Chapter 7

FSRA: A Feature Selection and Rank Aggregation Framework for CPS Attack Classification

7.1 Introduction

Technological developments and innovations in various industries have brought about revolutionary changes at an unprecedented rate. These changes are significant to the day-to-day operations of human lives due to parameters such as convenience and ease of use [243]. Particularly, ease of use has largely impacted industrial sectors such as transportation, banking, gas and petroleum, health care, nuclear power plants, smart grids, water treatment and distribution, and electricity facilities [244]. Supervisors, developers, and end users substantially depend on technology for efficient governance and interactions to provide and obtain uninterrupted services. In today's world, services provided by industrial facilities are mere a click away.

With the advancements in modern information and control systems, a new generation of systems have emerged, featuring a combination of independently developed cyber and physical processes. These systems are called Cyber Physical Systems (CPS). A CPS is composed of various interacting elements that monitor and control the physical processes through a communication network. The various elements include software systems, communication technology, and sensors or actuators that interact with the real world. These systems have impact on every aspect of our society. Hence, are regarded as Critical Infrastructure (CI). Security is an issue of

prime importance for any CI. More so, because of the impact that it can have in the day to day activities.

Traditionally, these industries were not well protected in terms of security issues because they did not face as much of a threat as the conventional computing systems, which served as targets to the perpetrators. However, in recent times, the security issues and cyber attacks faced by these industries have escalated to a point where normal lives of human beings are at stake. Ill-motivated attackers jeopardize the expected functioning of the facilities as a result of which anticipated services either are halted or disrupted. Attacks on a Iranian nuclear facility[1], the Ukranian power grid[2], and the David-Besse Nuclear plant[3] are only a few unfortunate instances which illustrate the gravity of the security issues faced by such facilities. The operations and performance of these industries directly influence the smooth running of basic societal infrastructure. Furthermore, these issues impact upon the well-being of human lives in that particular geographical region, area, or organization [245]. Performance related issues, possible security breaches, and discontinuation of services can drastically hamper the associated habitual lives of human beings in the region.

During the process of transferring the data from physical world to the cyber world and again forwarding decisions from the cyber world to the physical world, a question that may be asked is where may a security breach occur. One of the many answers to this question may be in the communication network connecting the cyber and physical worlds. Attackers may carry out a Man-in-the-Middle attack (MITM attack) to manipulate the data being transferred. Analyzing the data in a normal situation and how it differs during an attack, can be a significant research challenge. When an attack takes place, the normal flow of data is disrupted and false data may be injected instead, thereby misleading the experts and preventing them from taking appropriate and timely decisions. Thus, everything boils down to the data in the critical infrastructure. Data manipulation and integrity attacks such as code injection attacks and false data injection attacks are common in critical infrastructure. It is not difficult to assume that, erroneous data instances are similar to one another because they are generated by a program and at the same time they are different from normal data. So, distinguishing the two with the help

[1] https://www.csoonline.com/article/562691/stuxnet-explained-the-first-known-cyberweapon.html

[2] https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf

[3] http://large.stanford.edu/courses/2015/ph241/holloway2

of statistical or machine learning methods is likely to detect attacks on critical infrastructure.
In this chapter, a defense framework, CPSAD (CPS Attack Detection Framework) which can help detect CPS attacks over optimal feature space is presented. A feature selection method called Feature Selection using Rank Aggregation (FSRA) is also proposed. FSRA is an ensemble feature selection method built over base feature selection methods. The proposed ensemble feature selection method ensures cost-effective classification of CPS attacks without compromising the performance.

7.1.1 Related Work

7.1.1.1 Cyber Physical Systems

A CPS intrusion detection system is proposed by Quincozes et al. [246] where the authors use a greedy metaheuristic approach called Greedy Randomized Adaptive Search Procedure (GRASP) [247] for selecting discriminating features to detect an attack in each of the three layers of CPS i.e. in the application layer, transmission layer as well as the perception layer. The conducted empirical study proves that such an approach is able to outperform traditional filter-based feature selection methods. A similar approach is presented in [248], where F1-score is used as a criteria along with the GRASP technique to detect binary attack classes as well multi-class attacks. However, the proposed approach is confined to detecting attacks only in the perception layer. On the other hand, in [249] the authors introduce a security tool which functions as an adaptive neuro-fuzzy inference system. The heart of the intrusion detection system relies on examining the incoming network traffic and selecting relevant attack features based on chi-square test.
Over the years, several researchers have conducted systematic reviews on Cyber Physical Systems widely focusing on the design mechanisms, security issues and system flaws, challenges faced in designing detection mechanisms, and various risk mechanisms involved in securing a CPS. [250][64][251][68]. Yan et al. [252] presents a detection method which detects attack in the physical layer of a heavy duty gas turbine facility by monitoring behavioral changes in the physical processes. The physical processes in a Cyber Physical System provide measurements from which salient features are extracted with an underlying belief that these features can discriminate an attack and a normal measurement or instance. On the contrary, Saghezchi et al. [253] construct data driven models to detect DDoS attacks in

industrial CPSs. Unlike others, the authors collect data from a real world semiconductor production factory which helps in emphasizing how effective the developed method is. Panigrahi et al. [254] presents a signature-based intrusion detection technique for CPSs. The approach is based on the combination of both Decision tree and Naive Bayes and is capable of handling multi-class attacks. Multi-objective Evolutionary Feature Selection (MOEFS) [255] method is used in the preprocessing stage for selecting a total of five features which aids in detecting the attacks.

7.1.1.2 Feature Selection

Ensemble approaches rely on the assumption that decision given by many experts is always better than the decision of a single expert. In ensemble feature selection, base feature selection methods provide subsets of features which are then combined using a consensus function to get the final feature subset. The authors in [256] propose an ensemble feature selection method which considers filter, wrapper and embedded methods for combination in medical datasets. Two strategies are presented for selection of feature subsets. In the first strategy, two different types of feature selection methods are combined (filter-wrapper, filter-embedded, and wrapper-embedded). In the second strategy, all the three feature selection methods are combined together. For combination of the feature subsets the authors use both union and intersection. The method is then evaluated with datasets of varying dimensions containing categorical, numerical and mixed data types. Hashemi et al. [257] propose Ensemble Feature Selection - Multi Criteria Decision Making (EFS-MCDM) approach which makes use of different ranker algorithms. Each ranker algorithm provides a rank based on some score to each feature and a decision matrix is formed. The individual ranks of these algorithms from the decision matrix is then combined using the VIKOR approach proposed by Opricovic [258][259]. The final output is a ranked list of features.
Basir et al. [260] introduces an ensemble feature selection method based on a bio-inspired search technique. The method tries to find the optimal feature subset using multi-objective algorithms ENORA [261][262] and NSGA-II [263]. A similar method is proposed in [264] where the authors try to improve the generalization capability of the ensemble by trying to minimize the training error and sensitivity with the NSGA-III algorithm [265]. Sanjalawe and Althobaiti [266] propose an ensemble feature selection and hybridized detection technique to detect DDoS attacks

in the cloud. The base feature selection techniques include four algorithms namely Particle Swarm Intelligence (PSO) [267], Grey Wolf Optimizer (GWO) [268], Krill Herd (KH) [269], and Whale Optimization Algorithm (WAO) [270]. Each of the base selectors provide a ranked list which is then aggregated using mutual information for optimal feature selection. For detecting the attack traffic efficiently, the authors employ a hybridized approach combining both Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). A similar metaheuristic ensemble approach is presented by Dey et al. [271] to detect attacks in IoT networks. For selecting optimized set of features, two base feature selection approaches are considered namely, Binary Gravitational Search Algorithm (BGSA) and Binary Grey Wolf Optimization (BGWO). Additionally, a proposed fitness function help measure the quality of the solution. The selected features are then passed to Random Forest and Adaboost ensemble classifiers.
Authors in [272] introduce an ensemble approach in the classification level. To remove unwanted features, an algorithm called Correlation-based Feature Selection - Bat Algorithm (CFS-BA) is proposed which is responsible for choosing a feature subset based on feature-feature correlation. In the classification level, using voting mechanism the decisions of three base classifiers are combined, namely, C4.5, Random Forest and Forest by Penalizing Attributes. Interestingly, Kolukisa and Gungor [273] in addition to proposing an ensemble feature selection method (with seven base feature selectors) for diagnosing coronary artery disease also perform Grid Search parameter optimization for finding the best possible set of hyperparameters for the learners. Hoque et al. [17] propose a method named Ensemble Feature Selection - Mutual Information (EFS-MI) where five base filter feature selectors are combined using a greedy combiner based on mutual information. If a feature is a top ranked feature and is commonly chosen as top ranked by all the base selectors then it is directly part of the final subset. However, features which are not commonly ranked, for them the feature-class mutual information is measured. Feature which has the highest feature-class mutual information is selected and its feature-feature mutual information is calculated with all other features from the final subset. If the calculated value is less than a given threshold, the feature becomes a part of the final subset.

