Asset Criticality and Risk Prediction for an Effective Cyber Security Risk Management of Cyber Physical System

Halima Ibrahim Kure[1], Shareeful Islam[1], Mustansar Ghazanfar[1], Asad Raza[2], Maruf Pasha[3]
[1]School of Architecture, Computing and Engineering, University of East London, UK
[2]Abu Dhabi Poly Technic, institute of Applied Technology, UAE
[3]Department of Information Technology, Bahauddin Zakariya University, Pakistan
h.kure@uel.ac.uk, shareeful@uel.ac.uk, m.ghazanfar@uel.ac.uk, asad.raza@adpoly.ac.ae, maruf.pasha@bzu.edu.pk
Corresponding author: Maruf Pasha (Maruf.pasha@bzu.edu.pk)

Abstract: Risk management plays a vital role in tackling cyber threats within the Cyber-Physical System (CPS). It enables identifying critical assets, vulnerabilities, and threats and determining suitable proactive control measures for the risk mitigation. However, due to the increased complexity of the CPS, cyber-attacks nowadays are more sophisticated and less predictable, which makes risk management task more challenging. This paper aims for an effective Cyber Security Risk Management(CSRM) practice using assets criticality, predication of risk types and evaluating the effectiveness of existing controls. We follow a number of techniques for the proposed unified approach including fuzzy set theory for the asset criticality, machine learning classifiers for the risk predication and Comprehensive Assessment Model (CAM) for evaluating the effectiveness of the existing controls. The proposed approach considers relevant CSRM concepts such as asset, threat actor, attack pattern, Tactic, Technique and Procedure (TTP), and controls and maps these concepts with the VERIS community dataset (VCDB) features for the risk predication. The experimental results reveal that using the fuzzy set theory in assessing assets criticality, supports stakeholder for an effective risk management practice. Furthermore, the results have demonstrated the machine learning classifiers exemplary performance to predict different risk types including denial of service, cyber espionage, and crimeware. An accurate prediction of risk can help organisations to determine the suitable controls in proactive manner to manage the risk.

KEYWORDS: Cyber Security Risk Management, Risk Prediction, Machine Learning, Fuzzy theory, Feature Extraction, Control, and Cyber Physical System.

1. INTRODUCTION

The primary objective of CPS is resilience by delivering it's users an uninterrupted services based on relying on the most valuable assets such as information and communication networks, and digital data for reliable service delivery [1, 2]. These assets require the attainment of stability, reliability, efficiency which need tight integration of computing, communication and control technological systems [3]. However, CPS faces different types of cyber threats which are constantly evolving and more sophiscated which makes the risk management task more challenging [4]. A recent survey result from experian shows that almost half of business organisations suffer at least one security incident per year [5]. Therefore, global cybersecurity spending is continuously rising to 96 billion US dollars in 2018 [6]. Despite of the efforts for implementing controls to secure the CPS, large organisations are still facing cyber attacks which could pose severe business interruption. It is really challenging to eliminate the cyber

attacks, but organisations should aim to predicate the risks so that necessary actions can be taken for its mitigation. We advocate considering the risk prediction as a part of overall risk management practice. Machine Learning (ML) can particularly be beneficial for predicting risk. There are number of works that proposed prediction models which allowed for the adoption of preventive actions to avoid the disruption of critical services. These papers examined the demographics of users' and network connectivity behaviour [6], web browsing behaviour [7], website features [8], network mismanagement details [9]and historical incident reports of organisations [10] to predict cyber incidents. Despite of these contributions, there is a lack of focus on integrating ML for predicting risk types to support overall risk management process .Additionally, there is a need to determine the effectiveness of existing controls taking into account the predicated risks so that organisation can identify the additional controls to tackle the risks.

Within the above context, this paper contributes for an effective risk management practice and its novelty is in four folds. Firstly, we propose to use fuzzy logic to determine Asset Criticality (AC). In doing so, five primary security goals are used as input factors i.e. Confidentiality (C), Integrity (I), Availability (A), Accountability (ACC) and Conformance (CON). The AC is the fuzzy output based on the assessment outcomes of identified assets. Secondly, ML models such as K-Nearest Neighbours (KNN), Neural Networks (NN), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Naïve Bayes Multinomial (NB-Multi) and Naive Bayes (NB) are used to predicate the risk types. We extract the features based on CSRM concepts such as threat actor, assets, controls and TTP for the risk prediction. Thirdly, we consider Comprehensive Assessment Model (CAM) to determine the effectiveness of existing controls and propose additional controls to tackle the predicated risks. Finally, we use VERIS community database (VCDB) to predicate the risks. The result shows that asset criticality and risk predication can effectively support the overall risk management process. The result also confirms that some controls such as network intrusion, authentication, and anti-virus show high efficacy in controlling risks by following the CAM approach.

The rest of the paper is structured as follows. Section 2 we provide a brief introduction of the related works. Section 3 explains the concepts necessary for "cybersecurity risk management (CSRM)". In section 4, we introduced the experimental methodology for determining the cybersecurity risk type. Section 5 explains the experimental results obtained from the different classifiers using the VCDB dataset. Section 6 concludes our research and provides future work based on our findings.

2. RELATED WORKS

This section provides state-of-the-art contributions which are relevant with our work in the area of CSRM and ML classifiers for the overall cyber security .

2.1. Machine learning and cybersecurity

Machine learning classifiers are widely used in several application domains such as text categorisation[11], internet traffic classification [12], recommender systems[13], and malicious "uniform resource locator (URL)" detection [14]. However, the benefit of machine learning techniques for risk management is still at an early stage. In [15] proposed an intrusion detection system (IDS) for synchro-phasor systems that detect cyber-attacks but is limited to man-in-the-middle (MITM) and denial of service (DoS) cyber-attacks against synchro-phasor devices only. In the work of[16], they applied multiple learning algorithms to Modbus return terminal unit (RTU) data in order to demonstrate an ability to discriminate command and data injection attacks on the

supervisory control and data acquisition systems (SCADA) of a pure gas pipeline system. In [17], the authors proposed a Siamese Network Classification Framework (SNCF) that can map the Siamese network to a classification based on the similarity to alleviate imbalance for risk prediction. However, comprehensive evaluation for other ML classifiers was not carried out to see which one gives the best predictive accuracy result. The work of [18] presents a RiskTeller system that analyses binary file appearance logs of machines to predict future machines that are at risk of infection. However, the RiskTeller is only able to predict a risk level and the not the specific risk type. [19] Presents an algorithm model to predict cyber risks by using social media big data analytics and statistical machine learning. However, the proposed algorithm only uses vulnerability information to predict risk types, other features such as: TTP, IOC, and Assets etc. are not considered. In [20], the authors proposed a model that integrates fault tree analysis, decision theory and fuzzy theory to ascertain the current causes of cyber-attack prevention failures and determine the vulnerability of a given cybersecurity system. However, predicting risk type within a risk management framework is not the focus of this paper. In[21]a novel multi-model-based hazardous incident prediction approach is designed which has the ability to assess the risk caused by unknown attacks. However, the model has no ability for self-learning, and the sub second computation time cannot meet some hard real-time systems requirements. Machine learning techniques have been used in cloud computing for different purposes. In [22] the authors proposed a Periodicity-based Parallel Time Series Prediction (PPTSP) algorithm for large-scale time-series data and implemented in the Apache Spark cloud computing environment. The result shows that the PPTSP algorithm is significant in predicting accuracy and performance [23]. However, this algorithm has not been used for the risk type prediction. The authors in [24] used a realistic patient data to develop a patient treatment time consumption model. The model is developed based on important parameters to calculate different waiting times for different patients based on their conditions and operations performed during treatment.

2.2. Cybersecurity risk management (CSRM)

The authors in [25] discussed the challenges for securing critical infrastructure and analysed security mechanisms for prevention, detection and recovery, resilience and deterrence of attacks for securing CPS. In [26], a layered approach is proposed for evaluating risk based on security to prevent, mitigate and tolerate attacks both on real power applications and cyber infrastructures. In[1], the authors proposed a quantitative risk assessment model that provides users with attack information such as the type of attack, frequency, and target and source host identity. Authors in [27] proposed a new approach for critical infrastructure asset identification using multi-criteria decision theory to resolve the challenges of identifying critical assets. The approach didn't provide a systematic process for arriving at criticality decision. In [28] a framework that can automatically identify critical components and dependency structural risks is presented. The framework models the connections of assets and devices to depict their interdependencies on a company's business process to reduce their overall risk against cybersecurity threats. However, this framework doesn't predict risk types and implement effective control measures.

There are existing industry specific standards that focus on providing guidelines for risk management and cyber security improvement. NIST framework [29] is considered a practical approach to improving cyber security focusing on the the Critical Infrastrcuture (CI). The framework considers four implementation tiers(i.e., partial, risk informed, repeatable, adaptive)

to demonstrate the organisation view about cyber-security risks and the processes in place to manage those risks. The tiers consider three main components. i.e., risk management process, program and external participation. However, the framework doesn't provide any detailed guideline how the tier should be measured and move from lower tier to higher tier. The ISO 27005:2011 [42] standard provides for a detailed guideline for the information security risks management. The rationale for choosing these methods is that they are widely accepted standards for raising security awareness by identifying some of the most severe cyber-physical organisations' faces. ISO 27001 : 2017 [47] also emphasizes on the information security risk assessment and treatment process for the overall security management.

To summarise the literature mentioned above, there are several contributions that uses the ML approach in different application domains. However, a little effort is taken relating to how risk prediction can be integrated to support the risk management activities and adoption of the effectiveness of existing controls. The existing standards only provide a high level guideline for the risk assessment and management. Finally, there is a lack of guideline on how to determine the asset criticality for CPS. Our work contributes to address these limitations by proposing an effective CSRM approach based on asset criticality, risk prediction and effectiveness of security control.

3. Proposed Unified Approach

The unified approach aims for an effective risk management practice using asset criticality, risk predication and effectiveness of existing controls based on a number of CSRM concepts, as presented in Figure 1. This work extends our previous work [31] by integrating the use of fuzzy logic to determine critical assets and ML for the risk type prediction. Fuzzy logic is a powerful tool to handle the uncertainty and provides solution where there are no sharp boundaries and precise values. It provides a way of absorbing the uncertainty inherent to phenomena whose information is unclear and uses a strict mathematical framework to ensure the precision and accuracy. Additionally , it is flexible to deal with both the quantitative and qualitative variables [32]. The proposed approach considers Fuzzy logic based on the relative importance among the fuzzy input values (C, I, A, CON, ACC) by assessing individual assets to different levels of criticality and ranks the values within each output category simply by using the fuzzy inputs. The proposed approach also integrates the ML classifiers for the risk type predication. ML is particularly beneficial of using large data to discover hidden patterns. Hence, the unified approach aims to predicate the risk types that can potentially affect an organisation using the ML techniques.

