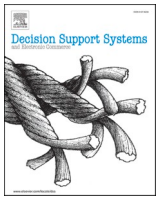
Decision Support Systems 177 (2024) 114102

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss



A hybrid framework using explainable AI (XAI) in cyber-risk management for defence and recovery against phishing attacks

Baidyanath Biswas [a], Arunabha Mukhopadhyay [b], Ajay Kumar [c,*], Dursun Delen [d,e]

[a] Trinity Business School, Trinity College Dublin, Ireland
[b] IT and Systems Area, Indian Institute of Management, Lucknow, India
[c] EMLYON Business School, Ecully, France
[d] Center for Health Systems Innovation, Department of Management Science and Information Systems, Spears School of Business, Oklahoma State University, Tulsa, Oklahoma, USA
[e] Department of Industrial Engineering, Faculty of Engineering and Natural Sciences, Istinye University, Istanbul, Turkey

ARTICLE INFO

Keywords:
Information security
Explainable AI
Cyber insurance
Bivariate distributions
Copula

ABSTRACT

Phishing and social engineering contribute to various cyber incidents such as data breaches and ransomware attacks, financial frauds, and denial of service attacks. Often, phishers discuss these attack vectors in dark forums. Further, the probability of phishing attacks and the subsequent loss suffered by the firm are highly correlated. In this context, we propose a hybrid framework using explainable AI techniques to assess cyber-risks generated from correlated phishing attacks. The first phase computes the probability of expert phishers within a community of similar attackers with varying expertise. The second phase calculates the probability of phishing attacks upon a firm even after it has invested in IT security and adopted regulatory steps. The third phase categorises phishing and genuine URLs using various machine-learning-based classifiers. Next, it estimates the joint distribution of phishing attacks using an exponential-beta distribution and quantifies the expected loss using Archimedean Copula. Finally, we offer recommendations for firms through the computation of optimal investments in cyber-insurance versus IT security. First, based on the risk attitude of a firm, it can use this explainable-AI (XAI) framework to optimally invest in building security into its enterprise architecture and plan for cyber-risk mitigation strategies. Second, we identify a long-tail phenomenon demonstrated by the losses suffered during most cyber-attacks, which are not one-off incidents and are correlated. Third, contrary to the belief that cyber-insurance markets are ineffective, it can guide financial firms to design realistic cyber-insurance products.

1. Introduction

Phishing involves the social engineering of data over the Internet to acquire personal and business information from users. Attackers exploit a user's susceptibility and deceive them into divulging critical information [94], such as login credentials for social networks [91], banking applications [12], credit cards [55] and healthcare [49] through emails, corrupt URLs or text messages. During the COVID-19 pandemic, phishing attacks contributed to other secondary cyber disasters, such as data breaches, ransomware attacks, business email compromises, tech-support frauds, and tax refund scams, with over USD 2.07 billion losses. Attackers crafted special phishing emails[1] to spoof the COVID-19 loan relief funds in the USA[2] and financial support packages in the UK. Although malicious agents execute these attacks through technological steps, information systems (IS) scholars perceive *phishing* as more of an economic problem with related financial risks [1,42]. For instance, Ju et al. [45] reported an annual productivity improvement of 7.6 million USD in South Korean firms by preventing phishing emails. The European Union Report on cyber-insurance also highlights that phishing emails pose the most frequent cyber-risks to businesses.[3] Thus, phishing

* Corresponding author.
E-mail addresses: baidyanath@gmail.com (B. Biswas), arunabha@iiml.ac.in (A. Mukhopadhyay), akumar@em-lyon.com (A. Kumar), dursun.delen@okstate.edu (D. Delen).

[1] COVID-19 Exploited by Malicious Cyber Actors
[2] Malicious Cyber Actor Spoofing COVID-19 Loan Relief
[3] Cyber Risk for Insurers – Challenges and Opportunities

https://doi.org/10.1016/j.dss.2023.114102
Received 22 March 2023; Received in revised form 7 September 2023; Accepted 2 October 2023
Available online 4 October 2023

attacks pose high-level cyber-risks and necessitate proactive intervention from firms[4,5] using ML and AI, particularly explainable AI (XAI) techniques.

While cyber-attacks remain among the most severe threats businesses face worldwide, they must mitigate the losses, hence the necessity for cyber-insurance.[6] For example, AIG Cyber Coverage Policy offers first-party and liability coverage for a failure of a company's network security or a failure to protect confidential information. In addition, cyber-insurance schemes may include costs to notify clients who suffered data breaches or cyber-attacks, restore their files electronically, and handle claims from stakeholders, regulators, and clients. These agents also conduct regular cyber-risk assessments that include measuring threat likelihood from external sources, using the data to model residual risks, calculating the probability of impact and loss ranges for each attack type, and finally, implementing controls.[7]

1.1. How can explainable AI (XAI) play a significant role in cyber-risk management?

Explainable AI (XAI) defines the process that permits users to comprehend how an AI system decides, predicts, and performs its operations. Therefore, XAI helps identify the strengths and weaknesses of the rules associated with decision-making using AI/ML [33,70,71,81]. While scholars argue that sometimes the need for explanation may be too difficult to achieve, nevertheless, building AI-based systems for critical applications [34, 47. 81, 95, 105] and lucid explanations of these AI-based rules are essential for the system users to understand, trust, and effectively manage them [33]. However, a quick scan of the extant IS literature shows a gap with scant articles published on *explainable*, *interpretable*, and *transparent* AI/ML. Fig. 1 presents the yearly count of publications in the AIS Senior Scholars' List of Premier Journals[8] with *explainable*, *interpretable*, and *transparent* AI[9] (derived from SCOPUS). The plot shows that *Decision Support Systems* leads the list with twenty-one articles published across 2000–2023.

Next, we briefly discuss the scope of XAI-based research in cyber-risk management, which has yet to receive much attention. First, to tackle phishing problems, firms aim to reduce the probability of phishing attacks, where they can implement XAI-based techniques to improve upon traditional cybersecurity measures. For instance, the recent literature proposes deep-learning methods to apply character-level information for generic text classification and subsequent categorisation of URLs into *phish* and *legitimate* [13,88,105]. While traditionally ML-based techniques were easily interpretable [61], they fell short of scalability and adaptability while adding new URLs (because phishing URLs are short-lived, and many had a life span of less than a day[10]). Recently, the NLP-based hybrid techniques (i) have often involved a huge number of variables in model-building [39,61,73], and (ii) do not apply feature-engineering and variable importance schemes. Further, many of these advanced detection frameworks rely on complex deep-learning techniques and must be more interpretable because they involve many features to build the ML model. Therefore, they need to be more tractable and explainable.