Unlike others, authors in [274] consider three filter-based and two embedded feature selection methods as base selectors. To establish the degree of divergence among the base selectors, a correlation study is carried out using Spearman's cor-

relation among the obtained feature ranks from the selectors. The ranked lists of features are combined using a bunch of aggregators like SVM-Rank aggregator [275], geoMean, Stuart to name a few. At the classification level, each combined feature list provided by each aggregator is compared against each other for identifying the best feature subset. Hashemi et al. [276] introduce a Pareto-based ranking ensemble feature selection method, which considers both feature-class relevance and feature-feature redundancy as the two objectives. To achieve this objective, three base feature selection methods are used and the final list of features is obtained by using the crowding distance of features. On the other hand, in [277], the authors build an ensemble of bayesian classifiers in random feature subspaces. To achieve better performance, Hill Climbing search algorithm is used for refining the results iteratively. From the experiments, it is concluded that in addition to producing better results than the individual learners, the ensemble also yields low generalization error.

7.1.2 Motivation

Critical Infrastructures play an immense role in making several critical aspects of our day to day operations or activities easy and convenient for us. These infrastructures provide useful services to people of a geographical area with the help of an integrated system consisting of a number of cyber and physical elements. These elements and their integrated functioning are potential vulnerable points and hence, pose serious risk in terms of security. If security is breached, not only does the associated components fall under the influence of the attacker but also the end-users are at the mercy of the attacker. Malicious users may either exploit the existing vulnerabilities or launch a fresh attack to disrupt the ongoing functions. In such a scenario, identifying the affected element at the earliest is prime and single point of focus. Devastating and recurring attacks on critical infrastructures over the years has motivated to catalog solutions based on appropriate use of statistical and machine learning techniques. Therefore, the aim is to develop a solution which could identify the infected element in a critical facility in near real time.

7.1.3 Contributions

Following are the two major contributions reported in this chapter:

1. A framework called CPSAD for cost-effective CPS attack classification over

optimal feature space with high accuracy and minimum false alarm.

2. A method called FSRA to identify an optimal subset of relevant features to help detect CPS attacks at minimum cost without compromising the performance.

Table 7.1 depicts the symbols and notations used to describe the proposed method.

Table 7.1: Symbol Table for the Proposed Method

Symbol	Symbol Meaning	Symbol	Symbol Meaning
D	Dataset	RF_{list}	Ranked list of feature as given by a base feature selection method
s	no. of samples in D	n	no. of base feature selection methods
d	no. of features in D	$rank_i$	rank of a feature as given by ith base feature selection method
F	Original feature set of D	$RF_{agglist}$	Aggregated ranked list of features
f_i	a feature in D	$F_{commlist}$	Common feature subset
B_{FS}	Base feature selection method	$F_{optimal}$	Optimal feature subset

7.2 Problem Formulation

For a given dataset D with s samples and d features, the problem is to find the optimal subset of features, $F_{optimal}$ from the original set of features F. Features in $F_{optimal}$ should help discriminate an attack instance from the normal instances in the domain of CPSs. The subset of features should be selected in such a way that any increase in the number of features does not improve performance of the learning model and any decrease in the number of features deteriorates the performance.

7.3 CPSAD: Proposed Attack Detection Framework

CPSAD is a framework to detect CPS attacks with high accuracy. It has four modules, namely detection module, alarm generation module, feedback analyzer

module, and reference or rule generation module. Each module has its own functionality and may or may not depend on other module for input. The framework is illustrated in the Figure 7.1 and the modules are described in length below.

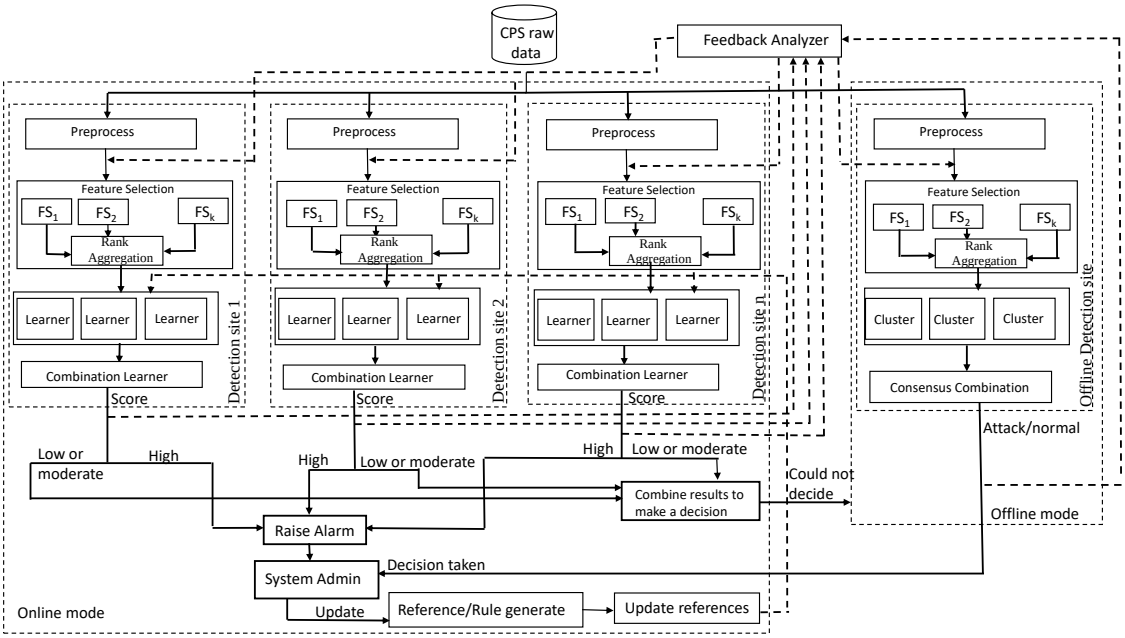


Figure 7.1: CPSAD: An attack detection framework

7.3.1 Detection Module

The aim of this module is to discriminate attack instances from normal instances. It is the heart of the proposed framework which operates in two modes, online and offline. While online mode is used to detect known attacks, offline mode detects unknown varieties (doubtful traffic, can be a novel attack or new normal). For online mode of operation, the module employs an ensemble feature selection method to choose an optimal subset of features. This feature subset is used by the learners whose results may then be used by a combination learner to produce an anomaly score. For unknown or suspicious instances, the framework operates in the offline mode, where after obtaining the optimal subset of features clustering approaches may be performed and a consensus may be built to carry on the task of detection. There can be a number of sub-modules for this module as described below.

1. *Preprocessing sub-module:* It carries out the basic tasks of preprocessing such as missing value estimation, removal of duplicates, removal of zero variance attributes, and normalization.

2. *Feature selection sub-module:* It helps to choose an optimal subset of features which helps discriminate between normal and attack instances. The proposed ensemble feature selection method is described in Section 7.3.2.

3. *Learner/cluster sub-module:* Builds supervised as well as unsupervised prediction models based on the selected subset of features to identify known and unknown attacks.

4. *Combination learner/consensus combination:* Combines the results of the learners or individual clusters to generate an unbiased final decision. The final output may be a score or a flag specifying the instance as normal or attack.