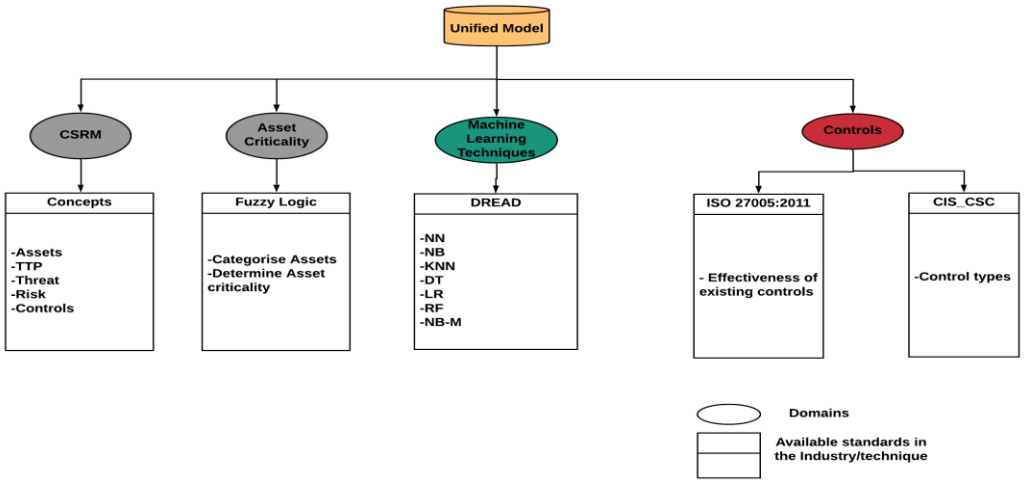

Figure 1: Proposed Approach

3.1. CSRM Conceptual view

Concepts serve as a common language for describing the properties necessary for CSRM to proactively assess and manage risks. This section presents the CSRM concepts and its unique properties that are important for risk prediction.

Actor: An actor is an entity, generally a human user, a system, an organization, or a process each with a specific strategic goal within its organizational setting and carries out specific activities to generate cybersecurity risk management actions or receive the generated cybersecurity risk management actions by another actor. Threat actor is a special type of actor with malicious intent. Their identity can characterise them such as suspected motivation, intended goals, skills, resources, past activities, tactics, techniques and procedures (TTP) used to generate a cyber-attack and their location (within a network, adjacent network, local network or physical) within the organisation. All these are unique properties of the threat actor that serve as features and passed to the classification algorithms for the process of risk prediction.

Assets: Assets are entities which are necessary and have values to the critical infrastructure organisation. The asset properties include server, network, media, people, terminal, user device. All these assets are aimed by threat actors to attack and cause a significant impact on the organisation. However, some assets are more critical than other assets and require a high level of controls because they are more likely to be attacked and when attacked they cause more loss to the organisation. So, predicting the risk type helps organisations to protect those assets way before any attack is carried out on them.

Goals: The goal of any CPS includes; the concealment of sensitive data against unauthorised users, ensuring the assets of the organisation are made available and accessible to the end-users, and the ability of the assets to perform their required functions effectively and efficiently without any disruption or loss of service. Therefore, this concept identifies the goals of each asset in terms of security and organisational context, and it is carried out by the security analyst.

Tactics Techniques and Procedures (TTP): TTP involves the pattern of activities used by a threat actor to plan and manage an attack, thereby compromising assets. They are used to help categorise attacks, generalise specific attacks to the patterns that they follow and provide detailed information about how various software tools perform attacks; they include malware, hacking, misuse, social and many other mentioned in section 4, which serve as the features for machine learning classifiers. In order to predict risk and to know the appropriate controls to be used to protect the assets of the organisation, information about TTP must be known.

Controls: These are the course of action taken either to prevent an attack or to respond to the attack in progress Centre for Internet Security and Critical Security Controls (CIS_CSC) provides basic controls that mitigate the most common attacks against systems and networks and achieve cybersecurity. We categorised the controls types into; detective controls designed to detect irregularities or errors which have already occurred and to assure immediate correction and corrective controls help to mitigate damage once a risk has materialised. Preventive controls are designed to keep errors or irregularities from occurring. This means that the level of attack determines the type of control to be used and the effectiveness of the existing controls is evaluated.

Indicator of Compromise: Indicator concept contains a pattern that can be used to detect suspicious or malicious cyber activity. They are detective in nature and are for specifying conditions that may exist to indicate the presence of a threat along with relevant contextual information. Organisations should be aware of the data associated with cyber-attacks, which are known as indicators of compromise (IOC). IOC is commonly partitioned into three distinct sub-classes. The sub-classes include network indicator, host-based indicator and email indicator. These sub-classes have their own sub-classes. For instance, email indicators have sub-class email attachment, email link. Network indicators have sub-class IP address.

Incidents: The incident is the type of event that represents information about an attack. The incident is defined by its types and linked with the indicator and the actor.

Vulnerability: Vulnerability is the weakness or mistake in an organisations security program, software, systems, networks or configurations that are targeted and exploited by a threat actor to gain unauthorised access to an asset (system or network) using TTP. It consists of sub-classes such as vulnerability types and assets targeted.

Threat: Threats are potential dangers that might exploit vulnerability within the critical infrastructure and cause possible harm to one or many asset components to deter security goals or hinder the business process. Each risk is associated with a specific threat, and the threats are categorised to evaluate their severity to assets. Also, threats are considered from different sources that elaborate more about security threats associated with critical infrastructure such as ENISA[25].

Risks: The risk is defined as the potential consequence of failure that obstructs the achievement of goals, which mainly caused by threat actors. Due to the evolving nature of the threat landscape, it is challenging for the organisation to mitigate all possible cybersecurity risks completely. It is the role of the actors to ensure that risks are kept to a minimum level to achieve the overall business continuity. It includes properties like type, level, and control.

The Meta-model, illustrated in figure 2 , shows the relationship among the concepts. The actor is represented as having an interest in the organisation's assets. The threat actor is a type of actor with malicious intent characterised by their motivation, skills, resources available to carry out a successful attack. Assets in general have security goals such as confidentiality, integrity and Availability and the attainment of the goals is based on the specific organisation context. Vulnerability is the weakness within the security program, software, systems, networks, or configurations targeted and exploited by a threat actor to gain unauthorised access to an asset (system or network) using TTP. Risk is the failure of an organisation or individual to achieve its goals due to the malicious attempt to disrupt its critical services by a threat. The incident is the type of event that represents information about an attack on the organisation. The components determine the type of incident include threat types, threat actor's skill, capability, asset, and location. With a specific attack pattern, the organisation tends to think broadly by developing a range of possible outcomes to increase their readiness for a range of possibilities in the future. With Indicators, a pattern that can be used to detect suspicious or malicious cyber activity is gathered. Finally, there are controls which aim to mitigate the risk.

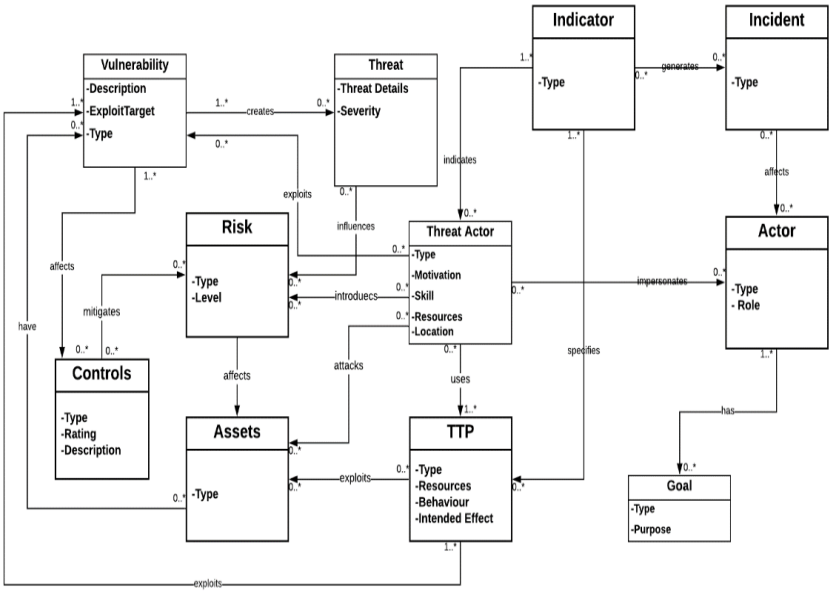

Figure 2. Conceptual Meta Model

3.2. Asset Criticality

The asset criticality is the first activity of the unified approach aims to identify and prioritise critical asset by assessing the primary security goals of those assets. The criticality assessment of all assets is carried out by a team of experts within the organisation. To ensure validity, consistency and support stakeholders in assessing the criticality of each asset, a decision support system using fuzzy set theory is created. Fuzzy set theory plays a vital role in the decision process enhancement it helps to deal with or represent the meaning of vague concepts usually in situation characterisation such as linguistic expressions like “very critical”. Fuzzy logic introduced by[26], is one of the best ways to deal with all the types of uncertainty including lack of knowledge or vagueness[27]. This section includes an running example to demonstrate how the asset criticality is determined.

3.2.1. Running Example

This running example is from the data set which is explained in the section 3.3.1. A highly skilled external attacker gained access to the master terminal unit (MTU) of the power grid system through a remote access point by exploiting the weak password and firewall. The attacker was able to disrupt communications, access database storing company and customer critical data such as passwords and operating plans as well as the SCADA system. Thereby, monitoring the status of the system and injecting malicious control commands as well as forging data into the control centre. This action led the system operators into taking inappropriate actions that interrupted the availability of electricity.

3.2.2. Development of a Fuzzy Asset Criticality System (FACS)

Criticality is the major indicator used to determine the importance of the assets to the organisation. After the different assets have been identified, we determine the criticality based on their relative importance using Fuzzy Asset Criticality System (FACS).

Fuzzification: FACS determines asset criticality by using (C, I, A, CON and ACC) as the five fuzzy inputs for assessing the criticality of individual assets and assigning level of criticality. Each input is assigned five fuzzy labels Very Low (VL), Low (L), Medium (M), High (H) and Very High (VH) for assessing the level of the fuzzy output Asset criticality (AC) value which is assigned five fuzzy labels Very Low Critical (VLC), Low Critical (LC), Medium Critical (MC), High Critical (HC) and Very High Critical (VHC) of individual assets. Table 1 shows the numerical ranges which fuzzy sets are selected based on them. The membership functions for AC also are depicted in a scale of 1 to 5.

