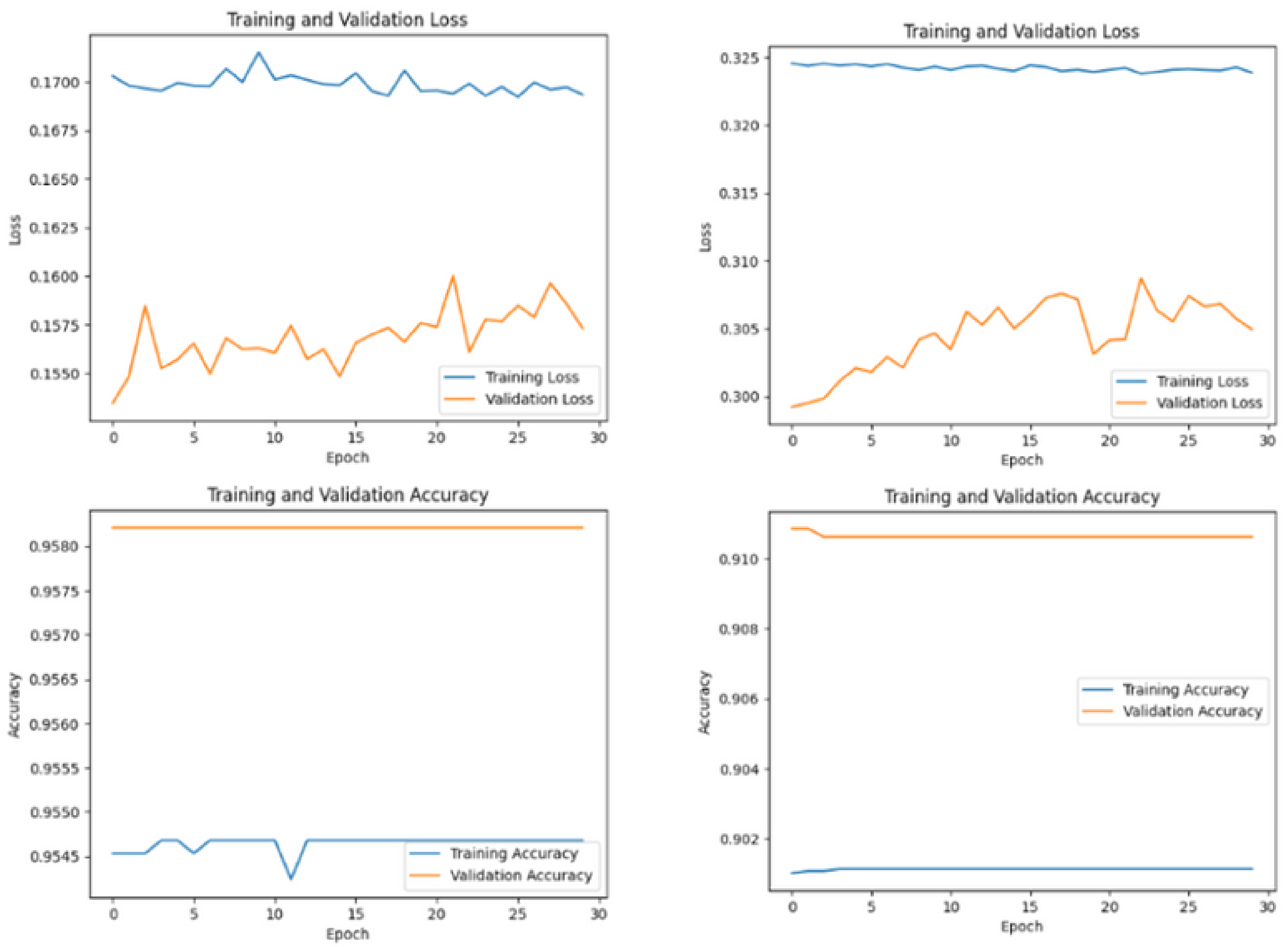
Water storage tank system



Results

The plotting of training and validation loss and accuracy over epochs using Matplotlib:

Neural network model



Gas pipeline system dataset

Water storage tank system dataset

The first plot shows the training loss and validation loss over each epoch, with the x-axis representing the epoch number and the y-axis representing the loss value.