CPSAD functions in both online and offline mode. In online mode, it is important to note that if a particular detection site gives a sufficiently high anomaly score compared to a threshold then it is declared to be an attack at that site itself. In such a case, the defense system does not wait for the results from other detection sites. Once an attack is detected, an automated alarm is sent to the system administrator along with associated essential information (if any). However, for low or moderate anomaly scores the attack sensing detection sites deployed at multiple points co-ordinates among each other and finally a decision is taken to flag the instance as either normal or suspicious. To define the point of attack, the following cases are considered.

1. A coordinated attack detection is carried out at all the sites. In this case, the deployed detection modules along with all their anomaly scores raise alarms to the system administrator, who in turn takes the necessary steps to tackle the attack.

2. An attack takes place at only one site, in such a case the system administrator after receiving the alarm carefully disconnects the locally attacked network from the rest of the network in order to stop the propagation of the attack.

7.3.2 FSRA: Proposed Ensemble Feature Selection Method

Feature Selection Using Rank Aggregation (FSRA) focuses on unbiased combination of the decisions given by the feature ranker algorithms to obtain an optimal

subset of features which help discriminate between normal and attack instances. The framework is illustrated in the Figure 7.2. Following definitions are useful to

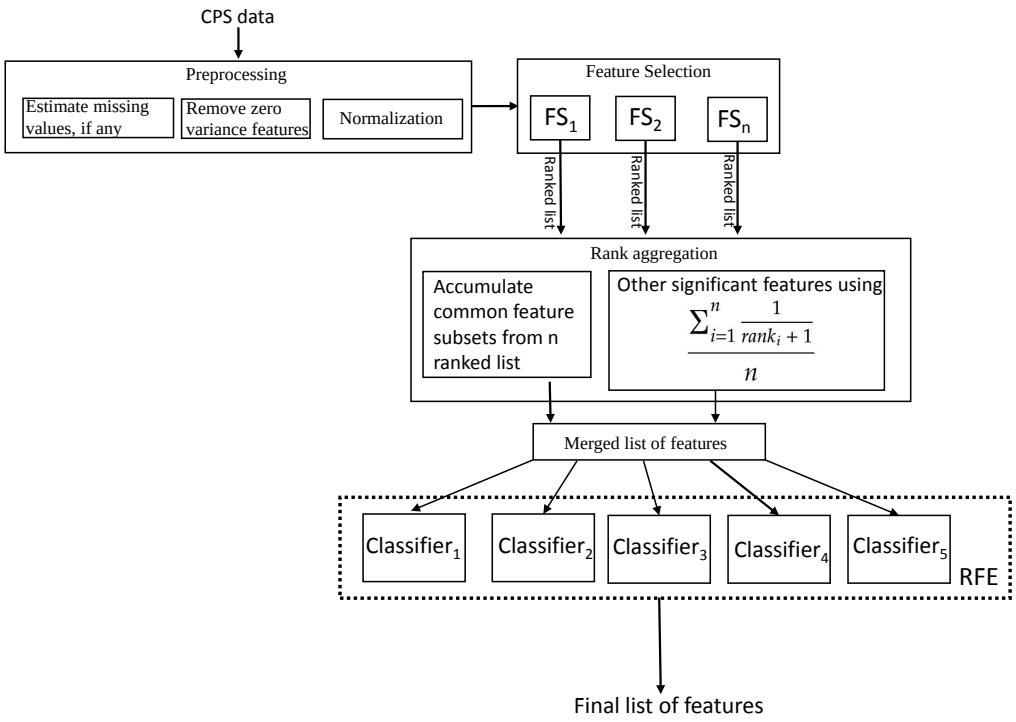


Figure 7.2: FSRA: Feature Selection using Rank Aggregation

describe the proposed method.

Definition 7.1 (Feature of a CPS dataset)**.** A feature f_i represents the characteristics of a CPS dataset, D given by the sensors or actuators.

Definition 7.2 (Ranked feature list)**.** It is an organized list of features based on the ranks given by each feature selection algorithm.

Definition 7.3 (Rank aggregated feature list)**.** It is the subset of features obtained after aggregating the base ranked feature lists given by each ranker algorithm.

Definition 7.4 (Optimal feature subset)**.** It is the subset of features, $F_{optimal}$ obtained after recursive elimination of less significant features from the aggregated ranked feature list.

Definition 7.5 (Known attack)**.** An attack instance is considered to be known if it has already been encountered and its reference is available for use.

Definition 7.6 (Suspicious instance)**.** An instance is considered to be suspicious if it has not been encountered yet and reference is not available for use. It can either be a new attack or even a new normal instance.

7.3.2.1 FSRA framework

The FSRA framework includes a sequence of steps as presented below.

1. After the CPS data is collected, it undergoes preprocessing using several techniques. Missing values if any are estimated, zero variance features (if any) are removed and feature values are normalized using *min-max* normalization (all values between 0 to 1).

2. The preprocessed dataset is then used as input to n base feature selection methods to obtain highly informative features. Each base feature selection method, B_{FS} gives a ranked list of features, RF_{list} containing features from most important to least important.

3. The ranked lists of features are then given as input to the proposed rank aggregation module to obtain an aggregated ranked list of features, $RF_{agglist}$. The rank aggregation module functions as follows.

   (a) From the obtained n ranked lists from the feature selection algorithms, first, the common features are considered to obtain a common feature subset, $F_{commlist}$.

   (b) The other significant features are identified using the equation 7.1:

   $$\frac{\sum_{i=1}^{n} \frac{1}{rank_i+1}}{n} \tag{7.1}$$

   Here, n signifies the number of base feature selection methods, and $rank_i$ denotes the rank of a feature given by the i^{th} base feature selection method. This ensures that if a particular feature is given sufficiently high rank by a method but is not common to all, that feature is preserved in the aggregation process. The output of this step is a merged list of features, $RF_{agglist}$.

4. The merged list of features are given to the learners to obtain an optimal final subset of features, $F_{optimal}$ using Recursive Feature Elimination (RFE).

Proposition 7.1. The subset of features $F_{optimal}$ given by FSRA for a given dataset is relevant and optimal.

Proof. Let $f_i \epsilon F_{optimal}$ is a feature, selected by FSRA which is not relevant and redundant. However, a feature is selected by FSRA only when-

(a) f_i is commonly declared by all the base ranker algorithms as relevant, or

(b) f_i is specified by any (or some) ranker algorithms as highly relevant (due to assignment of high rank).
So, f_i selected by FSRA is relevant.

Similarly, the subset $F_{optimal}$ is optimal because

(a) Any increase in the number of features does not improve the classifier performance. This can be seen in the results of SWaT dataset (Figures: 7.3, 7.4, 7.5, 7.6 and 7.7)

(b) Any decrease (elimination) in the number of features deteriorate the performance. This can be seen in the results of Gas Pipeline dataset (Figures: 7.13, 7.14, 7.15 and 7.16, 7.17).
Hence, the assumption is false and hence the proof.

■

7.3.2.2 Complexity Analysis

The overall complexity of the proposed method depends on the individual learning algorithms and the rank aggregation process. So, the complexity of FSRA will be the dominating complexity of the two. The rank aggregation process which is calculated by Equation 7.1 will have a linear complexity. The training complexity of the learning algorithms on the other hand, will largely depend on the number of training samples (*say* n) and the number of features (*say* d) considered. Hence, the complexity of FSRA can be concluded as $\mathcal{O}(n^2d)$.

7.3.3 Alarm Generation Module

This module receives input from individual detection sites and output is a possible alert to the system administrator in case of a potential attack. When the anomaly scores are received from different sites, they are combined to form a definite decision. The system administrator is alerted by this module in case of an attack.

7.3.4 Feedback Analyzer Module

The framework promises to minimize the false alarms through the use of feedback analyzer module. Input is provided as feedback from the detection module and output is in the form of updates to the detection module or its sub-modules. It takes the feedback (if any) from the detection modules at multiple points and updates the training database. Subsequently, the feature subsets may be updated. Such a module is needed to effectively and continuously improve the learning process and to reduce false alarms.