Second, we look at the literature on hacker forums and darknet analysis [8,6,22,74,76]. Online forums share phishing tools and techniques such as online deception [91], ransomware, and denial-of-service (DDoS) payloads deliverable through emails [76,77]. While recent studies [22,72] also encourage the application of advanced deep-learning models, they often still need to be discovered, and their findings are difficult to interpret by business users. Therefore, explainable AI can play a role in both cases by making the associated rules and features more easily understandable to users, if not by making a minor compromise in the accuracy of the models [105].

1.2. Role of cyber-insurance in cyber-risk management

Cyber-insurance is a unique and viable tool to mitigate the residual losses since it helps indemnify the losses in a phishing attack [8,64,108]. In many cases, cyber-insurers indemnify financial losses arising from phishing attacks (i) if the related loss is significant [29,54], (ii) by investing in IT security tools [64], and (iii) the firm does not suffer from moral hazard problems [67,68]. Additionally, a cyber-insurer's security scans during the underwriting and post-issuance phases led to a 65% reduction in ransomware claims faced during the cyber-attacks.[11] The cyber-insurance market in the USA is estimated at US$ 2.5 billion per annum and is expected to grow up to $7.5 billion over the next decade.[12] For instance, Nationwide Mutual offers cyber-insurance products for retail users to ensure identity theft protection and small businesses to cover financial damages arising from phishing, cyber-stalking, and malware attacks[13,.14] These criteria make a strong case for a firm to insure its phishing-related risks [82].

1.3. Objectives of this study

This study proposes a three-stage framework to assess cyber-risks originating from phishing attacks and subsequent mitigation by balancing investments in technology and cyber-insurance. We follow the lifecycle of a phishing attack to identify the following: (i) probability of expert attackers, (ii) probability of phishing attacks, and (iii) probability of failed detection of phishing attacks. Next, our framework applies the utility theory and Copula to determine (i) optimal premium for cyber-insurance, and (ii) complementary IT security under different risk attitudes. Essentially, we seek an answer to the following research questions:

RQ1. What is the joint probability of a phishing attack?

RQ2. What is the expected loss of a firm where the probability of misdetection and loss are correlated?

This study contributes to the literature by first identifying expert and novice phishers using a relatively simpler yet interpretable XAI-based technique. Next, it identifies the absence of regulatory and legal factors as unique predictors of phishing attacks. Then, it identifies the NLP-based features of phishing URLs and selects the top ten among them using variable importance. Then, it applies a Beta distribution to fit the misclassification rate during phishing detection, followed by the application of Archimedean Copula to examine the correlated loss. Finally, the study offers mitigation strategies for firms (e.g., investment portfolios with IT security tools and complementary cyber insurance) to mitigate cyber-risks.

The remainder of this paper is organised as follows. Section 2 presents an overview of existing studies on phishing and cyber-insurance and identifies relevant research gaps. In Section 3, we build the proposed framework and methodology adopted. In Section 4, we describe the data. In Section 5, we execute the model and present the results. Then, in Section 6, we discuss the managerial and academic implications of the results. Finally, in Section 7, we conclude this study and highlight the future scope for extension.

[4] APWG Phishing Attack Trends Reports
[5] MillerSmiles Phishing Scam Archives
[6] Cyber Risk for Insurers – Challenges and Opportunities
[7] AIG Cyber RiskAssessment
[8] https://aisnet.org/page/SeniorScholarListofPremierJournals
[9] Search Key: (TITLE-ABS-KEY ((“explain” OR “interpret” OR “transparent”) AND (“artificial intelligence” OR “machine learning”))
[10] The life cycle of phishing pages
[11] Cyber insurer's security scans reduced ransomware
[12] Insurance 2020 and beyond
[13] ID theft protection
[14] What is cyber insurance?

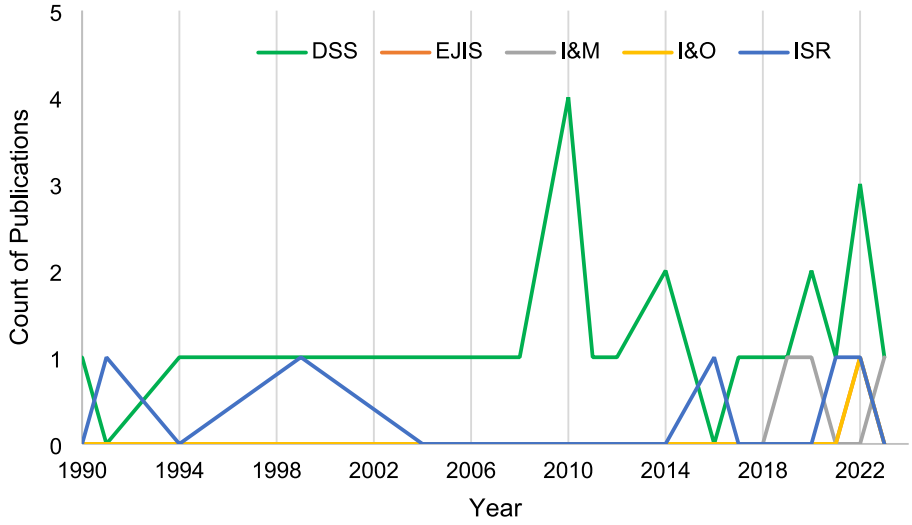


Fig. 1. Yearly publications based on Scopus Data.
DSS=Decision Support Systems; EJIS = European Journal of Information Systems; I&M = Information and Management; I&O=Information and Organization; ISR = Information Systems Research.

2. Literature review

In this Section, we discuss contemporary studies on (i) *darknets and hacker forums*, (ii) *phishing attacks*, and (iii) *cyber-insurance* in the context of our study.