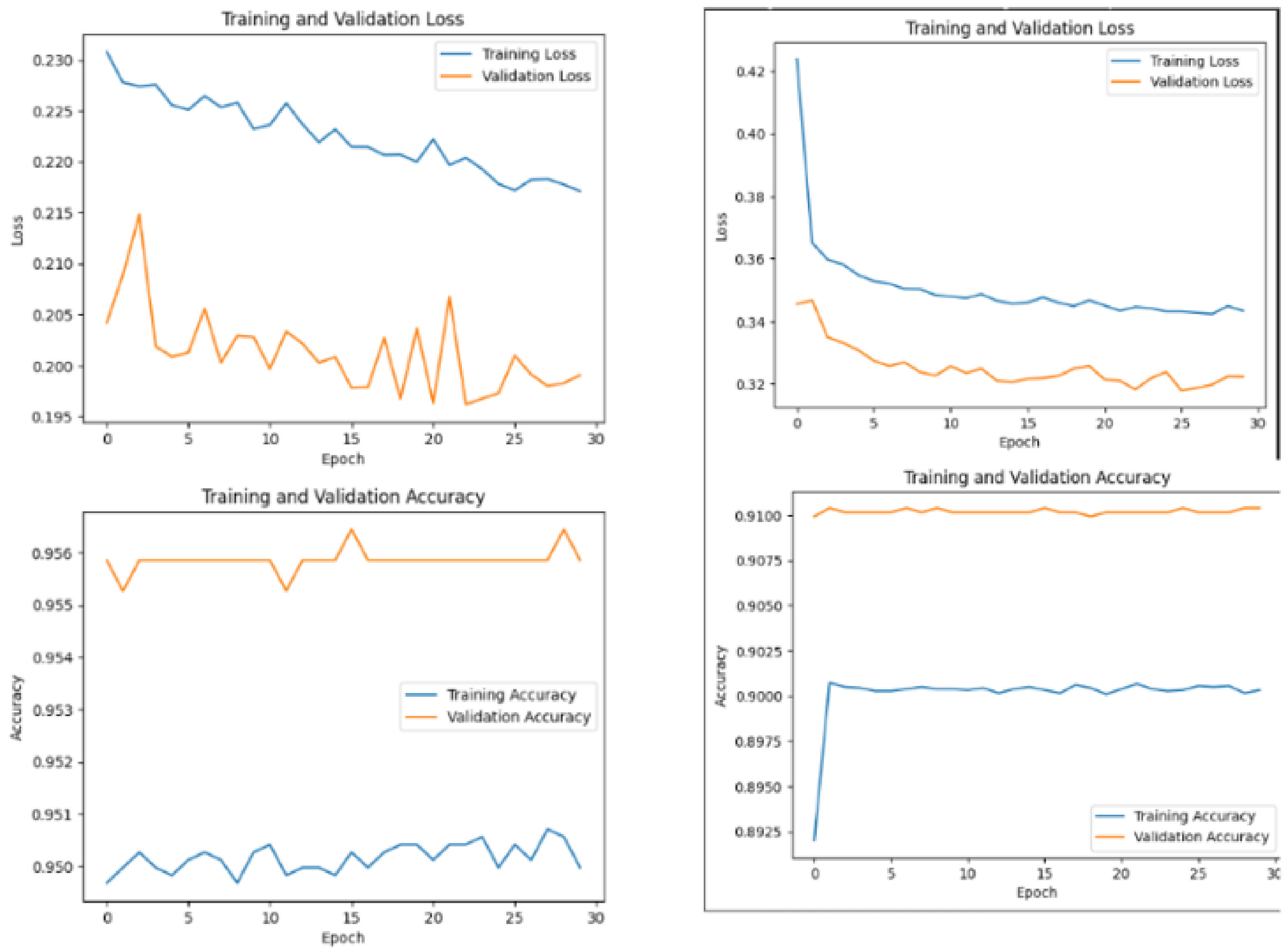
The second plot shows the training accuracy and validation accuracy over each epoch, with the x-axis representing the epoch number and the y-axis representing the accuracy value.