7.3.5 Reference or Rule generation Module

The reference or rule generation module consists of the references of known attacks and is used by the system administrator. Accordingly, the output of the reference or rule generation module is used by the learners in the detection module. Unknown instances are recognized by the system administrator (or domain expert) either as new normal or new attack. Conclusively, the primary goal is to detect and discriminate anomalous instances from normal instances as correctly as possible over an optimal subset of features. This module works offline under the supervision of the system administrator.

7.4 Experimental Results

Three feature selection methods namely, Conditional Mutual Information Maximization (CMIM) [200], Minimum Redundancy and Maximum Relevance (mRMR) [228] and Mutual Information based Feature Selection (MIFS) [193] are selected as the base feature selection methods after an exhaustive empirical study. The preprocessed datasets are fed as input to these methods and each then provides a ranked list of features. The ranked list of features (ranked index of a feature) for the datasets are shown in Table 7.2.

After obtaining the ranked list of features, the next task is to find the aggregated feature list (as mentioned in step 3) as shown in Table 7.3 and 7.4. The column *index_of_feature* signifies the index number of the feature in the particular dataset, score indicates the score obtained by the feature using Equation 7.1, and the common column denotes if the feature is common to all the feature selection methods.

Table 7.2: Ranked List of Features for the Datasets

Dataset	Base Feature Selection Method	Ranked features
SWaT Dataset	CMIM	22, 16, 12, 20, 9, 23, 14, 21, 10, 11, 1, 18, 7, 8, 24, 19, 3, 6, 17, 13, 5, 2, 4, 0, 15
	MIFS	45, 8, 37, 39, 27, 21, 18, 7, 2, 0, 35, 19, 46, 1, 31, 20, 29, 40, 48, 26, 28, 25, 47, 17,3
	mRMR	45, 39 , 8, 20 ,46, 17, 28 ,37, 35, 26, 27, 9, 21, 18, 11, 7, 38, 2, 48, 5, 0, 19, 1, 10, 30
Gas Pipeline	CMIM	25, 24, 8, 18, 19, 1, 21, 0, 6, 10, 22, 2, 4, 3, 5, 9, 12, 20, 7, 11
	MIFS	25, 1, 18, 20, 10, 3, 7, 11, 13, 14, 15, 16, 17, 23, 4, 6, 0, 2, 8, 22
	mRMR	25, 1, 18, 8, 3, 5, 19, 9, 12, 24, 20, 0, 10, 6, 4, 2, 7, 11, 13, 14
Water Storage	CMIM	21, 22, 1, 14, 17, 0, 10, 4, 7, 19, 16, 15, 13, 3, 5, 9, 12,20
	MIFS	21, 1, 14, 3, 13, 10, 0, 2, 6, 8, 11, 18, 19, 15, 7, 16, 5, 9
	mRMR	21, 1, 14, 3, 5, 9, 12, 13, 16, 4, 22, 17, 15, 20, 10, 0, 7, 2

If the feature is common to all, it is marked as 'Yes' and otherwise 'No'.

Table 7.3: SWaT Dataset Aggregated List of Features

SWaT Aggregated List of Features							
Sl No.	Index of feature	Score	Common ?	Sl No.	Index of feature	Score	Common ?
1	45	0.666667	NO	14	7	0.088141	YES
2	22	0.333333	NO	15	17	0.086988	YES
3	8	0.301587	YES	16	2	0.070707	YES
4	39	0.25	NO	17	1	0.068605	YES
5	20	0.1875	YES	18	35	0.06734	NO
6	16	0.166667	NO	19	19	0.063763	YES
7	37	0.152778	NO	20	28	0.063492	NO
8	21	0.122863	YES	21	0	0.063095	YES
9	12	0.111111	NO	22	11	0.055556	NO
10	18	0.099206	YES	23	23	0.055556	NO
11	27	0.09697	NO	24	10	0.050926	NO
12	9	0.094444	NO	25	26	0.05	NO
13	46	0.092308	NO				

The aggregated list of features are given as input to the five popular classifiers namely Adaboost [175], Random Forest [226], XGboost [278], Extra Trees [227] and Gradient Boosting [224] and Recursive Feature Elimination (RFE) is applied to this process. This process is done so as to know which classifier gives better performance in terms of accuracy and F1-score for optimal number of features.
For example in the SWaT dataset, the highest accuracy of 99.9% is obtained in case of XGBoost classifier with 8 features as shown in Figure 7.3. Random Forest and Extra Trees classifier show similar performance as shown in Figure 7.4 and 7.5 with

Table 7.4: Aggregated Ranked List of Features for Gas Pipeline and Water Storage Dataset

Gas Pipeline Aggregated List of Features			Water Storage Aggregated List of features		
Index of feature	Score	Common?	Index of feature	Score	Common?
25	1	Yes	21	1	Yes
1	0.388889	Yes	1	0.444444	Yes
18	0.305556	Yes	14	0.305556	Yes
8	0.211988	Yes	22	0.19697	No
24	0.2	NO	3	0.190476	Yes
3	0.146032	Yes	13	0.133974	Yes
20	0.132155	YES	10	0.125397	Yes
10	0.125641	YES	0	0.124008	Yes
19	0.114286	No	5	0.108497	Yes
0	0.089052	Yes	9	0.094907	Yes
7	0.084771	Yes	17	0.094444	No
6	0.08168	Yes	16	0.088173	Yes
5	0.077778	No	7	0.078867	Yes
11	0.076852	Yes	15	0.077228	Yes
4	0.070085	Yes	4	0.075	No
2	0.06713	Yes	12	0.067227	No
9	0.0625	No	2	0.060185	No
12	0.056645	No	19	0.058974	No
13	0.054581	No			
14	0.05	No			

8 features. However, Adaboost and Gradient Boosting classifiers show accuracy $<$ 99% and with 14 features as shown in Figure 7.6 and 7.7. Therefore, in case of SWaT dataset it can be concluded that XGBoost gives better performance than rest of the classifiers. For the same dataset, F1-scores are illustrated in Figure 7.8 for XGBoost, Figure 7.9 for Adaboost, Figure 7.10 for Extra Trees, Figure 7.11 for Gradient Boosting, and Figure 7.12 for Random Forest classifier.

On the other hand, for the Gas pipeline dataset, Gradient Boosting achieves highest accuracy of 94.9% with 6 features as shown in Figure 7.13. It is important to note that, Adaboost achieves an almost similar accuracy of 94.3% with 4 features as can be seen in Figure 7.14. So, there is a trade off between accuracy achieved and the number of optimal features. Also, both Extra Trees and XGBoost classifiers achieve accuracy of 92.6% and 91.5% respectively with 4 features each as shown in Figures 7.15 and 7.16. Random Forest classifier obtains the lowest accuracy of 84.5% with 7 features as shown in Figure 7.17. Thus, with respect

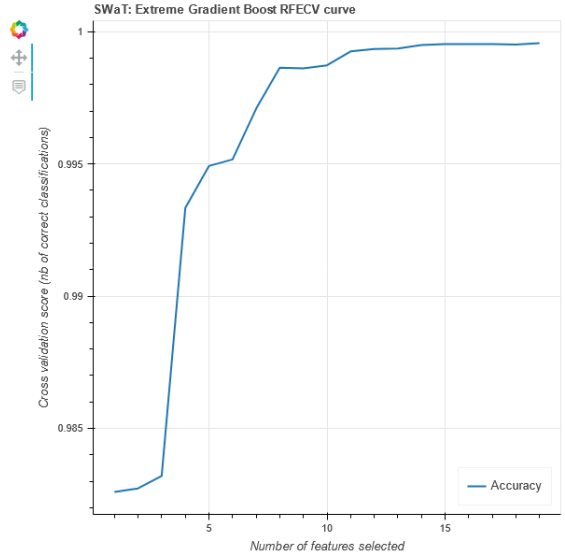


Figure 7.3: RFE with XGBoost classifier for SWaT dataset (Accuracy)