Table 1: Fuzzy Ratings

Features	Asset Factors	Description	Linguistic Terms	Crisp Rating	Fuzzy Rating	Interpretation
Input	Confidentiality (C)	How much data could be disclosed and how sensitive is it?	Very High (VH)	5	$3.5 \leq C \leq 5$	All data disclosed
			High (H)	4	$2.5 \leq C < 5$	Extensive critical data disclosed
			Medium (M)	3	$1.5 \leq C \leq 4.5$	Extensive non-sensitive data disclosed
			Low (L)	2	$1 \leq C \leq 3.5$	Minimal critical data disclosed
			Very Low (VL)	1	$1 < C \leq 2.5$	Minimal non-sensitive data disclosed
	Availability (A)	How many services could be lost and how vital is it?	Very High (VH)	5	$3.5 \leq A \leq 5$	All services completely lost
			High (H)	4	$2.5 \leq A < 5$	Extensive primary services interrupted
			Medium (M)	3	$1.5 \leq A \leq 4.5$	Extensive secondary services interrupted

			Low (L)	2	$1 \leq A \leq 3.5$	Minimal primary services interrupted
			Very Low (VL)	1	$1 < A \leq 2.5$	Minimal secondary services interrupted
	Integrity (I)	How much data could be corrupted and how damaged is it?	Very High (VH)	5	$3.5 \leq I \leq 5$	All data corrupt
			High (H)	4	$2.5 \leq I < 5$	Extensive seriously corrupt data
			Medium (M)	3	$1.5 \leq I \leq 4.5$	Extensive slightly corrupt data
			Low (L)	2	$1 \leq I \leq 3.5$	Minimal seriously corrupt data
			Very Low (VL)	1	$1 < I \leq 2.5$	Minimal slightly corrupt data
	Accountability (ACC)	Are the threat actors traceable to an individual?	Very High (VH)	5	$3.5 \leq ACC \leq 5$	Completely anonymous
			High (H)	4	$2.5 \leq ACC < 5$	Fully traceable
			Medium (M)	3	$1.5 \leq ACC \leq 4.5$	Highly traceable
			Low (L)	2	$1 \leq ACC \leq 3.5$	Possibly Traceable
			Very Low (VL)	1	$1 < ACC \leq 2.5$	Minimal Traceable
	Conformance (CON)	How much deviation from specified behaviour constitutes conformance?	Very High (VH)	5	$3.5 \leq CON \leq 5$	Full variation
			High (H)	4	$2.5 \leq CON < 5$	High profile variation
			Medium (M)	3	$1.5 \leq CON \leq 4.5$	Clear variation
			Low (L)	2	$1 \leq CON \leq 3.5$	Low variation
			Very Low(VL)	1	$1 < CON \leq 2.5$	Very low variation
Output	Asset Criticality (AC)	How critical is the asset to the organisation?	Very Critical (VC)	5	$3 \leq AC < 5$	Extremely critical and high value to the Critical Infrastructure(CI) organization, requires an extreme level of protection
			Highly Critical (HC)	4	$2 \leq AC < 4$	High importance to the organization and requires a high level of protection.
			Medium Critical (MC)	3	$1 \leq AC \leq 3$	Moderately important to the organization and requires moderate protection

			Low Critical (LC)	2	$0 < AC \leq 2$	Minimal importance and does not require many levels of protection.
			Very Low Critical (VLC)	1	$0 \leq AC \leq 1.5$	Non-critical and requires a very low level of protection

Rules : We follow the Min–Max fuzzy inference method proposed by Mamdani due to the following advantages [28, 29].

- It is suitable for engineering systems because its inputs and outputs are real-valued variables
- It provides a natural framework to incorporate fuzzy IF–THEN rules from human experts
- It allows for a high degree of freedom in the choices of fuzzifier, fuzzy inference engine, and defuzzifier, so that the most suitable fuzzy logic system for a particular problem is obtained. It provides a natural framework to include expert knowledge in the form of linguistic rules.

We used 125 IF-THEN rules to provide a database by mapping between five input parameters (C, A, I, CON and ACC) and AC value. The rules are designed to follow the logic of the Asset criticality evaluator. A number of the IF-THEN rules of the developed system are shown in Figure 3.

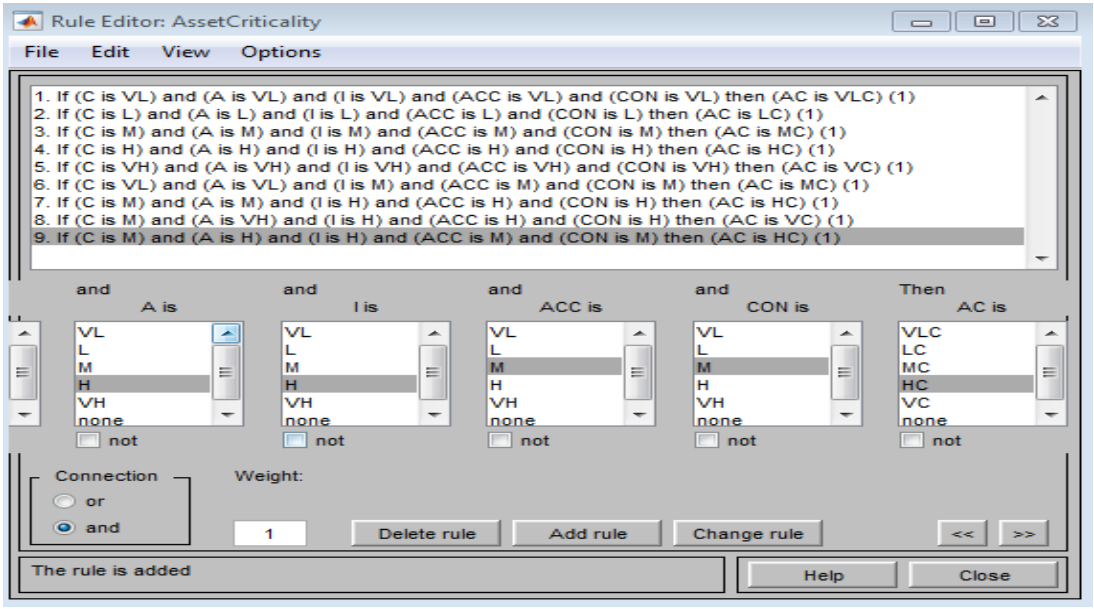

Figure 3: Rules Set for FACS

Inference Engine: An inference engine attempts to create solutions from the database. In this paper, the inference engine maps input fuzzy sets (C, A, I, ACC and CON) into fuzzy output set (AC). For example, if SCADA system is given the input values (C=1), (I=2), (A=1), (ACC=3), (CON=4) by an assessor, the output value will be (AC = "2.5") using the FACS. Traditionally, different sets of fuzzy input (C, A, I, ACC, CON) may generate an identical value of the fuzzy

output (AC); however the assets may not necessarily be the same. Figure 4 shows a number of IF-THEN rules in order to provide a more understanding the proposed FACS model.

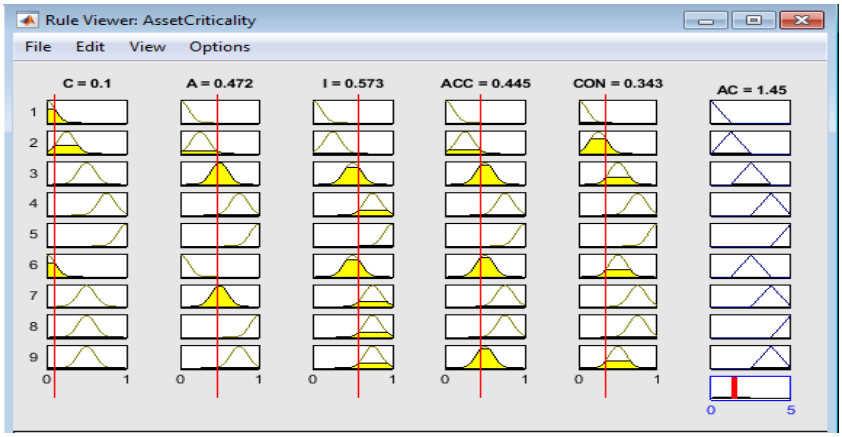

Figure 4: Sample of Rules

Defuzzification: There are different methods for converting the fuzzy values into crisp values such as, Centre of Gravity (COG), Maximum Defuzzification Technique and Weighted Average Defuzzification Technique. One of the most commonly used defuzzification method is COG. The COG technique can be expressed as follows, where x* is defuzzified output, $\mu_i(x)$ is aggregated membership function and x is the output variable.

$$X^* = \frac{\int \mu_i(x) x \, dx}{\int \mu_i(x) \, dx} \qquad (2)$$

3.3. Machine learning classifiers for predicting cybersecurity risk

As mentioned before, the proposed unified approach considers ML classifiers for the risk predication. This section provides an overview of the risk predication. Figure 5 shows how the features are extracted from the data sets and used by ML classifiers for the purpose of risk prediction. The data were partitioned into 80% training and 20% testing. We used the widely known 5-fold cross-validation scheme to split the given data into testing and training set and reported the average results obtained over the five folds. Predictions are carried out on the testing dataset and accuracy measures the prediction. The performance of each algorithm is being assessed. This section considers three sequential steps for prediction of the risk type and provides an overview of the experimental steps.

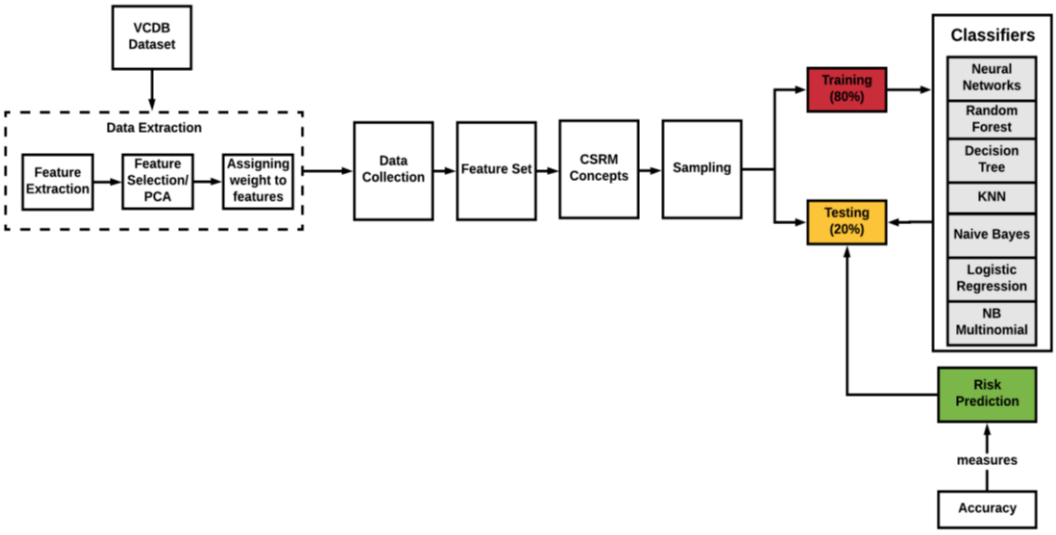

Figure 5: Feature Extraction and Risk Prediction

3.3.1. Dataset Description

We used the dataset from "Veris Community Database (VCDB)"[30] which aims to collect and disseminate data breach information for all publicly disclosed data breaches, to test our classifiers. It provides some of the enormous available collection of datasets that consists of a collective intelligence report datasets allowing us to test the performance of the classifiers in predicting risk type. We further created a mapped version of this dataset by selecting some features in the dataset and mapping them to TTP, Threat Actor, Asset and Control categories. For example brute force is mapped to TTP. We extracted the features in VCDB that are of interest in training and testing our classifiers. A validation team is formed to support this mapping The total features are 1,122 and the sample size is 7,834. In [31], data on reported cybersecurity incidents are needed to serve as ground-truth for their study. Such data is required to train the classifiers as well as assess its accuracy in predicting incidents. Therefore data from the VCDB is collected to obtain proper coverage. In [32] VCDB is used to train and test a sequence of classifiers/predictors. The data for each CSRM feature is mapped to the VCDB dataset as shown below:

- **Discovery and Response:** This entry in the VCDB dataset is our Control feature. It focuses on the timeline of the events and how the incident was discovered. It provides useful insights into the detection and defensive capabilities of the organisation and helps identify corrective actions needed to detect or prevent similar incidents from occurring.
- **Incident Description:** this entry is mapped to our Threat Actor, TTP and Assets features. It focuses on "whose actions affected the assets", what actions affected the assets" and which assets were affected". Threat Action (TTP) describes what the Threat Actor did to cause or contributes to the incident such as Malware, Hacking and Misuse. Actors (Threat Actors) are entities that cause or contribute to any particular incident and their actions can be malicious, intentional or unintentional. Threat Actors are recognised in VCDB as external, internal and partner. Assets (Assets) describe the information assets that were compromised during an incident. Compromised means the loss of confidentiality, integrity, availability and authenticity. Assets are categorised into Variety (such as SCADA), Ownership, Management, Hosting, Accessibility and Cloud.