2.1. Darknets and hacker forums – "dark side of IT"

Hacker forums and darknet analysis constitute a significant part of IS/IT literature [6,8,74,76], often termed the "dark side of IT". Using these forums as platforms, attackers often gain access to technical artefacts such as malicious codes and malware files to execute cyber-attacks, leading to financial losses [8,63,64]. We briefly review the relevant IS/IT literature exploring hacker forums. To identify key actors, Benjamin and Chen [5] applied an array of forum features such as *length of posts*, *number of replies* and age *of membership* to determine the reputation of member hackers. Next, Benjamin et al. [6] built DICE-E to describe an ethical way of collecting data from darknet forums. In their analysis, Benjamin et al. [6] delved into various darknet forums to acquire forum features to calculate the status of different members.

Samtani and associates have conducted a series of studies to build empirical models and predict the status (or reputation) of member hackers in dark forums. Samtani and Chen [75] applied social network analysis to identify top actors in darknet forums using social network-based metrics. Then, Samtani et al. [76] tracked artefacts within one English and six Russian hacker forums using topic modelling. Ebrahimi et al. [22] built a deep-learning framework to learn from English darknets, apply the knowledge, and detect artefacts in non-English forums. Then, Grisham et al. [32] applied deep learning and text classification techniques to find significant actors in darknet forums and extended past studies [5,75]. Next, based on the mode of knowledge transaction (*acquisition* or *provision*), Zhang et al. [55] examined darknet posts and identified the presence of various ranks of hackers based on their reputation.

2.2. URL-based detection of Phishing Attacks

While the recent literature on phishing can be categorised across three broad dimensions - *behavioural*, *financial impact*, and *technological mitigation*, we restrict ourselves to the third strand of literature. Anti-phishing literature may involve a range of features to differentiate phishing content from genuine URLs, such as URL structure, web pages, emails and text messages. Next, we briefly discuss the extant literature on this topic.

Technology-based solutions to detect phishing websites have drawn considerable attention among practitioners and the research community [42,100]. Zhang et al. [104] proposed a content-based algorithm, CANTINA, to detect phishing websites based on TF-IDF. Joshi et al. [44] proposed a novel algorithm using a browser plugin *PhishGuard* to identify a forged website and tested it in Mozilla Firefox 3.0. Xiang et al. [100] proposed the CANTINA+ algorithm by employing machine-learning techniques to detect phishing with HTML pages, search engines, and third-party services. Mohammad et al. [61] identified thirty predictors of phishing URLs that could accurately classify any URL into genuine, suspicious, or phishing links. Tan et al. [89] proposed *PhishWHO*, a phishing web page-detection technique using classification algorithms by comparing a web page's targeted and actual identities. Babagoli et al. [2] used support vector machines to detect a phishing website. Jain and Gupta [39] used a random-forest-based classifier to detect phishing websites by extracting client-side features. Smadi et al. [85] built an online system to detect phishing emails using dynamically evolving neural networks based on reinforcement learning. Bozkir et al. [13] built a phishing URL detection system with a unique dataset through N-gram embeddings using neural networks.

2.3. Cyber insurance as an IS risk mitigation tool

Firms suffering from cyber incidents and financial losses resort to cyber insurance, which helps them cover IS security risks [108,109]. Siegel et al. [82] suggested that an insurance framework must support technology and process controls to manage IT security successfully. Gordon et al. [30] proposed a cyber-insurance framework to mitigate IS risk by balancing investments in security spending and insurance premiums. Böhme [9] built an indemnity model for claims against correlated losses and showed that software monoculture and correlated risks hindered a broad cyber-insurance market. Öğüt et al. [67] analysed the interdependency of IS risks across firms and proposed information-sharing strategies for social optimality. Majuca et al. [57] examined the evolution of the cyber insurance market the effects of moral hazards and identified adverse selection as a reason for the slow adoption of cyber insurance. Böhme and Kataria [10] analysed the impact of correlated cyber-risk on cyber-insurance markets. Shetty et al. [79]

integrated the impact of breaches across all types of losses and extended Bandyopadhyay et al. [3] to explain the challenges that plagued the extensive use of cyber insurance. Herath and Herath [36] proposed a cyber-insurance pricing model using Gumbel and Clayton Copula techniques based on the time of the breach and the insurance premium paid.

Next, Mukhopadhyay et al. [63] presented a utility-based preferential-pricing model for cyber-insurers with the help of a Gaussian copula-based Bayesian Belief Network with vulnerability assessment and subsequent cyber-risk computation. Further, they computed the insurance premium under different risk profiles (averse, neutral, constant) to help the firm decide whether to insure IT. Laszka and Grossklags [52] proposed a market structure where insurance providers are encouraged to invest in security research through vulnerability reward programs. Srinidhi et al. [86] built a decision model to guide managers in the allocation of resources to productive as well as security processes. Mukhopadhyay et al. [64] estimated the probability of a cyber-attack, then the expected loss, and subsequently proposed cyber insurance or self-protection as a risk-mitigation strategy for firms. Bandyopadhyay and Mookerjee [4] demonstrated the challenges of using cyber insurance by studying the optimal purchase decisions after primary and secondary breaches faced by a firm.

Sharma and Mukhopadhyay [78] examined the losses multiplayer online gaming firms suffered from DDoS attacks. They proposed an ensemble of cyber-risk mitigation strategies such as technology investment, cyber-insurance, or both. Next, we examined a few studies that applied the Gordon-Loeb model (GL) to propose firm-level decisions to choose between cybersecurity investments or cyber-insurance or both: a one-period utility-based scenario using various utility functions (Skeoch [83]); a quantitative cyber-risk assessment framework in critical infrastructures (Young et al. [103]); extends Young et al. [103] combining insurance and security investments, where the latter contribute to reducing the insurance premium (Mazzoccoli & Naldi [58]); a single profit-maximising insurer with voluntarily participating interdependent insured firms (Khalili et al. [46]).

Table 1 presents the summary of the extant literature on cyber insurance and highlights the lack of a single study that combines *cyber-risk quantification* through (i) estimation of the probability of expert phishers by mining dark-net messages, (ii) estimation of the probability of cyber-attacks, (iii) estimation of the probability of detection of phishing-attacks, and (iv) *cyber-risk mitigation* through cyber-insurance. So, we position this study at the intersection of the above gaps in the literature.