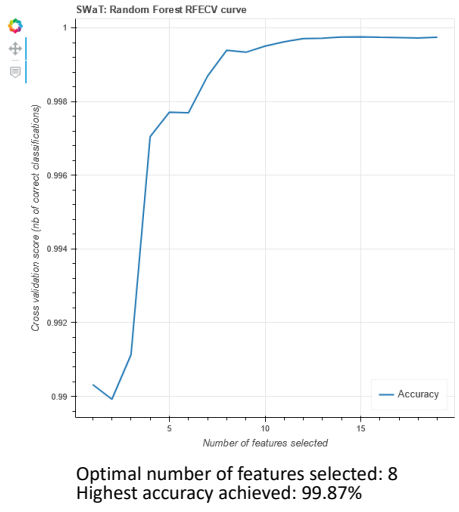


Figure 7.4: RFE with Random Forest classifier for SWaT dataset (Accuracy)

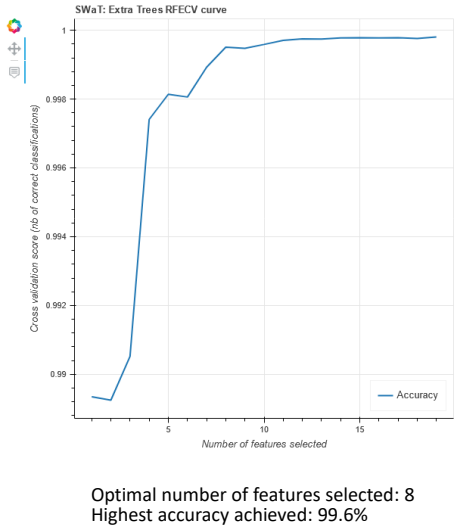


Figure 7.5: RFE with Extra trees classifier for SWaT dataset (Accuracy)

to highest accuracy achieved, it is concluded that Gradient Boosting gives better performance in case of Gas Pipeline dataset. For the same dataset, F1-score results are illustrated in Figure 7.18 for XGBoost, Figure 7.19 for Extra Trees, Figure 7.20 for Random Forest, Figure 7.21 for Gradient Boosting, Figure 7.22 for Adaboost.

For the water pipeline dataset, XGBoost classifier provides best performance with 89.2% and 4 optimal features as shown in Figure 7.23. Random Forest and Extra Trees achieve 88.1% and 87.9% with 8 features as presented in Figure 7.24 and 7.25 respectively. Adaboost on the other hand, achieves 71% accuracy with

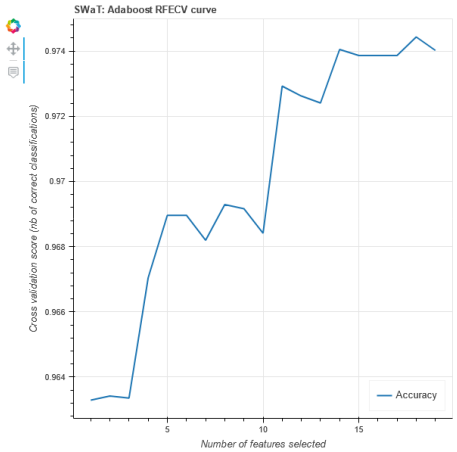


Figure 7.6: RFE with Adaboost classifier for SWaT dataset (Accuracy)

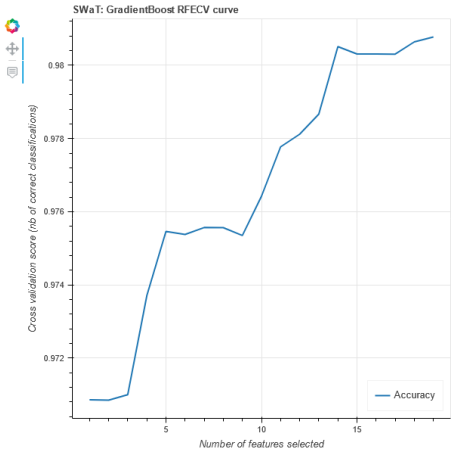


Figure 7.7: RFE with Gradient Boosting classifier for SWaT dataset (Accuracy)

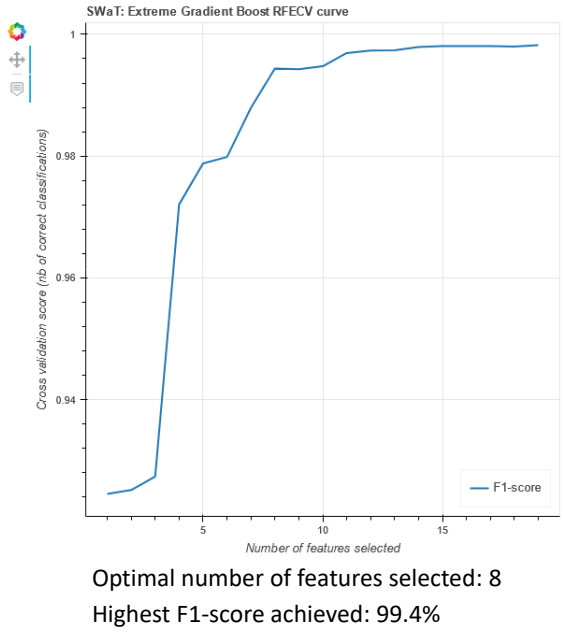


Figure 7.8: RFE with XGBoost classifier for SWaT dataset (F1-score)

11 features as seen in Figure 7.26. Of special mention is the Gradient Boosting classifier which achieves 72% accuracy with 1 feature only as depicted in Figure 7.27. However, it cannot be said that it gives optimal performance because XGBoost obtains higher accuracy than Extra Trees with 4 features. All the other classifiers (except Gradient Boosting) require more than 4 features to achieve the given performance. Hence, XGBoost is chosen to be the best classifier with optimal performance in case of water pipeline dataset. For the same dataset, F1-scores are illustrated in Figure 7.28 for XGBoost, Figure 7.29 for Adaboost, Figure 7.30 for

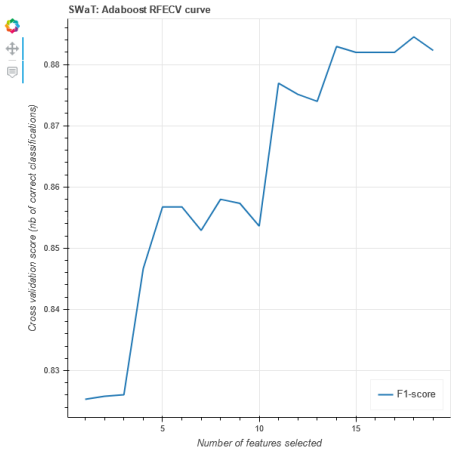


Figure 7.9: RFE with Adaboost classifier for SWaT dataset (F1-score)

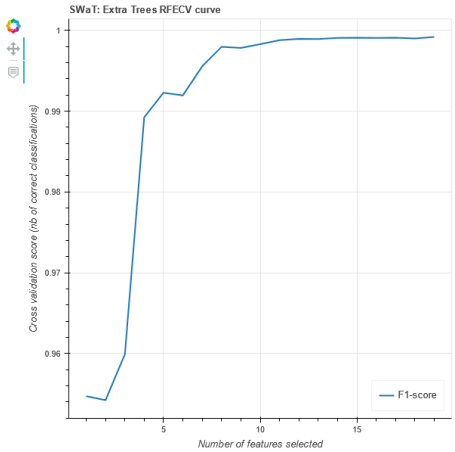


Figure 7.10: RFE with Extra trees classifier for SWaT dataset (F1-score)

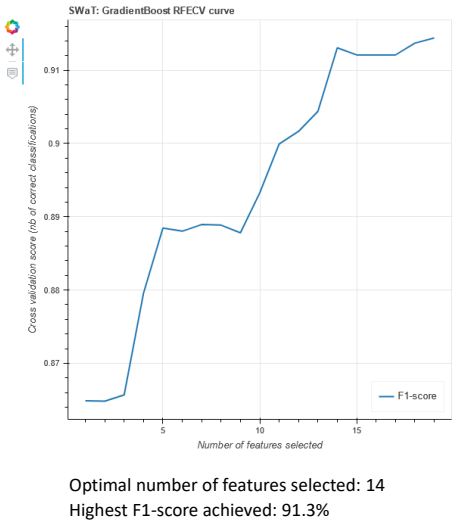


Figure 7.11: RFE with Gradient Boosting classifier for SWaT dataset (F1-score)

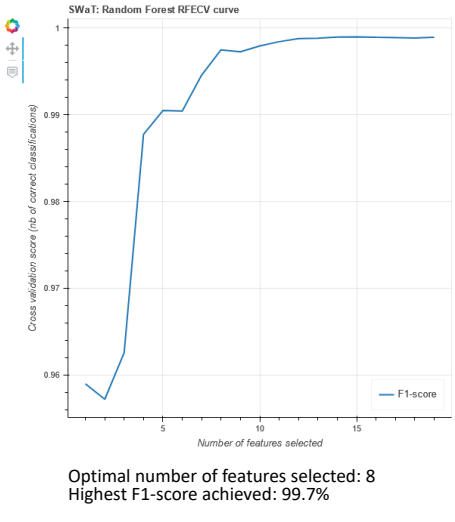


Figure 7.12: RFE with Random Forest classifier for SWaT dataset (F1-score)

Extra Trees, Figure 7.31 for Gradient Boosting, and Figure 7.32 for Random Forest classifier.