3.3.2. Mapping

In this section, we explain how we map the existing dataset features to the CSRM concepts. The features extracted from VCDB are used for training and testing our classifiers. Details documented in the incidents include the TTP used, assets compromised, threat actor type and motive and controls in place. The list of features extracted from the VCDB dataset that are mapped to CSRM concepts are shown in tables 2, 3, 4 and 5below.

- **Threat Actor:** The first set of mapping is information regarding the threat actor including individual, group of individuals or organisations that are believed to have operated with malicious intent, as shown in Table 2. Therefore, each incident is put in one of the four categories: External, internal, partner and unknown threat actor types. Each category includes additional features that further differentiate the threat actor type. For instance, an external threat actor is further categorised as organised crime, former employee, competitor, espionage and grudge. The Partner is further categorised as the industry. The internal threat actor is categorised as hired, demoted, personal issues, resigned, auditor, cashier and developer. Therefore, we train our classifiers based on the threat actor responsible for the incident. Predicting risk requires information about the threat actor type and motive, this allows organisations to determine the policies to educate their employees, access to their data, safeguard their networks from attackers and perform due diligence when selecting partners as third party.

Table 2: Feature vector for threat actor for VCDB dataset

Threat Actor Type	Espionage	Competitor	Grudge	System Admin	Financial	Fun	End-User	……	Developer
Number of features	1	2	3	4	5	6	7	……...	80

- **Assets:** The asset mapping considers six categories of asset types: server, media, user device, terminal, people and networks, shown in Table 3. Knowing the type of assets that are more likely to be affected can help organisations to improve their ability to predict risk following security incidents significantly. Organisations can further implement appropriate controls such as network administrators keeping regular backups on media and server assets.

Table 3: Feature vector for asset for VCDB dataset

Asset Type	Disk drive	Documents	Access reader	LAN	Router/Switch	Patch Management	RTU	.....	Database
Number of features	1	2	3	4	5	6	7	…...	234

- **TTP:** This set of mapping relates with the type of attack the threat actor exploited taking into account the Tactic, Technique and Procedure(TTP). We consider seven general categories of TTP including Environmental, error, hacking, malware, misuse, physical and social, as shown in Table 4. Each category of TTP includes additional features that can help to differentiate incidents further. For instance, SQL injection and brute force are identified as hacking. Hacking incidents involve data breach through compromised credentials. Knowing the TTP type can provide organisations with valuable information on the types of preventive measures to be used to reduce risk.

Table 4: Feature vector TTP for VCDB dataset

TTP Type	Remote access	Ransomware	Remote injection	SQL injection	Spyware/ keylogger	Brute force	Buffer overflow	…	Email attachment
Number of features	1	2	3	4	5	6	7	…	155

- **Controls:** The control types fall into one the two categories detective and corrective controls , as shown in Table 5. We train our classifiers based on the controls available at the time of the attack. We further categorise detective into sub-categories: Internal (log review, antivirus, data loss prevention, fraud detection) and external (actor disclose, incident response, monitoring service, suspicious traffic). Assessing the risk associated with controls prompts organisations to determine the set of security protections or countermeasures further to minimise risk. Some of the controls might be insufficient to mitigate risk, so, these different control types that were compromised at the time of the attack are the properties that serve as features for machine learning classifiers to predict risk type and appropriate controls implemented.

Table 5: Feature vector for control for VCDB dataset

Control Type	Fraud detection	Incident response	Monitoring service	Anti-virus	IT Review	Log Review	Security alarm	….	Law enforcement
Number of features	1	2	3	4	5	6	7	….	42

3.3.3. Experimental Setup

In our experiments, we used VCDB dataset because it has been used in literature providing easier benchmarking and we have feature information about cybersecurity. Further, in our experiments we used PyCharm and python 3.6 interpreters to run our codes.. The procedure works as follows: the dataset is divided into sub samples. A sample is chosen as testing data and the remaining sample as training data.

3.3.4. Feature Extraction

Feature extraction is the first step to start a machine learning process because it is a technique that aims at finding specific pieces of data in natural language and then converts them into a suitable format for machine learning classifiers. Our research draws from a variety of data sources that collectively characterise the security posture of organisations as well as the security incident report used to determine their security outcomes. In this step, we extract all the necessary features from the dataset to map our CSRM concepts, which are presented in the previous section. Every concept has properties, and those properties are considered as features, for example:

- Asset concept features include; Server, media, people, networks, user device and terminal.
- Threat actor features include; External, Internal and supply chain partner.
- Control features include; corrective, detective and preventive.
- TTP features include; Malware, hacking, social, physical, environmental, misuse and error.

The features are further converted into a format suitable for the machine learning classifiers by assigning a weight between 1 and 0.

3.3.5. Features and classification labels

This step presents the values of the data type used in the experiments and includes a list of features extracted from the dataset. The reason for choosing these feature types is because they are salient, straightforward and intuitive, and any machine learning classifier can be trained over them. Asset,

threat actor and controls are assigned binary numerical data type and given a possible value between 0 and 1. It consists of two sub-steps.

3.3.5.1. Features weights and labels

Dataset is collected from the "Veris Community Database (VCDB)"[30]. We then mapped the features in the dataset to the CSRM concepts, which are used as features for the classification and assigned their weights. We used the feature extraction techniques coupled with human annotation for extracting the essential features from the dataset. The risk type is the output class we are predicting; an ordinal categorical data type is used with possible values from 1 to 10.(Refer to table 13).

Output Feature

We have used 10 output categories of risks and the value range for the features is from (R_1 = Crimeware, R_2 = Cyber espionage, R_3 = Denial of service, R_4 = everything else, R_5 = lost and stolen assets, R_6 = miscellaneous errors, R_7 = payment card skimmers, R_8 = point of sale, R_9 = privilege misuse and R_{10} = web applications) with possible classes. This is a multi-class problem and we have the following risk types as output features explained in Table 6. The input features are shown in tables 7, 8, 9 and 10. These features are used to categorise the input features (threat actor, control, assets and TTP) into ten categories. The classification model is trained on the following categories listed in the table below:

Table 6: Feature vector as output features for control for VCDB dataset

Feature name	Possible classes	Range of values
Crimeware Cyber Espionage Denial of Service Everything Else Lost and Stolen Assets Miscellaneous Errors Payment Card Skimmers Point of Sale Privilege Misuse Web Applications	$\boldsymbol{R} = \{R_1, R_2, R_3 \ldots R_{10}\}$ Where: $\boldsymbol{R_1}$ = Crimeware $\boldsymbol{R_2}$= Cyber Espionage $\boldsymbol{R_3}$= Denial of Service $\boldsymbol{R_4}$= Everything Else $\boldsymbol{R_5}$= Lost and Stolen Assets $\boldsymbol{R_6}$= Miscellaneous Errors $\boldsymbol{R_7}$= Payment Card Skimmers $\boldsymbol{R_8}$= Point of Sale $\boldsymbol{R_9}$= Privilege Misuse $\boldsymbol{R_{10}}$= Web Applications	$\{1,2,..10\}$

Input features

We consider different classes of input feature such as threat actor, asset, TTP, and control. Table 7 shows the threat actor feature types with possible classes $\{t_1, t_2, t_3\}$, which represents the different threat actor feature types. They are trained on the proposed classifiers, and the possible values are between$\{0, 1\}$.

Table 7: Threat Actor type feature detail

Feature name	Possible classes	Range of values
External, Internal Supply chain Partner	$\boldsymbol{t} = \{\boldsymbol{t_1}, \boldsymbol{t_2}, \boldsymbol{t_3}\}$ $\boldsymbol{t_1}$ = External	$\{0, 1\}$

	t_2= Internal t_3= Partner	

Table 8 shows the different asset feature types used as an input parameter for risk type prediction. The asset features are given as {A_1, A_2…A_6} representing the different asset types and are trained on the proposed classifiers. The possible values are between$\{0, 1\}$.

Table 8. Asset type feature detail

Feature name	Possible classes	Range values
Server, Terminal, Media People, Networks,User device	$\boldsymbol{A} = \{A_1, A_2 \ldots A_6\}$	$\{0, 1\}$

Table 9 below shows the list of the control feature types extracted from the applied dataset. The control features include the different control types such as detective, corrective and preventive which are used as part of the input parameters for predicting risk type. These types allow to choose the right control actions for mitigating the risks. They include {c_1, c_2, c_3} representing the different types of control types with possible values between$\{0, 1\}$.

Table 9. Control type feature detail

Feature name	Possible classes	Range of values
Detective Corrective Preventive	$\boldsymbol{C} = \{\boldsymbol{c_1}, \boldsymbol{c_2}, \boldsymbol{c_3}\}$ Where: $\boldsymbol{c_1}$ = Detective $\boldsymbol{c_2}$= Corrective $\boldsymbol{c_3}$= Preventive	$\{0, 1\}$

The list of the TTP features extracted from the applied dataset is shown in Table 10. The different TTP feature types used as input features that are trained on the classifiers. These are the possible technique that the threat actor can be exploited to attack any system. They are given possible values as {TTP_1, TTP_2….TTP_7} which represents the different TTP feature names and are given possible values between$\{0, 1\}$.

Table 10: TTP type feature detail

Feature name	Possible classes	Range of values
Malware, Hacking, Social, Physical Misuse, Error, Environmental	$\boldsymbol{TTP} = \{\boldsymbol{TTP_1}, \boldsymbol{TTP_2}, .., \boldsymbol{TTP_7}\}$	$\{0, 1\}$

3.3.5.2. Assigning Weights to Feature Vectors

This section presents the features used for the experiment. There are five feature vectors considered for the experiment and each feature vector is assigned with binary values of either 0 or 1 {0, 1} as shown in Table 11. A model can easily be trained over these feature vectors, which can predict any risk type.