3. Proposed framework and research methodology

Fig. 2 shows our proposed framework, which consists of an *assessment of phishing risks* and *mitigation through cyber insurance*. The assessment stage is a joint exercise that can be *internal* to the firm, such as using anti-phishing filters, or *external*, such as the expertise of the phisher (attacker), or a mix of both *internal* and *external* factors, such as the chances of a firm facing phishing attacks even though it has invested in security technologies and practised secure IT policies and adhered to regulations. PRA consists of three modules: PEA, PPA and DPA, which lead to the joint probability of phishing attacks. Finally, MCI offers a complementary solution to choose between cyber insurance and IT security investments. Table 2 presents the list of notations and abbreviations used in this study.

3.1. Phishing-Risk Assessment (PRA)

The PRA module calculates the probability of a given URL, aiming to classify it into - "genuine" or "malicious". Therefore, the effective probability of attack (p_a) is the joint probability of the three events – the *expertise of the phisher*, the *probability of phishing*, and the *misdetection of phishing URLs*. Mathematically, it is a combination of these three events as follows -

$$\text{p (Risk of Attack)} = \text{p(Expert Phisher)*p (Phishing)*p(Detection Failure)}$$

$$p_a = (p_{ex})^*(p_{ph})^*(1 - p_d) \quad (1)$$

Next, we will compute the probabilities of each of these events. Subsequently, it will lead us to answer our first research question - *RQ1. What is the joint probability of a phishing attack?*

3.1.1. PPA module: estimation of the probability of expert attackers (p_{ex})

Dark forums and phisher communities provide an interesting testbed where members share malicious source code files and discuss technical knowledge [6,8]. Often, expert phishers exchange messages containing "phishing-related" keywords to (i) reinforce their position and reputation as experts and (ii) disseminate phishing-related knowledge among peers. Our first module within risk assessment, PEA, classifies its members into four levels of phishers – *newbie*, *beginner*, *intermediate*, and *advanced* drawing on past literature [6,8]. In particular, we extend the past literature [6,8,76] by selecting those forum messages that could

Table 1
Extant literature on cyber insurance – methodology, theory, inputs, outputs.

Paper	Method		Theory			Input						Output	
	Q1	Q2	C	P	U	O	T	CR	R	M	A	IP	SI
Siegel et al. [82]	Y	–	–	Y	–	Y	–	–	Y	Y	Y	–	–
Gordon et al. [30]	Y	–	–	Y	–	Y	Y	–	–	–	–	–	Y
Böhme [9]	–	Y	–	–	Y	–	Y	Y	–	Y	–	Y	–
Öğüt et al. [67]	Y	–	Y	–	–	–	Y	Y	Y	Y	–	–	–
Majuca et al. [57]	Y	–	Y	–	–	Y	Y	–	Y	Y	–	–	Y
Böhme and Kataria [10]	–	Y	–	–	Y	–	Y	Y	–	Y	–	Y	–
Shetty et al. [79]	–	Y	–	–	Y	–	Y	Y	–	Y	–	Y	–
Herath and Herath [36]	–	Y	Y	–	–	–	Y	Y	–	–	–	Y	–
Mukhopadhyay et al. [63]	Y	Y	Y	Y	–	Y	Y	–	–	Y	Y	Y	Y
Laszka et al. [53]	–	Y	–	–	Y	Y	Y	–	–	Y	Y	–	Y
Srinidhi et al. [86]	–	Y	–	Y	–	–	–	–	–	Y	–	Y	Y
Bandyopadhyay and Mookerjee [4]	–	Y	–	–	Y	–	Y	–	–	Y	Y	Y	–
Mukhopadhyay et al. [64]	–	Y	–	Y	–	Y	Y	–	–	Y	Y	–	Y
Sharma and Mukhopadhyay [78]	–	Y	–	Y	–	Y	Y	–	–	Y	Y	–	Y
Skeoch [83]	–	Y	–	–	Y	Y	Y	–	–	–	Y	Y	Y
Mazzoccoli & Naldi [58]	–	Y	–	–	Y	Y	Y	–	–	–	Y	Y	Y
Young et al. [103]	–	Y	–	–	Y	Y	Y	Y	–	Y	–	Y	Y
Khalili et al. [46]	–	Y	–	–	Y	Y	Y	Y	–	Y	–	Y	Y
This Study	**–**	**Y**	**Y**	**–**	**Y**	**Y**	**Y**	**Y**	**Y**	**–**	**Y**	**Y**	**Y**

Note: Q1 = Qualitative; Q2 = Quantitative; C=Copula; P=Process; U=Utility; O = Organisational;
T = Technological; CR = Correlated Risk; R = Regulations; M = Market Forces; A = Attacker Efforts; IP=Insurance Premium; SI=Security Investments