From the illustrated results, it can be seen that Extreme Gradient Boosting (XGBoost) classifier shows a consistent performance in two out of the three datasets used. The reason for the same may be that it can successfully boost up weak learners to gradually improve the performance in comparison to its counterparts.

Furthermore, Table 7.5 shows the top 10 ranked features for each of the CPS dataset considered as given by the proposed ensemble feature selection method.

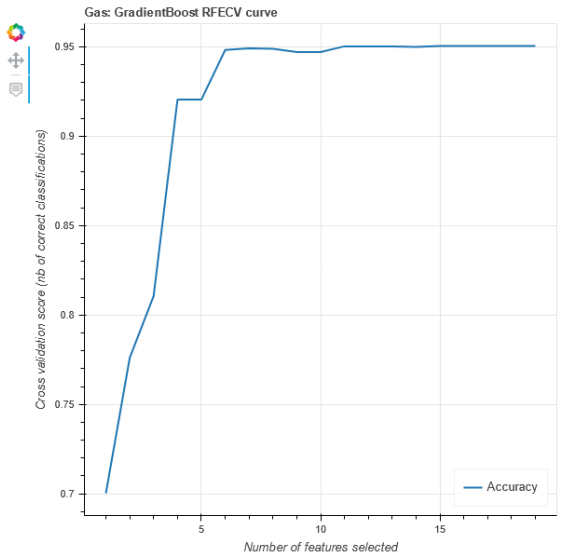


Figure 7.13: RFE with Gradient Boosting classifier for GAS pipeline dataset (Accuracy)

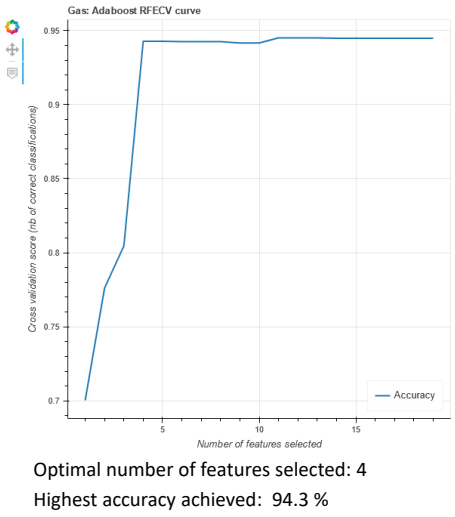


Figure 7.14: RFE with Adaboost classifier for GAS pipeline dataset (Accuracy)

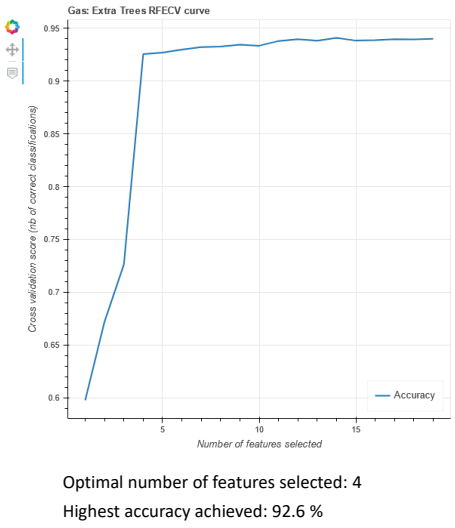
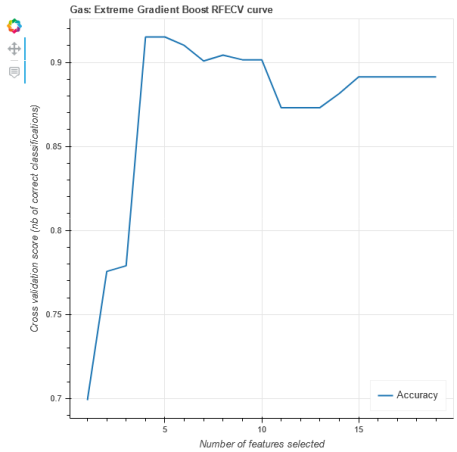


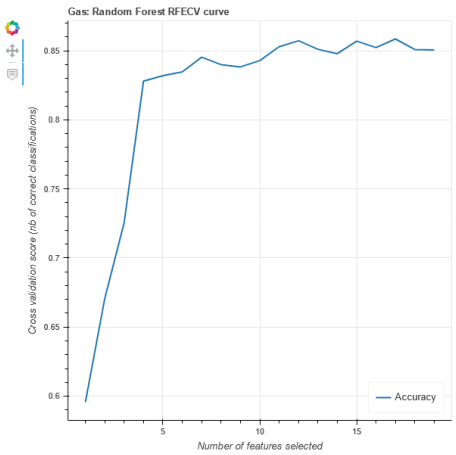
Figure 7.15: RFE with Extra trees classifier for GAS pipeline dataset (Accuracy)

The columns *feature name* and *index number* gives the name of the feature and the index number in the corresponding dataset. This table signifies which features are more informative in identifying an attack. In other words, FSRA is successful in identifying the point of attack in a CPS facility and the corresponding affected element. For example, for the SWaT dataset, FSRA identifies FIT201, LIT301, P203 and MV304 as informative features. FIT201 and LIT301 are indeed two point of attacks in the SWaT facility as described by Adepu and Mathur [279] and P203 and MV304 are the actuators affected because of it.



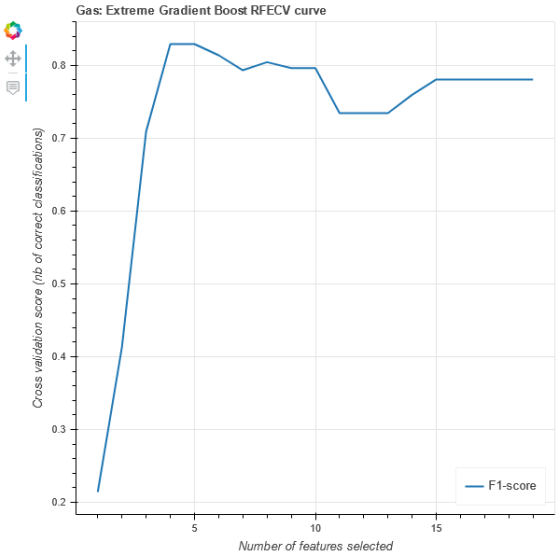
Optimal number of features selected: 4
Highest accuracy achieved: 91.5%

Figure 7.16: RFE with XGBoost classifier for GAS pipeline dataset (Accuracy)



Optimal number of features selected: 7
Highest accuracy achieved: 84.5%

Figure 7.17: RFE with Random Forest classifier for GAS pipeline dataset (Accuracy)