Table 11. Feature vector weights

Feature Vector	Feature vector Weights

TTP Control TA Asset Full	$V_B \in \{0,1\}$For a given feature vector $'F'$, the value $'v'$, of any feature $'x'$ is determined using the following rule: $v_B^x = \begin{Bmatrix} 1, if occur \\ 0, otherwise \end{Bmatrix}$ The output value for any feature is $\{1\}$ in case the corresponding feature occurs in the dataset. Otherwise, its value is recorded$\{0\}$

3.3.6. Classification

Classification is an essential step for machine learning to understand and assign data categories for accurate risk prediction. Once we have extracted all the features, the next step is to classify the features. In order to achieve the classification, we follow seven different algorithms to generalise our findings of integrating machine learning with CSRM to predict a certain risk type. The classifiers calculate both the likelihood and impact and can find complex relationships in data to produce better results than rest. To manually evaluate the effectiveness of each feature, we created four different partitions of the datasets as described above. Once the feature weights are defined, we train the machine learning classifiers over the training data. For given partition and classifier, the results are shown using the following notation: (refer to Table 12)

Table 12. Classification Models and feature description

Scenarios	Assets	Controls	Threat actor	TTP	Models
Set_B PCA is not applied	$Asset_B \in \{0,1\}$	$Control_B \in \{0,1$	$TA_B \in \{0,1\}$	$TTP_B \in \{0,1\}$	$Model_{Set_B}^{Classifier}$ where $Set_B \in$ {Assets, Controls, Threat actor, TTP}
Set_PCA_B PCA is applied	$Asset\_PCA_B \in \{0$	$Control\_PCA_B \in$	$TA\_PCA_B \in \{0,1\}$	TTP_{PCA_B} $\in \{0,1\}$	$Model_{Set\_PCA_B}^{Classifier}$ where $Set\_PCA_B \in$ {Assets, Controls, Threat actor, TTP}

We used the notation,Set_PCA to denote the feature sets that have been reduced by applying PCA. For example, the feature vector control is denoted by $Control_B$ moreover, in case, this feature vector has been transformed by PCA; we denote it by $Control\_PCA_B$. Similarly, the model built over feature set transformed by PCA is denoted by$Model_{Set_PCA}^{Classifier}$.

3.3.7. Training the machine learning classifiers

This section describes the training of machine learning classifiers using training data. We use extracted features enclosed in the training examples to find a model$\boldsymbol{M}: \boldsymbol{D} \rightarrow \boldsymbol{R}$, which approximates$\boldsymbol{T}$. The function $\boldsymbol{R}$ defines the class to which the learned model assigns the given sample d and is used for classification of new scenarios. The model $\boldsymbol{M}$ $(\boldsymbol{d})$ denotes a machine

learning classifier. The objective here is to find a model, which maximizes the accuracy (assigns a scenario to the most proper class).

Table 13: Notations used for building the classifier

Notation	Description
D	The collection of cyber-attack scenarios
$d' = \{d_1, d_2, .., d_N\}$	N number of scenarios to be classified
$R = \{R_1, R_2, R_3, R_4, R_5 \ldots R_{10}\}$	R is the number of possible risks categories
$d' = \{d_1, d_2, .., d_N\}$	The training set consisting of N scenarios with corresponding actual class labels $y = R = \{R_1, .., R_{10}\}$
T	A target concept T: D → R, which maps given a scenario to a class (we assume the categories are disjoint, i.e. each given scenario can only be categorized into one of the categories, and there is no overlapping between categories)
$M: D \rightarrow R$	A machine learning model, which approximates T (i.e. close to T)
$M\ (d)$	The model predicts unknown scenario 'd' (i.e. using a classification algorithm)

The Accuracy matric can be formally defined as:

$$Accuracy\ = \frac{\sum_{x \in d'} 1_{M(d)=R_d}}{|d'|} (3)$$

Where $|d'|$ is the size of the test set (number of scenarios to be classified), and $1_{M(d)=R_d}$ Is an indicator function that output one if the model predicted the class for test scenario is the same as actual test class and zero otherwise. Formally:

$$1_{M(x)=R_x} = \begin{cases} \text{One if } M(x) = R_x \\ 0 \quad \text{otherwise} \end{cases} \quad (4)$$

The right controls also increase the accuracy score, which corresponds to the low rate of a classification error.

3.3.8. Evaluation measures

There are several parameters used to measure the evaluation including precision, recall and F-1. These metrics are used for validating accuracy in different ways, yet they can be applied to other purposes also and are useful in describing how risk prediction methods are successful.

The precision gives us the probability that a selected value is true. It can be formally defined as:

$$\boldsymbol{Precision\ = \frac{True\ Positive}{Total\ predicted\ positive}} (5)$$

The Recall gives us the probability that the true value is selected. It can be formally defined as:

$$\textbf{Recall} = \frac{\textbf{True Positive}}{\textbf{Total Actual Positive}} (6)$$

The F1 Score is a function of the precision and recall and can be formally defined as:

$$\boldsymbol{F1} = 2 * \frac{\boldsymbol{Precision} * \boldsymbol{Recall}}{\boldsymbol{Precision} + \boldsymbol{Recall}} (7)$$

3.4. Determine the Control Effectiveness

The final step of the unified approach involves assessing the effectiveness of existing controls, determining the necessity of additional controls. The existing controls are assessed to ensure the its effectiveness. If a control is not effective , this may cause vulnerabilities leading to any potential risk. Therefore, the consideration should be given to the situation where a selected control fails in operation, whether there is need for additional controls for addressing the identified risk. There are various industry standards that provide recommendations on basic security controls were considered. For example, Critical Security Controls [33] publishes a set of 20 controls and best practice guidelines that organisations should adopt to control security risks. In assessing the effectiveness of existing controls, an assessment of each control objective is carried out by an assessor team. The controls are evaluated in terms of relevance, strength, coverage, integration, and traceability for this purpose by following the Comprehensive Assessment Model (CAM) as presented in [34] . The control rating and overall control effectiveness are in accordance with ISO 27005:2011 [41]. For each criterion, a rating score from 1 to 5 is given to measure which control addresses the specific control objective. Table 14 shows the five different criteria rating by following the CAM, which are adopted and reformed to the context of the critical infrastructure to serve as a control criteria for assessing the effectiveness of existing controls.

Table 14: Control criteria

Criteria	Description
Relevance	The level to which the control addresses the relevant control objectives under analysis.
Strength	The strength of the control is determined by a series of factors
Coverage	The level in which all significant risks are addressed.
Integration	The degree and manner in which the control reinforces other control processes for the same objective
Traceability	How traceable the control is, which allows it to be verified subsequently in all respects

The assessemnt is helped by criteria, each criterion, a rating score from 1 to 5 is given to measure which control addresses the specific control objective. Table 15 shows the five different control rating while table 16 shows the overall effectiveness of the controls.

Table 15: Control rating

Rating	Description

5	Adequate control	The control achieves the objectives intended to mitigate the risks.
4	Adequate control with some areas of improvement	The control achieves the objectives intended to mitigate the risks with evidence of some areas, though not critical, subject to improvement to meet the requisites of sound controls.
3	Generally adequate control, with some critical areas	The control mostly mitigates the risks intended to mitigate the risks. However, the characteristics of some of the controls are not entirely consistent with basic sound controls
2	Inadequate control, subject to significant improvement	The control partially achieves the control objectives intended to mitigate the risks
1	Insufficient control	The control is not sufficient to achieve the control objectives intended to mitigate the risks.

To find the overall evaluation of each control, equation four below is given:

$$\boldsymbol{OE = R + S + C + I + T} \quad (8)$$

Where:

OE = Overall Effectiveness , R = Relevance , S = Strength, C = Coverage , I = Integration T = Traceability

Table 16: Overall control effectiveness

Description	Overall Effectiveness
Insignificant	0-5
Minor	6-10
Moderate	11-15
Major	16-20
Critical	21-25

4. Experimental Finding and Discussion

This section presents the criticality level determined and assessed for assets in greater detail as part of asset criticality using the running example. It further shows the experiment results obtained from the different ML classifiers using the mapped CSRM features from the VCDB dataset to perform risk prediction. The aim of the experiment is to explore the ability of machine learning classifiers to:

- Predicate a risk type out of the ten risks
- Determine the accuracy of each of the classifiers in predicting risk type.

Six machine learning classifiers were used for the classification process. We have formed an assessor team and used the dataset that was selected from VCDB dataset with the objective of evaluating the performance of the classifiers in predicting known risks for future occurrence and how this can help in improving the classification accuracy. Lastly, we evaluate the effectiveness of existing controls and recommend new controls.

4.1. Asset Criticality

The result of the asset criticality is presented based on the running example presented in section 3.2.1. We follow the Fuzzy Asset Criticality System (FACS) to determine the asset criticality and result is shown in Table 17.

Table 17: Asset criticality results

Asset Name	Asset Description	Asset Goals					Fuzzy output	Asset Criticality Level
		Fuzzy input						
		C	A	I	CON	ACC		
Master Terminal Unit (MTU)	A controller that acts as a server that hosts the control software that communicates remote terminal units and programmable logic controllers over a network.	1	3	4	4	1	2.5	MC
Databases	Stores information about the organisations customers, personnel, marketing, transactions, assets, finances, and other information about the organisations business process.	4	4	3	4	5	4	HC
Company and customer data	Sensitive and private information about employees, finances, assets, etc.	3	3	3	4	4	3.5	MC
Firewalls	Network security system that monitors and controls incoming and outgoing network traffic.	1	3	3	1	1	2	LC
SCADA Systems	Gathers and analyses real time data	2	5	5	1	4	3	MC

4.2. Machine Learning Results

This section explains the experimental results obtained from the different classifiers using the datasets.

4.2.1. Risk type Prediction Result

Table 18 presents the accuracy performance details of the six classifiers in predicting the different risk types based on the given CSRM features (Assets, Controls, Threat Actor and TTP). Based on the Asset features, LR, DT and NB-Multi achieved 95%, 93% and 92% respectively for predicting

risk type “Lost and Stolen Assets”, “Everything Else”, “Crimeware”, “Cyber Espionage” and “Denial of Service”. They failed to identify risk types “Point of Sale” and “Web Application”. RF, KNN and NB achieved 87%, 86% and 71% respectively for predicting risk type “Crimeware”, “Cyber Espionage” and “Lost and Stolen Assets”. NN failed to predict any risk type and achieved 4%. Based on the TTP features, KNN, LR, NB-Multi and DT achieved an accuracy of 80% for predicting risk type “Denial of Service”, “Cyber Espionage” and “Everything Else”. RF achieved an accuracy of 72% for predicting risk type “cyber espionage” and “Everything Else”, NN failed to predict any risk type and achieved 4%. Based on the Threat Actor features, LR, NB-Multi and RF achieved 79% accuracy for predicting risk type “Everything Else”, “Cyber Espionage” “Privilege Misuse” and “Crimeware”. KNN could predict risk type “Everything Else”, “Cyber Espionage” and “Privilege Misuse” while DT could predict risk type “Everything Else”, “Cyber Espionage” and “Crimeware” both classifiers with 76% accuracy. The NB achieved 63% accuracy for predicting risk types “”Cyber Espionage” and “Privileged Misuse”. NN achieved 3% accuracy and failed to predict any risk type. Lastly, based on the control features, KNN achieved the highest accuracy of 40% in predicting risk type “Everything Else”. LR, DT, NB-Multi and RF achieved 39% for predicting risk type “Everything Else”. NB and NN achieved an accuracy of 5% and 3% respectively. Both classifiers failed to predict any risk type. Asset and TTP features performed well on all the different classifiers except NN. Comparing the performance of all the features, it shows that NB failed to perform risk type prediction based on control features and NN achieved very low risk type prediction based on all the features. Therefore, for the risk types “Everything Else”, “Privilege Misuse”, “Denial of Service” and “Cyber Espionage” all the input features achieved high prediction. Table 18shows that Asset and TTP are the best features to predict risk types presented in this work.