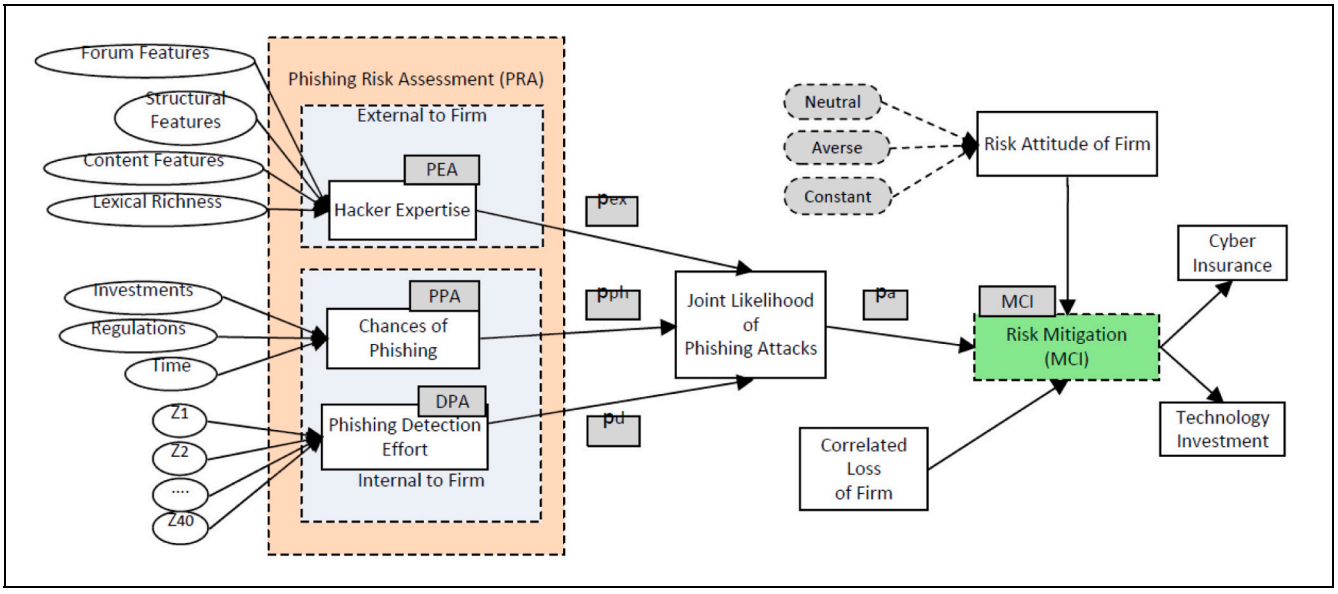


Fig. 2. Our proposed framework.

Table 2
Glossary of notations and abbreviations used in this study.

Notation	Meaning
PRA	Phishing Risk Assessment
MCI	Mitigation through Cyber Insurance
PEA	Probability of an expert attacker
PPA	Probability of phishing attacks
DPA	Detection of Phishing Attacks
RN	Risk-neutral
RA	Risk-averse
CN	Constant-risk
PEA Module	
$X_1, X_2, \ldots, X_{14}$	Determinants of an attacker's expertise based on features from Biswas et al. [8]
j	$j = 1, 2$ denotes two levels {*novice*, *expert*}
p_{ex}	Probability of accurately identifying an expert attacker (PEA); $0 < p_{ex} < 1$
PPA Module	
N_t	Number of firms participating in the CSI-FBI survey in a year "t"
Y_t	Number of firms reporting a cyber-attack in a year "t"
Sec_t	Installation of security technologies and compliance with industry standards
Reg_t	Introduction of legal and regulatory factors in a year "t."
p_{ph}	Probability of phishing attacks (PPA); $0 < p_{ph} < 1$. Adapted from Mukhopadhyay et al. [64]
DPA Module	
$Z_1, Z_2, \ldots, Z_{40}$	Determinants to classify email traffic into two categories based on the input features of the URL. Adapted from Sahingoz et al. [73]
p_d	Probability of successful detection in DPA
Loss Calculation	
p_a	The joint probability of a phishing attack
E(L)	Expected loss of a firm
$F_1(p_a)$	The continuous marginal cumulative distribution function of phishing risk
$F_2(L)$	Continuous marginal cumulative distribution function of a firm's loss
C	Copula function to combine two continuous marginal distributions
$F_{p_a,L}$	The joint cumulative distribution function of the expected loss
$\rho(p_a, L)$	Correlation between phishing-risk (p_a) and firm's loss (L)
MCI Module	
R	Revenue earned by the firm
B	Budget allocation for cybersecurity-related tasks
I	Premium paid for cyber insurance
CT	Cost of procuring cybersecurity-related technological measures
U(x)	Utility function
$U_{insured}$	Utility of the firm in the insured state
$U_{not-insured}$	Utility of the firm in a non-insured state

lead to possible phishing attacks only and by leaving aside other cyber-attack forms. Therefore, we compute the probability of an expert phisher using forum features as predictors, given by (2).

$$p_{ex} = P(Y = j | X_1 = \alpha_1, X_2 = \alpha_2, \ldots, X_{14} = \alpha_{14}) \quad (2)$$

In other words, we answer the following sub-research question within RQ1–.

RQ1 (a). What is the probability that a firm correctly detects an "expert" phisher based on dark-forum messages and discussions?

3.1.2. PPA module: estimation of the probability of phishing attacks (p_{ph})

For our second module within risk assessment, PPA, we chose the CSI–FBI [31] and Ponemon Surveys administered in government, private, and other diverse industries. According to these surveys, all N_t firms had used secure networks with active firewalls, data encryption, and up-to-date anti-virus software [29,31]. They also resorted to regulatory and legal compliance (such as auditing, up-to-date IT security policies, reported intrusion(s) to law enforcement agencies and reported intrusion(s) to legal counsels [25]. Yet Y_t out of N_t firms reported about a cyber-attack in a year "t", while the remaining $(N_t - Y_t)$ firms were secure. So, Y_t has two possible states (i.e., *attack*: $Y_t = 1$ or *no attack* $Y_t = 0$) with a probability of p_{ph} and $\left(1 - p_{ph}\right)$ respectively. And we build a generalised linear model (GLM) with the explanatory variables: t, Sec_{t-1}, Reg_{t-1}. We consider a lag of one year (i) from the time of security deployment, Sec_{t-1} and impacted users, Y_t, (ii) between the legal and regulatory factors Reg_{t-1} and impacted users Y_t.

$$P(Y_t = 1) = {}^{N_t}C_{Y_t} p_{ph}^{Y_t} \left(1 - p_{ph}\right)^{(N_t - Y_t)} \quad (3)$$

$$p_{ph} = P(Y_t = 1 | X = t, Sec_{t-1}, Reg_{t-1}) = 1/\left(1 + e^{-V_t}\right)$$

$$= 1/\left(1 + e^{-(\beta_0 + \beta_1 t + \beta_2 Sec_{t-1} + \beta_3 Reg_{t-1})}\right) \quad (4)$$

In other words, we answer the following sub-research question within RQ1–.

RQ1(b). What is the probability of a phishing attack on the firm even though it has invested in security technologies and practised secure IT policies and regulations?

3.1.3. DPA module: estimation of the probability of detection of phishing attacks (p_d)

For our third module, within risk assessment, DPA guides a firm to classify email traffic into two categories based on the input features, such as address-bar features, website URL abnormality, HTML/JavaScript, and website statistics, and classify URLs as *phishing* or *genuine*. We adopted the unique phishing dataset from Sahingoz et al. [73], which consists of 73,575 URLs with 36,400 legitimate and 37,175 phishing URLs for our experiments. Therefore, it emerges as a classification problem, and we answer the following sub-research question within RQ1–.