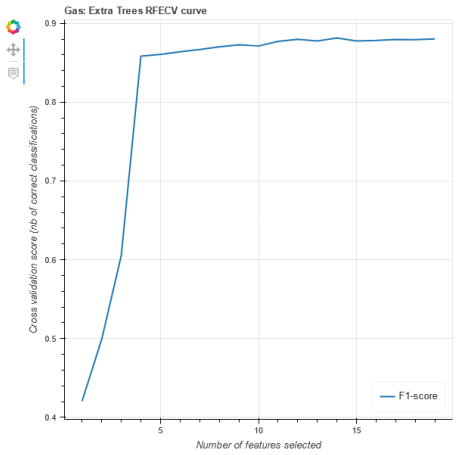
Optimal number of features selected: 4
Highest f1-score achieved: 83%

Figure 7.18: RFE with XGBoost classifier for GAS pipeline dataset (F1-score)

7.4.1 Comparison with Existing Methods

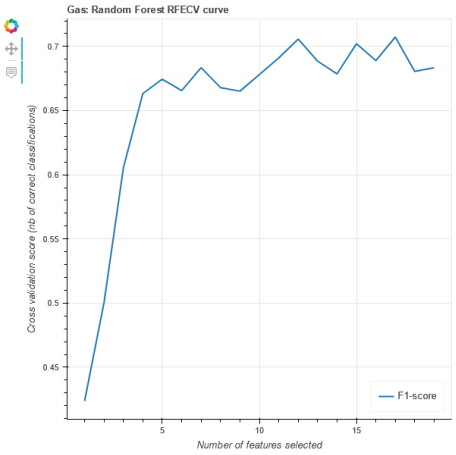
The proposed detection method which relies on identifying the informative features in a CPS facility is compared against some existing methods in Table 7.6.

1. Unlike [272], where the authors use ensemble learning (voting mechanism) at the classification level, the proposed method uses ensemble learning at the feature selection level.



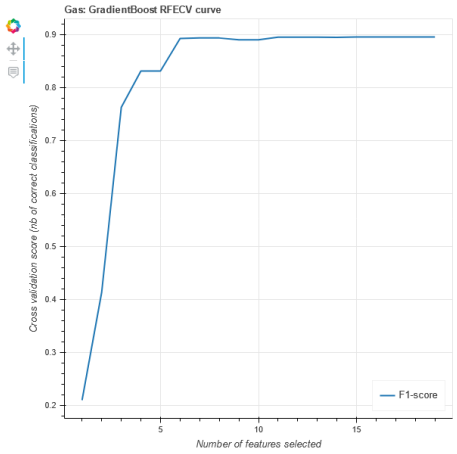
Optimal number of features selected: 4
Highest f1-score achieved: 86%

Figure 7.19: RFE with Extra trees classifier for GAS pipeline dataset (F1-score)



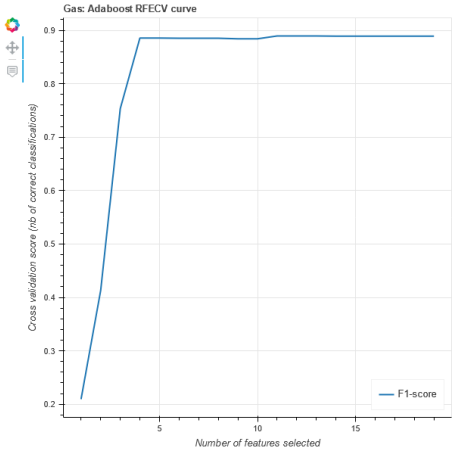
Optimal number of features selected: 12
Highest f1-score achieved: 70.5%

Figure 7.20: RFE with Random Forest classifier for GAS pipeline dataset (F1-score)



Optimal number of features selected: 6
Highest f1-score achieved: 89.3 %

Figure 7.21: RFE with Gradient Boosting classifier for GAS pipeline dataset (F1-score)



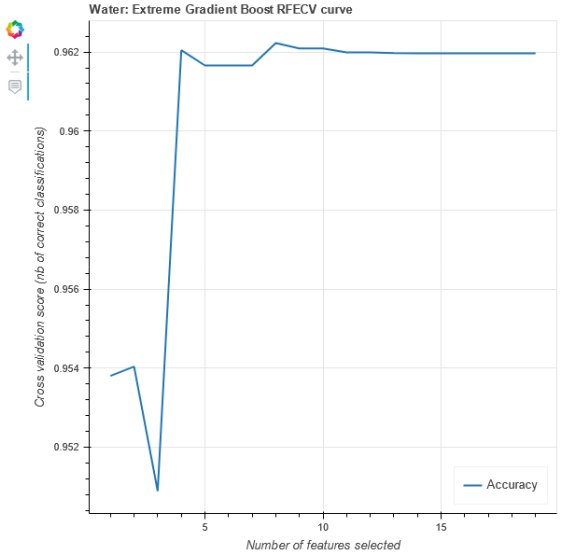
Optimal number of features selected: 4
Highest f1-score achieved: 88.3 %

Figure 7.22: RFE with Adaboost classifier for GAS pipeline dataset (F1-score)

2. Like [274], the proposed method is a heterogeneous feature selection method which tries to combine different feature rankers to obtain a final list of ordered features.

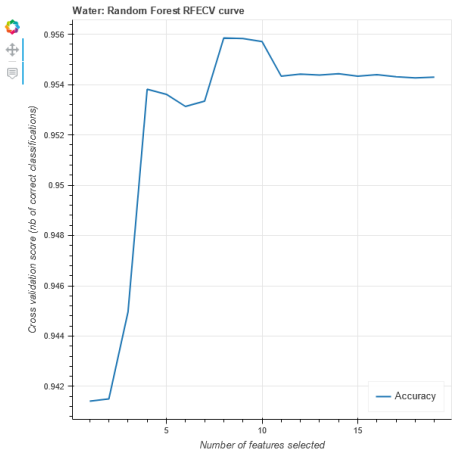
3. Unlike [277], where the authors consider only Bayesian learners for evaluation of the final feature subset, the subset generated by FSRA is evaluated with ensemble learners (tree-based ensembles and boosting ensembles).

4. Like [17], FSRA is also an ensemble feature selection method. However, for combining the base ranked lists the authors use feature-feature and feature-



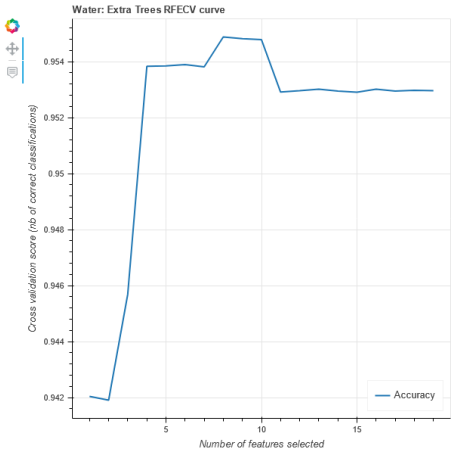
Optimal number of features selected: 4
Highest accuracy achieved: 96.2%

Figure 7.23: RFE with XGBoost classifier for Water pipeline dataset (Accuracy)



Optimal number of features selected: 8
Highest accuracy achieved: 95.6%

Figure 7.24: RFE with Random Forest classifiers for Water pipeline dataset (Accuracy)



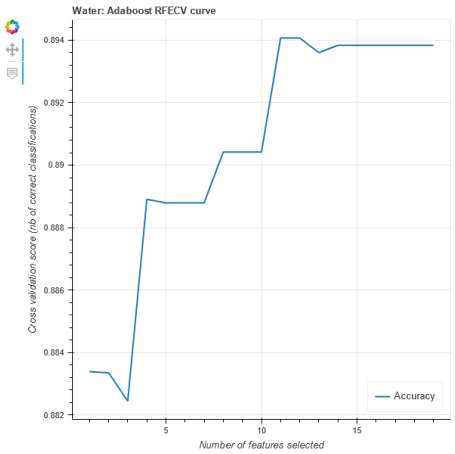
Optimal number of features selected: 8
Highest accuracy achieved: 95.5%

Figure 7.25: RFE with Extra trees classifier for Water pipeline dataset (Accuracy)

class mutual information.