Table 18: Performance of the features on each of the classifiers for predicting risk types

Accuracy	**Risk Type Prediction Features**			
	Asset	**TTP**	**Threat Actor**	**Control**
LR	95%	80%	79%	39%
DT	93%	80%	76%	39%
NB-Multi	92%	80%	79%	39%
RF	87%	72%	79%	39%
KNN	86%	80%	76%	40%
NB	71%	56%	63%	5%
NN	4%	4%	3%	3%

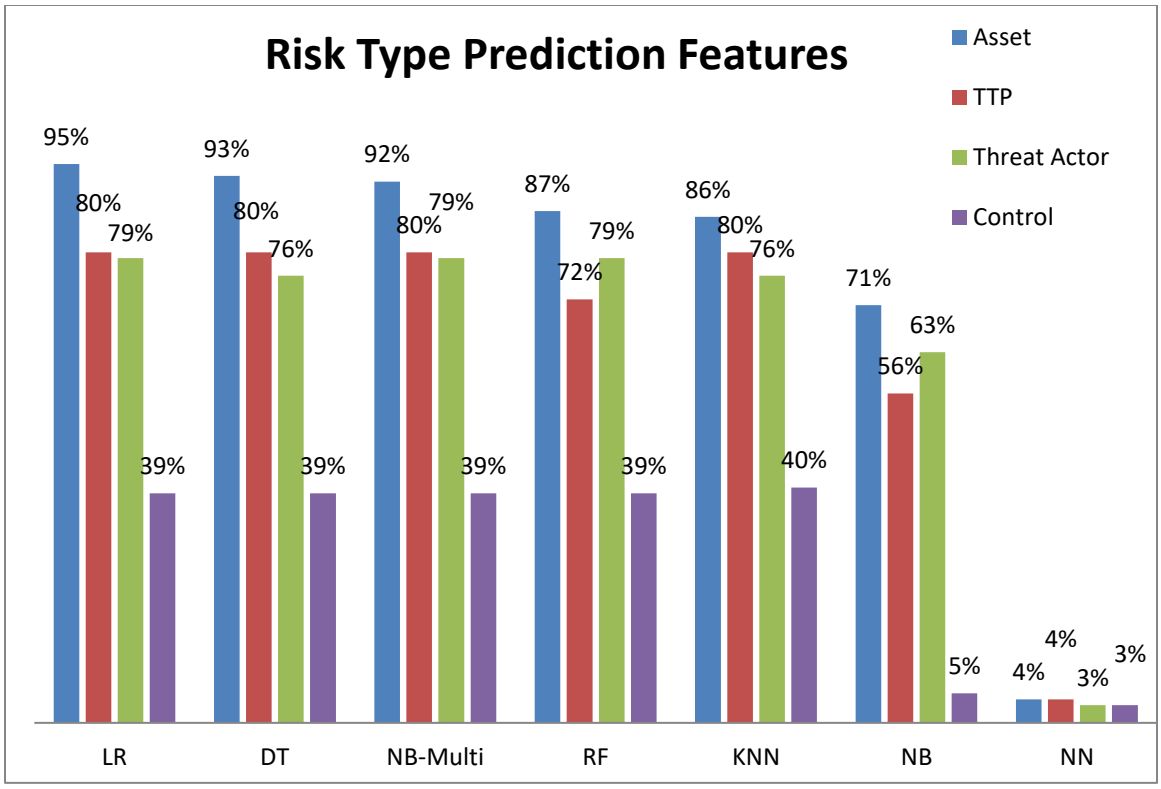

Figure 6: Performance of the features on each of the classifiers for predicting risk types

4.2.2. Prediction Accuracy

Once the risk type is predicated, the next step is to interpret the accuracy result of the different classifiers for various types of input features. The best overall predictive accuracy including all input features is recorded with Decision Tree (DT) for Asset is (92.92%) , Controls (79.26%), TTP (62.73%), Threat Actor (61.32%), and Full features (39.12%). The next best algorithm is NB Multi which gave us (91.90%) on asset , control (78.88%), threat actor (61.33%), TTP (59.54%) and full features gave us (39.05%). The third best algorithm is RF, it performed well on Asset features with (87.36%), control (78.75%), TTP (62.03%), Threat Actor (61.01%) and full features (38.93%). The fourth best algorithm is KNN, it performed well on almost all the input features, Asset features (85.77%), Controls (67.96%), TTP (58.07%), Threat Actor (56.80%) and the full features produced the least accuracy with (29.99%). The fifth best algorithm is the NB algorithm that performed well on the asset features with (71.03%), controls (55.90%), Threat Actor (19.85%), TTP (18.38%) and full features with (05.42%). The sixth algorithm which is NN didn't perform well on all the features, control features is (04.02%), Asset features is (03.51%), Full feature is (03.32%), TTP (03.13%) and threat actor (03.06%). This shows that the Asset features performed well with DT (92.92%), NB-Mult (91.90%), RF (87.36%), KNN (85.77%) and NB (71.03%). NN did not perform well with (03.51%). The control features also performed well with DT (79.25%), NB Multi (78.88%), RF (78.74%) and KNN (67.96%). On the other hand, Neural Networks (NN) and Naïve Bayes (NB) did not make satisfactory prediction accuracy on all the features. It can be noted that the most prominent features to detect risk types are Assets and control features. The result clearly shows that DT outperformed other classifiers giving the highest satisfactory accuracy for the VCDB dataset for risk type prediction.

4.2.2.1. Results of the different classifier for the input features

Figure 7 shows the accuracy results of different classifiers for the various kinds of input features. The most prominent features to detect the risk type are found to be Assets and Controls where accuracy is above 70%. From left to right (top to bottom), X-axis denotes different types of classifiers and Y-axis denotes the corresponding accuracy for a given feature set. It can be seen from the descriptive result shown in figure 8 below that based on the asset features KNN, NB Multi, RF and DT have produced the most accurate predictions by giving the accuracy value of above 70% compared to NB and NN classifiers. The predictive results for control features in the graph indicate that DT produced the maximum accuracy with a value of 79% compared to other classifiers. Therefore, DT for Control features is the best predictive classifier. The different algorithms were used to determine the predictive accuracy for Threat Actor features. DT, RF, KNN and NB Multi produced maximum accuracy, however, DT and NB Multi produced the most accuracy with (61%).Further, and we checked the performance of the different algorithms under the TTP features. DT, RF, KNN and NB Multi produced good accuracy but DT outperformed other algorithms with 62%. NN and NB algorithms did not give us excellent results. Lastly, the Full feature result in the graph below shows all the classifiers produced accuracy of 39% and less. The result shows that Full features did not perform well on all the classifiers. Therefore, we can conclude that the best algorithm that performed well on all the input features except the full feature is DT and NB Multi.

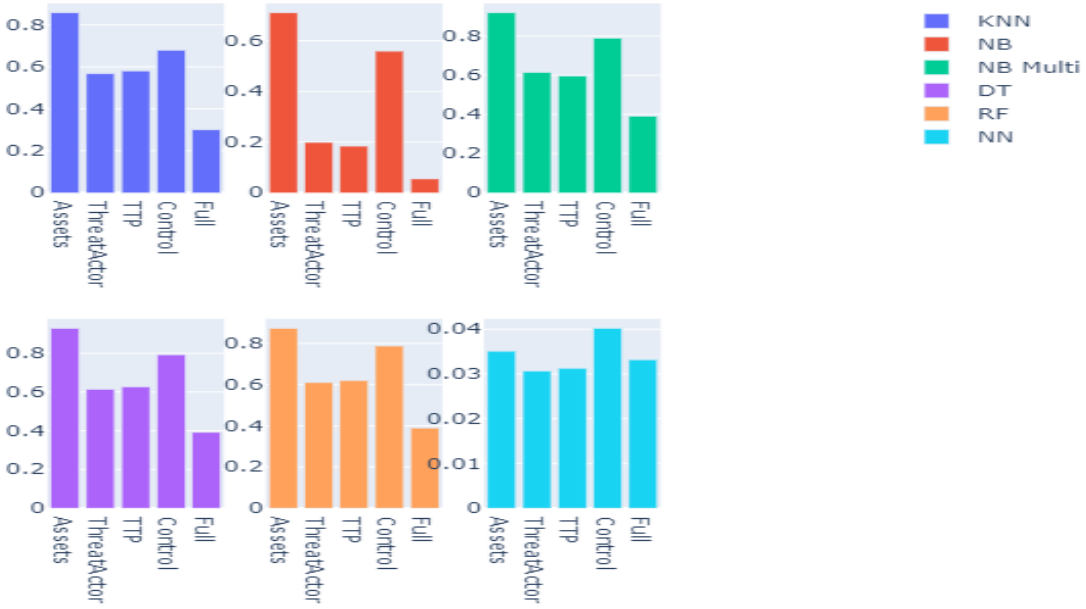

Figure 7: The accuracy of different classifiers for various types of input features.

4.2.2.2. Results of the different classifiers transformed by PCA

Figure 8 shows the results of different classifiers for various kinds of input features that have been transformed by applying PCA. We figure out that, PCA does improve accuracy for TTP and Control features where the accuracy is above 79%.

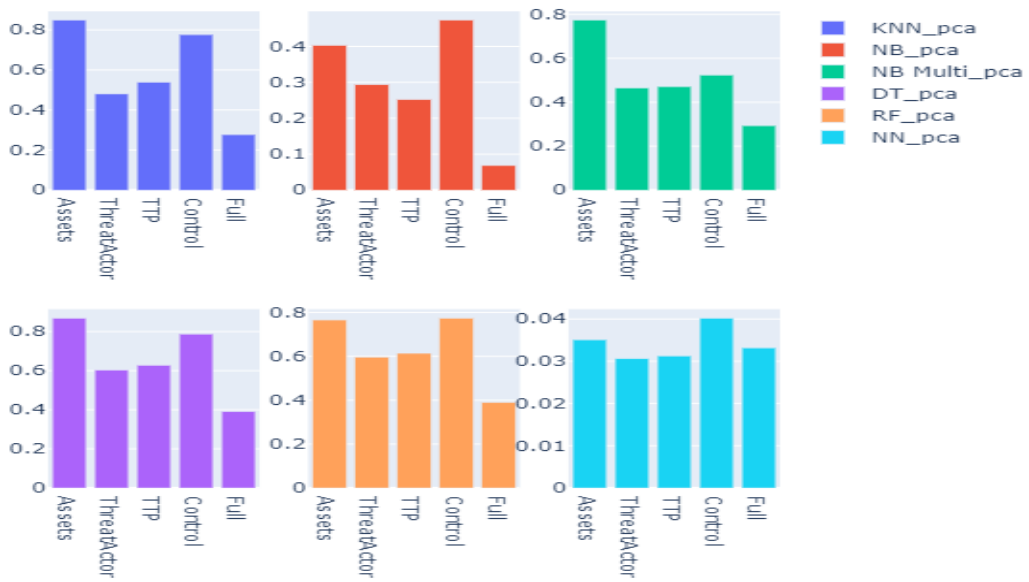

Figure 8: The accuracy of classifiers for various types of features transformed by applying PCA.