RQ1(c). What is the probability of successful detection of a phishing URL?

Mathematically, it can be represented as shown in (5).

$$p_d\,(Y = j|Z_{1=}\alpha_1, Z_{2=}\alpha_2, \ldots, Z_{n=}\alpha_{40}) = p(Y)\prod_{i=1}^{40} p(Z_i|Y) \Big/ p(Z_i) \tag{5}$$

3.1.4. Calculating the expected loss of a firm using Copula

Based on our proposed framework (Fig. 2), a firm needs to compute its expected loss after suffering a phishing attack. In this case, the loss of a firm (L) and the risk of a phishing attack (p_a) are correlated. Therefore, we need a method to combine both and then calculate E(L). Now, based on past literature, exponential distributions can well represent (i) the risk of a phishing attack (p_a) and related ROC of detection software [14]; (ii) loss suffered by a firm (L) due to cyber-attacks [64]. Therefore, given that the marginal distributions for p_a and L are exponential, we chose the Copula technique to combine the two continuous univariate distributions to create a joint distribution [64,84]. Archimedean Copula has been used by Yang et al. [101] to combine multivariate distributions with marginals that follow the univariate generalised beta distribution to model accident-liability claims, which are typically long-tailed data. Whelan [98] built multivariate Archimedean Copulas from univariate exponential distributions, including beta. Therefore, for this study, we chose Archimedean Copula function "C" to derive the c.d.f of the joint distribution for expected loss E(L) [50,99]. Archimedean Copula is well-suited for combining marginal distributions with long tails, as in our study, both p_a and L exhibit long-tail behaviour.

$$E(L) = F_1(p_a)^*F_2(L)^*C[F_1(p_a)F_2(L)\,,\rho(p_a, L)\,] \tag{6}$$

$$F_{p_a,L}(p_a, L) = C[F_1(p_a)F_2(L)\,] \tag{7}$$

Later, we verified that p_a follows beta distribution, while L follows exponential distribution (please see Figs. 6 and 7).

3.2. Mitigation through Cyber Insurance (MCI)

In the next step, we compute the required cyber-insurance premium or IT security investments needed by the firm, which, now having suffered a phishing attack, will need those tools to recover from its losses. Insurance is a policy tool that can reward individuals for adopting loss mitigation steps ahead of a cyber-attack by paying premiums and indemnifying their losses in case of cyber-attacks [51,108]. We propose a decision model based on utility theory to manage cyber-risks – either *transfer risk to a third-party* or *manage in-house*. If phishing detection fails, then the firm stands to suffer losses. Still, it can mitigate those losses by subscribing to cyber insurance policies. If a firm has opted for cyber insurance, it must have paid a premium and may recover fully from the loss under the condition of complete indemnification. However, the firm suffers financial loss if there is no cyber insurance. In addition, it could happen that due to the high risk of cyber-attacks (i.e., high p_a), a cyber-insurance firm refuses to provide insurance coverage or charge exorbitant premium fees to the focal firm (in this study). In those circumstances, the focal firm might think of a mixed strategy – first, reduce the high risk of cyber-attacks using security technologies and then purchase cyber-insurance. Based on these steps, we present the decision-making problem in Eqns. (8)–(10) and presented in Fig. 3. Based on these premises, we arrive at the next set of research questions –.

RQ2(a). What is the optimal premium for cyber insurance that the firm needs to pay?

RQ2(b). What is the rate of investment in complementary cybersecurity technology?

Mathematically, we represent the decision problem as follows:

Objective: To Find $(I, \Delta CT)$

$$\text{s.t.(i) } E\,[U_{insured}] \geq E\,[U_{not-insured}] \text{ and (ii) } B = (I + CT) \tag{8}$$

$$E\,[U_{insured}] = p_{ex}p_{ph}p_d U(R - CT - I) + p_{ex}p_{ph}(1 - p_d)U(R - CT - I)$$
$$+ p_{ex}(1 - p_{ph})p_d U(R - CT - I) + p_{ex}(1 - p_{ph})(1 - p_d)U(R - CT - I) \tag{9}$$

$$E\,[U_{not-insured}] = p_{ex}p_{ph}p_d U(R - CT) + p_{ex}p_{ph}(1 - p_d)U(R - CT - L)$$
$$+ p_{ex}(1 - p_{ph})p_d U(R - CT) + p_{ex}(1 - p_{ph})(1 - p_d)U(R - CT) \tag{10}$$

4. Data description and feature engineering

In this section, we briefly describe the empirical data for our study. We present the descriptive statistics for the data in each module – PEA, PPA, DPA, followed by MCI. Firstly, for the PEA module, we used data from hackhound.org [74]. It consists of messages generated between October 2012 and September 2015. We improved upon Biswas et al. [8] by (i) filtering out only those messages that could lead to possible phishing attacks and (ii) merging contiguous classes, leading to two major groups – "expert" and "novice". A quick inspection of the class data reveals a class imbalance problem, (iii) applying the majority-weighted minority oversampling to remove the class imbalance problem in the data [19]. Table 3 details the variables used to build the PEA classifier, while Table 4 presents the attacker roles in the input file for PEA.

Secondly, for the PPA module, we created a combined dataset from the CSI–FBI time series data and Ponemon Surveys [31,37]. This data has been previously used to predict the growth of various attacks based on the historical number of attacks and defensive efforts to reduce them yearly [64]. Table 5 reports the descriptive statistics. E.g., based on Table 5, an average of 137 firms suffered *denial-of-service* attacks. In contrast, 53 suffered *financial fraud* attacks, 459 firms implemented security measures, and 43 firms adopted legal and cybersecurity governance controls in the $(t\text{-}1)^{th}$ year.