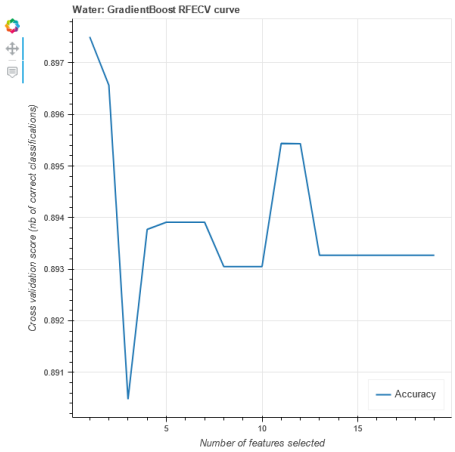
5. Unlike [274], during the aggregation process the proposed method ensures that a sufficiently high ranked feature is preserved even though it is not common to all the base feature ranked lists.

6. Unlike [246], where the authors propose a meta-heuristic greedy feature selection method for CPS attack detection, the proposed method is an ensemble feature selection method which mainly focuses on identifying different point



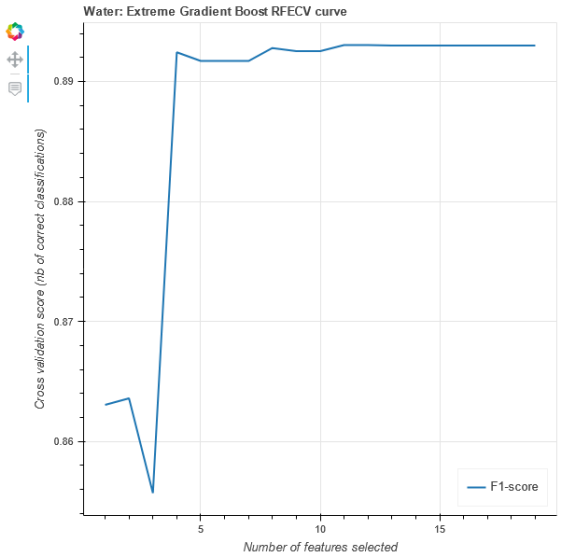
Optimal number of features selected: 11
Highest accuracy achieved: 89.4%

Figure 7.26: RFE with Adaboost classifier for Water pipeline dataset (Accuracy)



Optimal number of features selected: 1
Highest accuracy achieved: 89.7%

Figure 7.27: RFE with Gradient Boosting classifier for Water pipeline dataset (Accuracy)



Optimal number of features selected: 4
Highest F1-score achieved: 89.2%

Figure 7.28: RFE with XGBoost classifier for Water pipeline dataset (F1-score)

of attacks and the affected elements in a CPS facility.

7.5 Discussion

In this chapter, two frameworks namely CPSAD and FSRA are presented. CPSAD is a cost-effective and attack detection defense architecture which functions to detect known variety of attacks with high accuracy and minimum false alarm. On

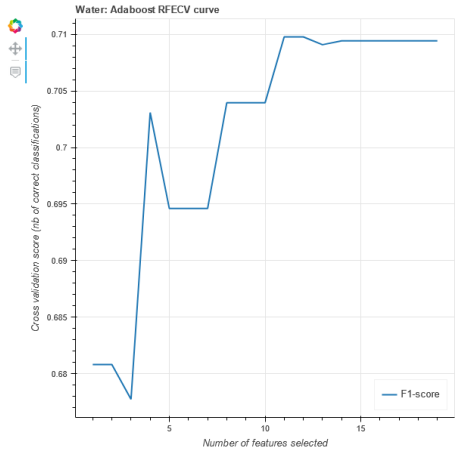


Figure 7.29: RFE with Adaboost classifier for Water pipeline dataset (F1-score)

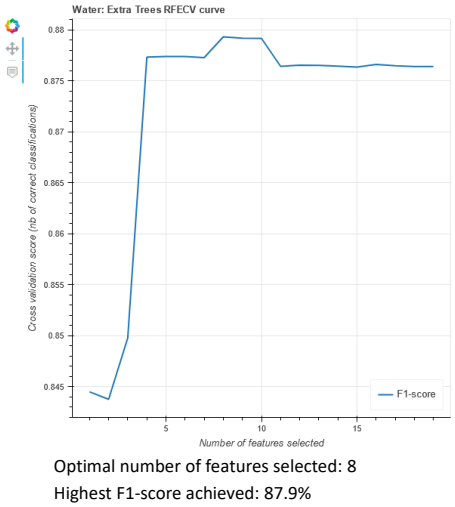


Figure 7.30: RFE with Extra trees classifier for Water pipeline dataset (F1-score)

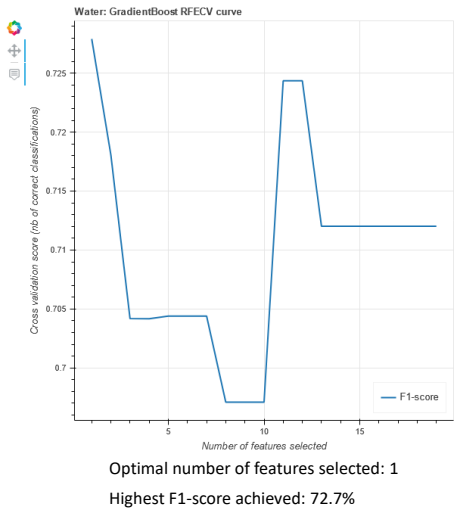


Figure 7.31: RFE with Gradient Boosting classifier for Water pipeline dataset (F1-score)

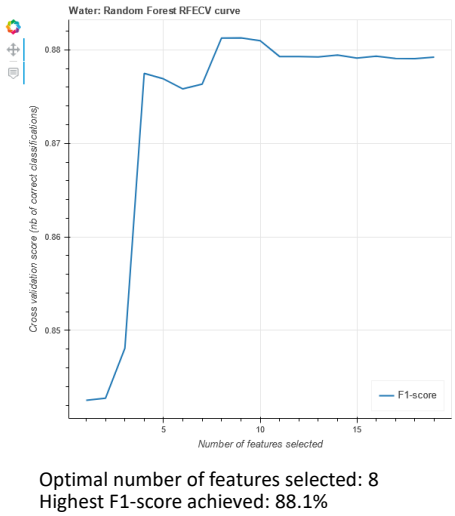


Figure 7.32: RFE with Random Forest classifiers for Water pipeline dataset (F1-score)

the other hand, FSRA is an ensemble feature selection method which helps identify the significant set of features to detect point of attacks and affected elements in a CPS facility. Three CPS datasets are considered and FSRA tries to find the optimal number of features required to detect an attack. Among all the classifiers considered, XGBoost stands out with its good performance in terms of accuracy and F1-score.

The next chapter presents the concluding remarks and the future work.

Table 7.5: Top 10 ranked features as given by the proposed method

Top 10 ranked features as given by the proposed method						
	Gas Pipeline		Water Storage		SWaT	
	Feature Name	Index Number	Feature Name	Index Number	Feature Name	Index Number
1	time	25	measurement	21	PIT502	45
2	response_address	1	response_address	1	MV304	22
3	setpoint	18	H	14	FIT201	8
4	resp_read_fun	8	time	22	FIT502	39
5	measurement	24	response_memory	3	MV302	20
6	response_memory	3	HH	13	DPIT301	16
7	control_scheme	20	sub_function	10	AIT504	37
8	sub_function	10	command_address	0	MV303	21
9	control_mode	19	response_memory_count	5	P203	12
10	command_address	0	resp_write_fun	9	LIT301	18

Table 7.6: Comparison with Existing Methods

Comparison with other methods on SWaT dataset		Comparison with other methods on Gas Pipeline dataset		Comparison with other methods on Water Storage dataset	
Method / Proposed by	F1-score	Method / Proposed by	F1-score	Method / Proposed by	F1-score
DNN-based Method [280]	80.28%	CPS-GUARD [281]	93.70%	Morris and Gao[140]	98.10%
SVM-based Method [280]	79.62%	Morris et al. [141]	98.80%	GRYPHON [282]	98%
NN-one class method [283]	87.00%	Beaver et al. [284]	75%	SOCCADF [285]	98.10%
DIF [286]	88.20%	Proposed Method	89.3%	Proposed Method	89.2%
AE [287]	52%				
FB [287]	36%				
Proposed Method	99.4%				