4.2.4. Results of Confusion Matrix

While accuracy provides a general indicator of classifier performance, recall, precision, and F measure values give a more complete picture of how the classifier produces errors. Recall measures the true positive rate, precision measures the positive predictive value, and the F measure is the harmonic mean of precision and recall. For these measures, values approaching 1.0 indicate strong classification performance. This section describes the performance of the classifiers on the test data for which the true values are known. This allows for the visualization of the performance of an algorithm. In this case, the best overall predictive accuracy was recorded with KNN which produced better result compared to other classifiers as shown in table 19.

Table 19: performance measure for KNN classifier for the various risk types

Output	Precision	Recall	F1-Score
1	1.000	0.525	0.689
2	0.700	0.687	0.693
3	0.729	0.501	0.694
4	0.766	0.578	0.659
5	0.735	0.561	0.636
6	0.614	0.340	0.438
7	0.820	0.432	0.566
8	0.815	0.373	0.512

9	0.950	0.710	0.813
10	0.264	0.711	0.385
Accuracy	**0.576**	**0.576**	**0.576**

4.2.4.1. Analyzing the results of the KNN algorithm for identifying the different types of risk
KNN provides better results comparing to other algorithms. This section provides precision metrics obtained from the KNN classifier. As presented in Figure 9, Crimeware risk type (R1) shows the highest precision which is almost 100%, cyber espionage (R2) shows 70% , while 73% precision for Denial of Service (R3).

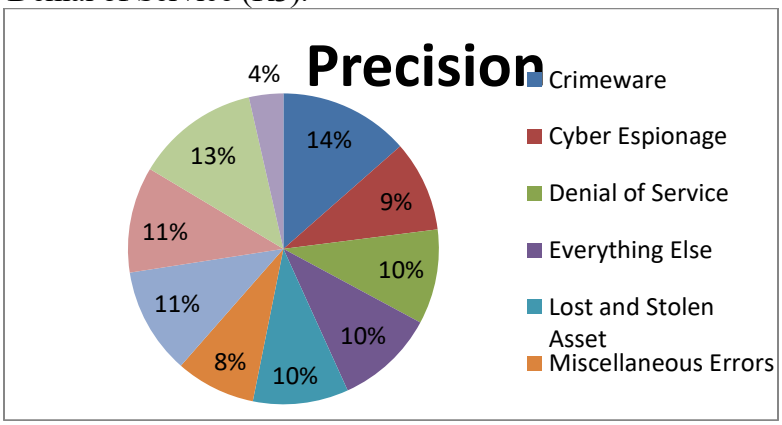

Figure 9: Precision result performance measure for KNN classifier for the various risk types based on the different features

The KNN classifier shows a recall of 53% in identifying Crimeware (R1), 69% recall was obtained for cyber espionage (R2), 50% recall for Denial of Service (R3), 58% recall for everything else (R4) and 56% recall for lost and stolen assets (R5). Recall of 34% for Miscellaneous Error (R6), Recall of 43% for payment card skimmer (R7), recall of 37% for point of sale (R8), recall of 71% for privilege misuse (R9) and a recall of 71% for web application (R10).

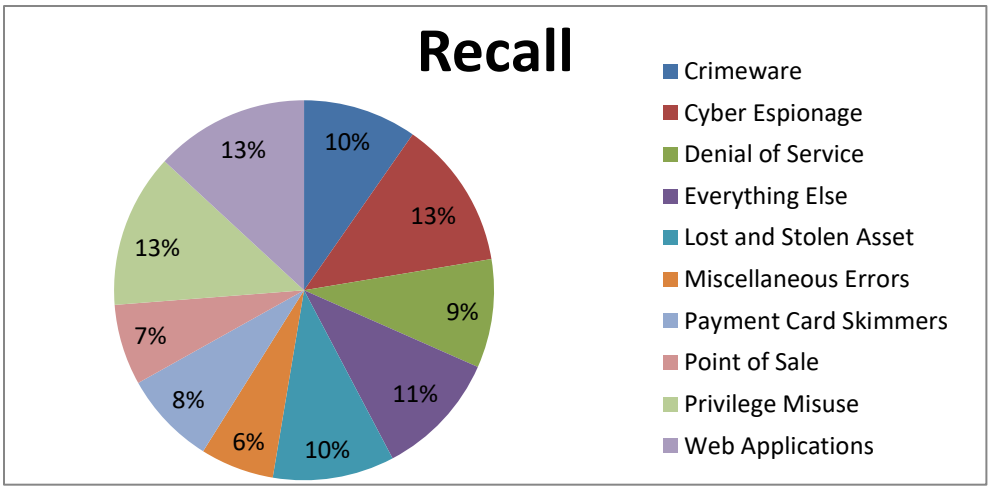

Figure 10: Recall result performance measure for KNN classifier for the various risk types based on the different features

The KNN classifier shows F1-score of 69% in identifying Crimeware (R1), 69% F1-score was obtained for cyber espionage (R2), 69% F1-score for Denial of Service (R3), 66% F1-score for everything else (R4) and 64% F1-score for lost and stolen assets (R5). F1-score of 44% for Miscellaneous Error (R6), F1-score of 57% for payment card skimmer (R7), F1-score of 51% for point of sale (R8), F1-score of 81% for privilege misuse (R9) and a F1-score of 39% for web application (R10).

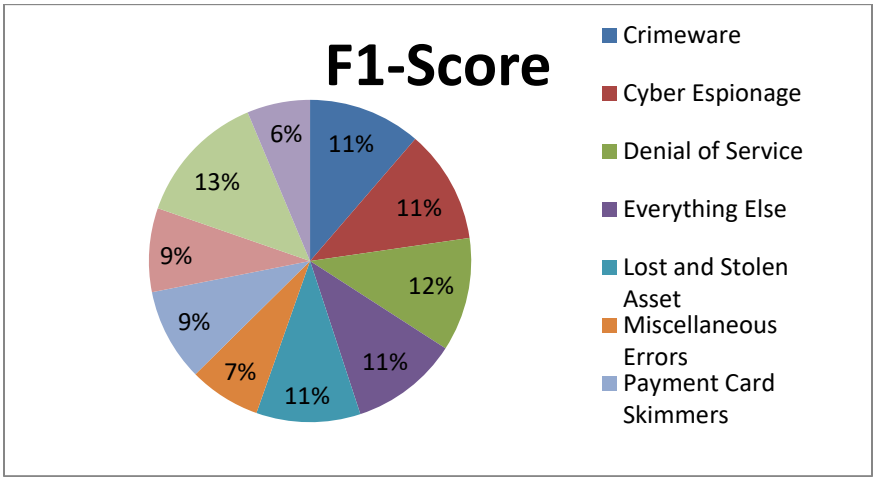

Figure 11: F1 result performance measure for KNN classifier for the various risk types based on the different features

Therefore, KNN achieved very high precision of 100% in identifying Crimeware (R1) and precision of 95% in identifying privilege misuse (R8). KNN achieved high Recall of 71% in both identifying privilege misuse (R9) and web application (R10). Finally, f1-score of 81% for identifying privilege misuse (R9) is achieved.

4.3. Controls

We identified the existing controls and determine the effectiveness of the controls. Table 20 presents the control and its types and overall effectiveness. Two factor authentication provides the highest effectiveness. It also provides a list of recommended controls.

Table 20: Control Effectiveness

Control Type	Control Description	Criteria					Overall Effectiveness	Recommended Controls
		S	R	C	I	T		
Preventive	Account lockout policies after a certain number of a failed login attempt to prevent passwords from being guessed.	4	4	3	4	3	18	Training and awareness for incident handling Relevant identify and access management
	Proper process, registry and file permission should be in place.	4	4	4	3	2	17	

Detective	Identify unnecessary system utilities or potentially malicious software.	3	4	4	3	2	16	Business continuity and incident recovery plan A balanced incident response team Recommend network segmentation
	Network intrusion prevention systems should be put in place.	5	4	3	4	3	19	
Corrective	Limit access to remote services through centrally managed VPNs.	4	4	5	2	1	16	
	Use strong two-factor or multi-factor authentication.	5	5	3	4	3	20	
	Ensure that administrator accounts have complex, unique passwords.	4	3	2	2	2	13	
	Use of two-factor authentication for public-facing webmail servers is recommended.	5	5	3	2	3	18	
	Training required for the DisCos employees to raise awareness.	3	4	5	3	4	19	
	Anti-virus to automatically isolate suspicious files	2	3	4	2	4	15	

5. DISCUSSION

Proactive management of cybersecurity risk is essential for the CPS. However, due to the constant changing of the threat landscape and sophisticated technology used to exploit the attack, this task becomes more challenging. The proposed unified approach aims to contribute for an effective cyber security risk management practice based on assets criticality, risk predication and effectiveness of existing controls. One of the most important aspects of CSRM is to determine the critical assets within a CPS that can be affected by potential risks. The use of fuzzy set theory allows us to determine the asset criticality based on the relative importance of security goals. The proposed approach considers various ML techniques and extracts CSRM features which are relevant for the risk prediction. The risk predication allows organizations to give an early warning of the security issues that needs adequate attention. Organisation of any size must understand the existing critical assets, cybersecurity risks, and effectiveness of existing controls. It helps to understand the current status of security control and undertakes strategic decision for the improvement of overall cybersecurity . The overall predication results of the different risk types based on the given CSRM features indicated that NB-Multi and DT are the best ML classifiers because they performed better by predicting seven different risk types such as “Crimeware”, “Cyber Espionage”, Denial of Service”, Everything Else”, “Lost and Stolen Assets”, “privilege Misuse” and “Point of Sale” while others predicted six or less. Following the above discussion, we observe that CSRM features (TTP, Assets, Controls and Threat Actor) types could actually be used to predict risk type. Therefore, as security threats grow, organisations need to identify cybersecurity threats and its trend and also be able to detect and respond to both known and unknown risks. This supports organisations to determine the right risk type and implement appropriate controls.

5.1. Comparison with the other study results

In this section, we compare the results of our approach with other study results from the literature to generalize our findings.In [43] Fuzzy Risk Analysis and Management for Critical Asset Protection (RAMCAP) is introduced in order to risk analysis and management for pipeline systems. However, the Fuzzy RAMCAP considers the relative importance among vulnerability, threat, and consequence but not the relevant goals for the assets. The authors in [44] focus on the manual and automatic asset identification, annotation and tracking as well as on the assignment of graded application security controls (ASCs) that can benefit from a comprehensive and formalized asset management. However, this process is not feasible in a heavily regulated business domain and can easily be a target to threat actors leading to a cyber-attack. Also, the work didn't consider the use of key primary indicators (KPI) to determine the criticality. A framework that models the connections of assets and identify critical components is presented [28]. However, the framework works best for network assets grouping only and doesn't consider other assets within the organisation. Also a robust calculation is required to determine critical assets. Our work presents FACS to analyze critical assets in CI. The main purpose was to investigate the major security risk associated to critical assets and effectively reduce and organisations overall risk against cyber security threats. The FACS considers the relative importance among the goals of the assets (C, I, A, CON and ACC). We test the implementation of FACS on a real-life power grid system and demonstrate its effectiveness. The result shows that FACS provides an accurate asset criticality ranking for risk analysis in CI.