These plots are useful for visualizing the performance of the neural network over the course of training, and can be used to identify issues such as overfitting or underfitting

Results

The plotting of training and validation loss and accuracy over epochs using Matplotlib:

Deep Belief Network model



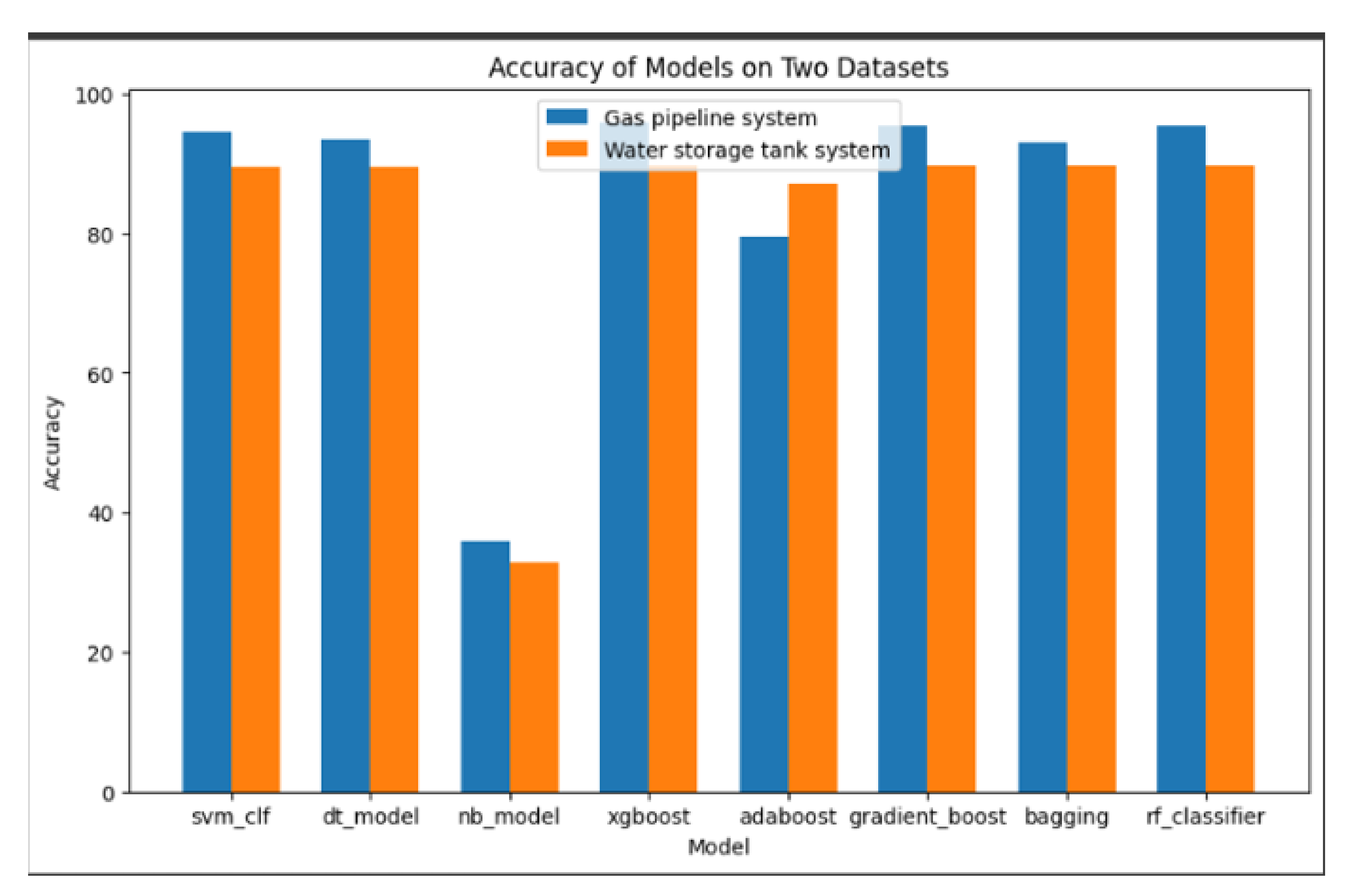
Gas pipeline system dataset

Water storage tank system dataset

These plots are useful for understanding how well the DBN model is learning from the training data and how well it generalizes to the validation data. If the training loss and accuracy are decreasing and increasing, respectively, while the validation loss and accuracy are also decreasing and increasing, respectively, this suggests that the model is learning from the training data and generalizing well to new data. If the validation loss and accuracy start to plateau or even increase while the training loss and accuracy continue to improve, this suggests that the model is overfitting to the training data and not generalizing well to new data.