Thirdly, in the DPA module, we adopted the phishing dataset from Sahingoz et al. [73], consisting of 73,575 URLs with 36,400 legitimate and 37,175 phishing URLs. We performed some data pre-processing first and extracted the forty NLP-based features from the dataset. First, let us look at a URL. It consists of some meaningful words, while the rest could be gibberish or special characters, which separate some components of the address. For instance, a dot mark (".") is used to separate the second-level domain (SLD) and the top-level domain (TLD). Similarly, characters can also be used in conjunction with "=", "?", "&". Next, we applied the *hunspell* package in R to check its spelling and whether it exists in the dictionary. We also calculated the Levenshtein distance between two word vectors using the *stringdist* package. Next, we undertook a feature-selection scheme based on the *varimp* package. Table 6 reports the ten predictors that emerged as significant. In addition, we executed the word vectorisation step to generate ten additional predictors. Finally, in Table 7, we present the descriptive statistics of the univariate exponential distributions for each attack type.

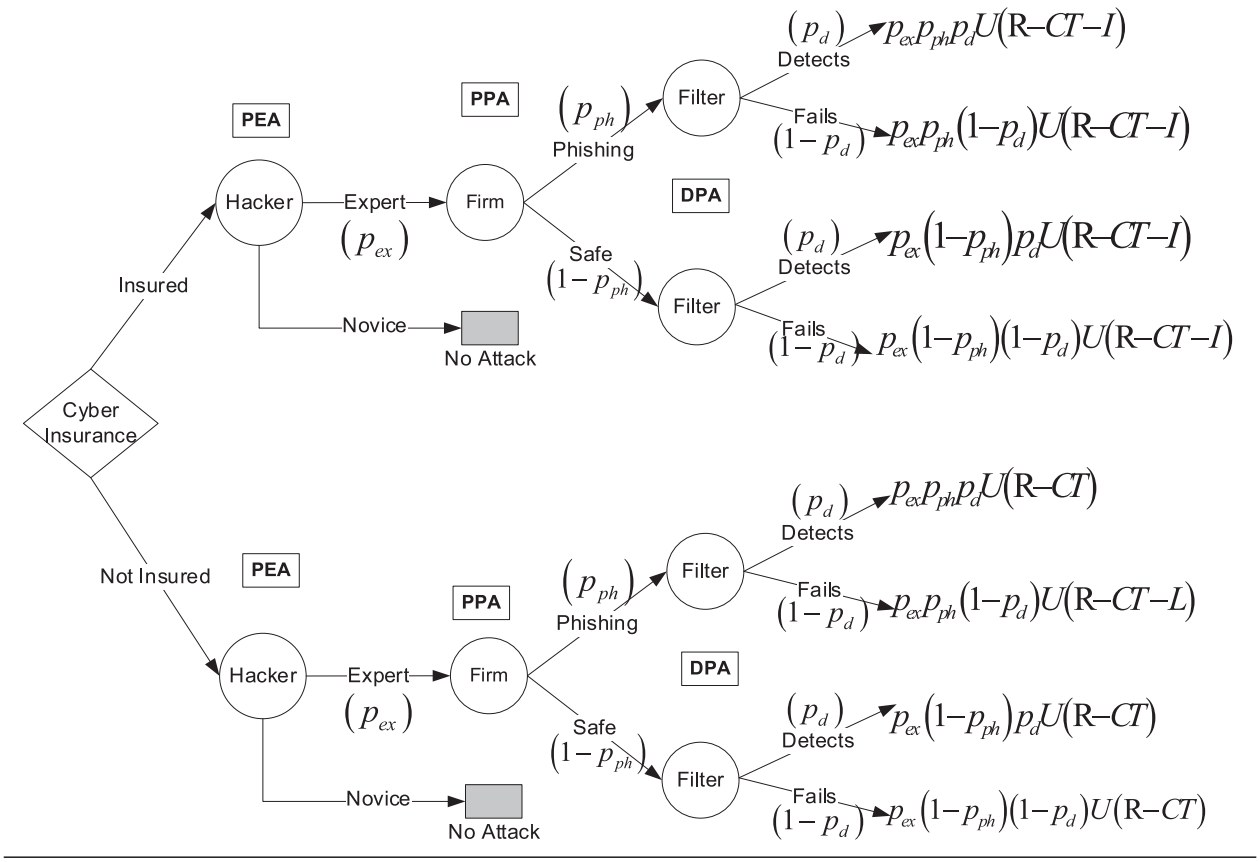


Fig. 3. – Decision tree for the insured and non-insured states of a firm.

Table 3
List of variables used to build the PEA Classifier.

Construct	Variable	Description	Literature Source
Forum Feature	X_1	Time spent (days)	Developed from Benjamin et al. [6]; Biswas et al. [8]
	X_2	Threads participated	
	X_3	Messages posted	
	X_4	Replies per thread	
	X_5	Discussions initiated	
	X_6	Sequence of messages	
	X_7	Length of messages	
	X_8	Attachments	
Structural Feature	X_{10}	Positive Sentiment	Developed from Biswas et al. [8]; Benjamin et al. [6]; Chen et al. [17]; Samtani et al. [76]
	X_{11}	Negative Sentiment	
	X_{12}	Keyword Usage	
	X_{13}	Attachments shared	
	X_{14}	Characters spent	

Table 4
Attacker roles in the input for PEA.

	Original Data (imbalance ratio = 0.078)			**After oversampling** (imbalance ratio = 0.817)		
Type	**Attacker Role**		**Count**	**Attacker Role**		**Count**
	Expert	Novice		Expert	Novice	
Training	38	489	527	412	489	901
Testing	10	123	133	88	123	211
Total	48	612	660	500	612	1112

Table 5
Descriptive statistics of the attack data used in the PPA module.

Measures	**Types of Attacks**					**Security**$_{t-1}$	**Legal**$_{t-1}$
	DoS	FF	SP	TPI	UA		
Mean	137	53	105	73	188	459	43
Std. Dev.	58	15	56	37	91	119	18
Count	18	18	18	18	18	18	18

DoS=Denial of Service; FF=Financial Fraud; SP=System Penetration; TPI = Theft of Proprietary Information; UA = Unauthorised Access

5. Discussion of results

In this Section, we first report the results from the PRA module, followed by the MCI module. PRA has three sub-modules - PEA, PPA and DPA. MCI consists of a loss module and CI. We performed our analyses using the R platform. In the PEA module, we used logistic regression to classify the phishers into "expert" and "novice", finally estimating p_{ex} . The training was performed with 901 records, while testing was performed with 211 records (~80:20 ratio). The PPA module used a generalised linear model to estimate p_{ph}. Here, training was performed with 14 records, while the testing was done with 4 records (80:20 ratio). The DPA module classified URLs as "phishing" and "genuine", thereby estimating p_d . We used six classifiers: CART, Gini DT, bagger DT, PCA DT, NBC, and SVM. The training was performed with 58,860 records, while the testing was performed with 14,715 records (80:20 ratio). In

Table 6
Top ten predictors of phishing URLs in the DPA module ($N = 73{,}575$ records).