In [12], an adaptive intrusion detection system is proposed that detects different types of attacks in adversarial network environments. However, the proposed framework needs to be applied to other information security problems. In [45]an investigation on detecting and categorizing anomalies is carried out using LR and RF ML techniques. The result demonstrates that RF technique with feature selection scheme can achieve 99% accuracy with anomaly detection. Much research has been carried out in this domain without paying attention to identifying risk and imposing appropriate countermeasures against different types of attacks. In [13], the authors reviewed the most commonly used machine learning algorithms, which are primary tools for analyzing network traffic, intrusion detection, DDoS attack detection, web applications, and detecting anomalies. However, detecting risk type is still an ongoing plan. In [40]the result demonstrates how and to what extent business details about an organisation can help forecast its relative risk of experiencing different types of data incidents using incident report collected in the VCDB to achieve some level of protection. In [9] RF classifier is used to train more than 1,000 incidents taken from the VCDB to predict an organisations network breaches. Our work also used ML techniques in the cybersecurity domain but differentiated from other existing works with specific focus on cybersecurity risk prediction. Also, [34] proposed a comprehensive assessment model (CAM) that provides ways to measure internal controls. However, the criteria for assessment are complex and include some predefined rules that may be hard to follow. Our observation is that it would be useful in determining critical assets, predicting risk types and evaluating the effective of existing controls for an effective CSRM.

6. CONCLUSION

The cyber threat landscape is evolving rapidly with new techniques and more sophisticated attacks and risk management certainly plays an important role to understand the threats and associated risks to choose the suitable controls. This paper proposes a novel unified CSRM approach that

systematically determines critical asset, predicts the risk types for an effective risk management practice and evaluate the effectiveness of the existing controls. Our experimental results identify five critical assets by following the fuzzy set theory. Our observation is that the input and output information in the fuzzy logic is described as linguistic terms, which are more realistic and flexible in reflecting real situations. Furthermore, risk types are analysed using a predictive model influenced by the vulnerabilities and relevant threat, and TTP of the organisation to provide accurate risk level. The results also revealed that decision tree-based algorithms (DT and RF) are well suited for the risk prediction problem with 93% accuracy and further to 96% using PCA. The reason being that the decision tree can easily identify the most prominent feature to construct tree and can stop induction of the model before over fitting happens, which give the better generalization error for the test set. We notice that the accuracy is comparatively lower in our case for the classifiers. One possible reason can be the nature of the data, which is highly imbalanced and sparse. As a future endeavor, we intend to use oversampling and sparsity reduction techniques before applying classification algorithms, which might increase the performance of various models. Also, the risk predication part only considers supervised learning method, which requires effort for the dataset labeling, so we plan to use our approach on unsupervised data. It is necessary to create a process for integrating machine learning for an effective cybersecurity risk management practice. We are also planning to handle the zero-day attacks by using our approach.

Acknowledgements
This work has received funding from the Nigerian Petroleum Development Trust Fund (PTDF).

Reference

[1] W. Wu, R. Kang, and Z. Li, "Risk assessment method for cyber security of cyber physical systems," in *Reliability Systems Engineering (ICRSE), 2015 First International Conference on*, 2015, pp. 1–5.
[2] K.-D. Kim and P. R. Kumar, "An overview and some challenges in cyber-physical systems," *J. Indian Inst. Sci.*, vol. 93, no. 3, pp. 341–352, 2013.
[3] M. Fossi *et al.*, "Symantec internet security threat report trends for 2010," *Vol. XVI*, 2011.
[4] Experian, "2015 second annual data breach industry forecast," 2015.
[5] S. Boyson, "Cyber supply chain risk management: Revolutionizing the strategic control of critical IT systems," *Technovation*, vol. 34, no. 7, pp. 342–353, 2014.
[6] T.-F. Yen, V. Heorhiadi, A. Oprea, M. K. Reiter, and A. Juels, "An epidemiological study of malware encounters in a large enterprise," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 1117–1130.
[7] D. Canali, L. Bilge, and D. Balzarotti, "On the effectiveness of risk prediction based on users browsing behavior," in *Proceedings of the 9th ACM symposium on Information, computer and communications security*, 2014, pp. 171–182.
[8] K. Soska and N. Christin, "Automatically detecting vulnerable websites before they turn malicious," in *23rd {USENIX} Security Symposium ({USENIX} Security 14)*, 2014, pp. 625–640.
[9] Y. Liu *et al.*, "Cloudy with a chance of breach: Forecasting cyber security incidents," in *24th {USENIX} Security Symposium ({USENIX} Security 15)*, 2015, pp. 1009–1024.
[10] K. Veeramachaneni, I. Arnaldo, V. Korrapati, C. Bassias, and K. Li, "AI^ 2: training a big

data machine to defend," in *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)*, 2016, pp. 49–54.

[11] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surv.*, vol. 34, no. 1, pp. 1–47, 2002.

[12] H. T. Nguyen and K. Franke, "Adaptive Intrusion Detection System via online machine learning," in *2012 12th International Conference on Hybrid Intelligent Systems (HIS)*, 2012, pp. 271–277.

[13] O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," in *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 2186–2193.

[14] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL detection using machine learning: A survey," *arXiv Prepr. arXiv1701.07179*, 2017.

[15] Y. Yang *et al.*, "Intrusion detection system for network security in synchrophasor systems," 2013.

[16] J. M. Beaver, R. C. Borges-Hink, and M. A. Buckner, "An evaluation of machine learning methods to detect malicious SCADA communications," in *2013 12th International Conference on Machine Learning and Applications*, 2013, vol. 2, pp. 54–59.

[17] D. Sun, Z. Wu, Y. Wang, Q. Lv, and B. Hu, "Risk Prediction for Imbalanced Data in Cyber Security: A Siamese Network-based Deep Learning Classification Framework," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–8.

[18] L. Bilge, Y. Han, and M. Dell'Amico, "Riskteller: Predicting the risk of cyber incidents," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 1299–1311.

[19] A. Subroto and A. Apriyana, "Cyber risk prediction through social media big data analytics and statistical machine learning," *J. Big Data*, vol. 6, no. 1, p. 50, 2019.

[20] A. P. H. de Gusmão, M. M. Silva, T. Poleto, L. C. e Silva, and A. P. C. S. Costa, "Cybersecurity risk analysis model using fault tree analysis and fuzzy decision theory," *Int. J. Inf. Manage.*, vol. 43, pp. 248–260, 2018.

[21] Q. Zhang, C. Zhou, N. Xiong, Y. Qin, X. Li, and S. Huang, "Multimodel-based incident prediction and risk assessment in dynamic cybersecurity protection for industrial control systems," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 46, no. 10, pp. 1429–1444, 2015.

[22] J. Chen, K. Li, H. Rong, K. Bilal, K. Li, and S. Y. Philip, "A periodicity-based parallel time series prediction algorithm in cloud computing environments," *Inf. Sci. (Ny).*, vol. 496, pp. 506–537, 2019.

[23] J. Chen *et al.*, "A parallel random forest algorithm for big data in a spark cloud computing environment," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 4, pp. 919–933, 2016.

[24] J. Chen, K. Li, Z. Tang, K. Bilal, and K. Li, "A parallel patient treatment time prediction algorithm and its applications in hospital queuing-recommendation in a big data environment," *IEEE Access*, vol. 4, pp. 1767–1783, 2016.

[25] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, 2009, vol. 5.

[26] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proc. IEEE*, vol. 100, no. 1, pp. 210–224, 2012.

[27] C. Livadas, R. Walsh, D. Lapsley, and W. T. Strayer, “Usilng machine learning technliques to identify botnet traffic,” in *Proceedings. 2006 31st IEEE Conference on Local Computer Networks*, 2006, pp. 967–974.
[28] G. Stergiopoulos, P. Dedousis, and D. Gritzalis, “Automatic network restructuring and risk mitigation through business process asset dependency analysis,” *Comput. Secur.*, p. 101869, 2020.
[29] C. I. Cybersecurity, “Framework for Improving Critical Infrastructure Cybersecurity,” *Framework*, vol. 1, p. 11, 2014.
[30] R. A. Martin, “Common weakness enumeration,” *Mitre Corp.*, 2007.
[31] H. I. Kure, S. Islam, and M. A. Razzaque, “An integrated cyber security risk management approach for a cyber-physical system,” *Appl. Sci.*, vol. 8, no. 6, p. 898, 2018.
[32] H.-J. Zimmermann, *Fuzzy set theory—and its applications*. Springer Science & Business Media, 2011.
[33] CIS_CSC, “The Critical Security Controls For Effective Cyber Defense,” *Cent. Internet Secur.*, 2018.
[34] C. Dittmeier and P. Casati, “Evaluating Internal Control Systems: A Comprehensive Assessment Model (CAM) for Enterprise Risk Management,” *Altamonte Springs, Florida Inst. Intern. Audit. Res. Found.*, 2014.
[35] C. C. ENISA, “Benefits, risks and recommendations for information security,” *Eur. Netw. Inf. Secur.*, 2009.
[36] L. A. Zadeh, “Fuzzy logic,” *Computer (Long. Beach. Calif).*, vol. 21, no. 4, pp. 83–93, 1988.
[37] A. S. Markowski and M. S. Mannan, “Fuzzy logic for piping risk assessment (pfLOPA),” *J. loss Prev. Process Ind.*, vol. 22, no. 6, pp. 921–927, 2009.
[38] O. Cordón, “A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems,” *Int. J. Approx. Reason.*, vol. 52, no. 6, pp. 894–913, 2011.
[39] S. Widup, “The veris community database.” 2013.
[40] A. Sarabi, P. Naghizadeh, Y. Liu, and M. Liu, “Prioritizing Security Spending: A Quantitative Analysis of Risk Distributions for Different Business Profiles.,” 2015.
[41] M. Firoiu, “General considerations on risk management and information system security assessment according to ISO/IEC 27005: 2011 and ISO 31000: 2009 standards,” *Calitatea*, vol. 16, no. 149, p. 93, 2015.
[42] ISO 27005:2011 Information Techniques- Information Security Risk Management, International Organization for Standardization (ISO) 2009.
[43] A. Alidoosti, A. Jamshidi, S. Yakhchali, M. Basiri, R. Azizi, and A. Yazdani-Chamzini, “Fuzzy logic for pipelines risk assessment,” *Manag. Sci. Lett.*, vol. 2, no. 5, pp. 1707–1716, 2012.
[44] K. Waedt, A. Ciriello, M. Parekh, and E. Bajramovic, “Automatic assets identification for Smart Cities: Prerequisites for cybersecurity risk assessments,” in *2016 IEEE International Smart Cities Conference (ISC2)*, 2016, pp. 1–6.
[45] T. Salman, D. Bhamare, A. Erbad, R. Jain, and M. Samaka, “Machine learning for anomaly detection and categorization in multi-cloud environments,” in *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*, 2017, pp. 97–103.
[46] ISO 27001:2017: Information Technology -Security Techniques-Information Security

Management System Requirements, International Organization for Standardization (ISO) 2011