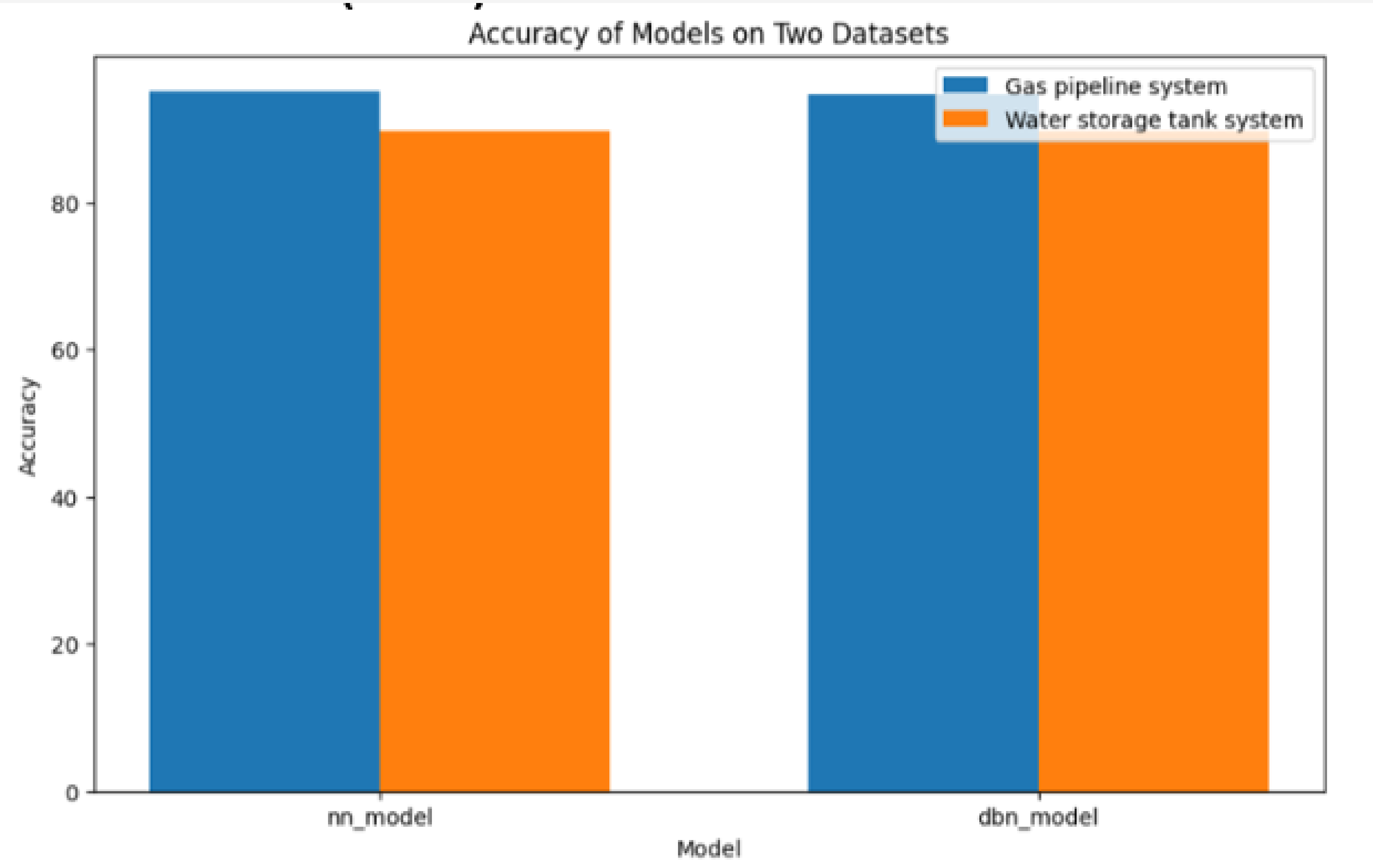
Results

Comparison of accuracies of different machine learning model that we implemented:



Results

Comparison of accuracies of Neural Network(NN) and Deep belief Neural Network(DBN):



CONCLUSION

- With the development of novel technologies and the integration of enormous number of IoT devices in IACS, the internet traffic is increased sharply, producing large-scale and multi-dimensional data, which makes the cyber-attacks scenarios more sophisticated.
- This project has demonstrated that deep learning techniques—including neural networks and deep belief network are more accurate in spotting cyberattacks than traditional machine learning methods.
- This is because deep learning models can analyze and learn from enormous amounts of complex data, which enables them to spot patterns and anomalies that conventional machine learning models might miss.
- Overall all the models perform well on gas pipeline dataset as compared to water storage tank dataset

THANK YOU