Feature No.	Importance Measure	Feature Description	Literature Source
Z8	4.623	The average length of the detected adjacent words	Derived from Sahingoz et al. [73], Jain and Gupta [39], Jain and Gupta [40], Mohammad et al. [61]
Z26	3.817	Top-level domains, e.g. ["com", "org", "net", "edu", "gov"]	
Z14	3.242	Number of words in URL created with random characters.	
Z15	3.083	Target firm(s) in the URL	
Z7	2.760	Number of adjacent words	
Z29	2.745	Use of Puny Code	
Z6	2.307	Std. dev. of word lengths in the word list.	
Z22	2.268	Is a registered domain created with random characters?	
Z1	2.262	Raw Word Count	
Z16	2.053	Number of target keyword in the URL	

Note: We present the top ten predictors only due to the limitations of space in the manuscript.

Table 7
Descriptive statistics for univariate exponential loss.

	DoS	FF	SP	TPI	UA
Mean	729	3870	675	5869	1460
Std. Dev.	729	3870	675	5869	1460
Count	18	18	18	18	18

DoS=Denial of Service; FF=Financial Fraud; SP=System Penetration; TPI = Theft of Proprietary Information; UA = Unauthorised Access.

the loss module, we computed the joint distribution of attack, where p_a represented an exponential-beta distribution, contingent on – *phisher expertise*, *act of phishing*, and *misdetection of phishing URLs*. Next, we applied Archimedean Copula to combine two univariate distributions [50]. In the CI module, we computed $(I, \Delta CT)$ for risk-neutral, risk-averse, and constant-risk firms.

5.1. Phishing-Risk Assessment (PRA)

5.1.1. PEA classifier and estimation of p_{ex}

Table 9 illustrates the regression coefficients from the logistic regression with target class "Expert" and "Novice" with probability defined in Eqns. (1) and (2). The B-coefficients determine the log-odds ratio. It shows that strongly supported variables X_1, X_2, X_3, X_4, and X_{10} are significant predictors of phisher expertise: X_6, X_8, and X_{11} are partially supported in our model, while X_5, X_7, and X_9 are not substantial. The Nagelkerke-R^2 and log-probability values of the logistic regression model are 0.653 and 142.873. Table 10 reports p_{ex} in terms of the model's classification accuracy for each group. The per-class classification accuracy is 89.475% for the "Expert" group of phishers and 97.414% for the "Novice" group of phishers, which is a significant improvement over [8]. The overall classification accuracy is 93.839%.

Table 10
Confusion matrix for PEA classifier (Testing Phase, $N = 211$).

	Predicted		Total	(1-p_{ex})	p_{ex}	Accuracy (%) (Biswas [8])
Observed	Expert	Novice				
Expert	85	10	95	0.895	0.105	80.000
Novice	3	113	116	0.974	0.026	79.110
Total	88	123	211	0.938	0.061	80.570

5.1.2. PPA classifier and estimation of p_{ph}

Table 11 summarises the coefficients, standard errors, and t-statistics of the linear model for each of the five secondary attacks originating from phishing. The time variable (t) is significant for all types and reduces with an increase in the dependent variable $\left(p_{ph}\right)$. The result indicates that, with passing time, firms become conscious about cyber-attacks, install relatively stronger access control mechanisms (Sec_{t-1}) and abide by regulatory frameworks, implement security controls and IT policies, take legal actions (Reg_{t-1}), leading to the reduction of attacks $\left(p_{ph}\right)$. We applied a binomial logit to generate the probability of DoS, FF, SP, TPI, and UA attacks (Table 11). Fig. 4 illustrates the logit classifier of PPA for the training and testing phases. We observe that the logit linear model performs the best for SP attacks and the worst for FF attacks.

5.1.3. DPA classifier and estimation of p_d

Table 13 reports that the classifier can successfully identify phishing URLs with a true positive (TP) rate of 92.35% and a true negative (TN) rate of 96.42%. Precision and recall values are 0.953 and 0.923, respectively. The performance metrics of our DPA module are comparable with [61,73] and better than Varshney et al. [92]. Table 12 shows that Bagger with Decision Tree achieves the highest classification accuracy among the six algorithms employed in our DPA module.

Table 11
Coefficient estimates of the PPA Logit Classifier (Training Phase: 1998 to 2011).

	Constant	*Time* (*t*)	Sec_{t-1}	Reg_{t-1}
	DoS Attacks			
β	−0.658***	−0.035***	−0.014	−0.001
S.E.	0.201	0.018	0.009	0.001
	SP Attacks			
β	−0.021	−0.127***	−0.006	−0.001
S.E.	0.330	0.024	0.009	0.001
	UA Attacks			
β	0.812***	−0.198***	−0.006	−0.001
S.E.	0.155	0.023	0.014	0.000
	FF Attacks			
β	−1.578***	−0.055**	−0.005	0.000
S.E.	0.343	0.022	0.011	0.001
	TPI Attacks			
β	−0.756**	−0.212***	−0.008	−0.001
S.E.	0.259	0.014	0.012	0.001

Note: ***, **, * denote significance at 0.01, 0.05, and 0.10 levels respectively for the 2-tail test.

Table 9
Coefficients of Logit Model for PEA classifier (Training Phase, $N = 527$).

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
B	1.941**	−1.594**	2.236**	−0.018*	−2.300	0.744*	0.062	−0.491*	−0.265	0.696***	0.338*
Exp(B)	6.966	0.203	9.356	0.982	0.100	2.104	1.064	0.612	0.767	2.006	1.402

Nagelkerke-$R^2 = 0.653$; Log Probability = 142.873
Note: ***, **, * - denotes significance at the 0.01, 0.05, and 0.10 level respectively, for the 2-tail test

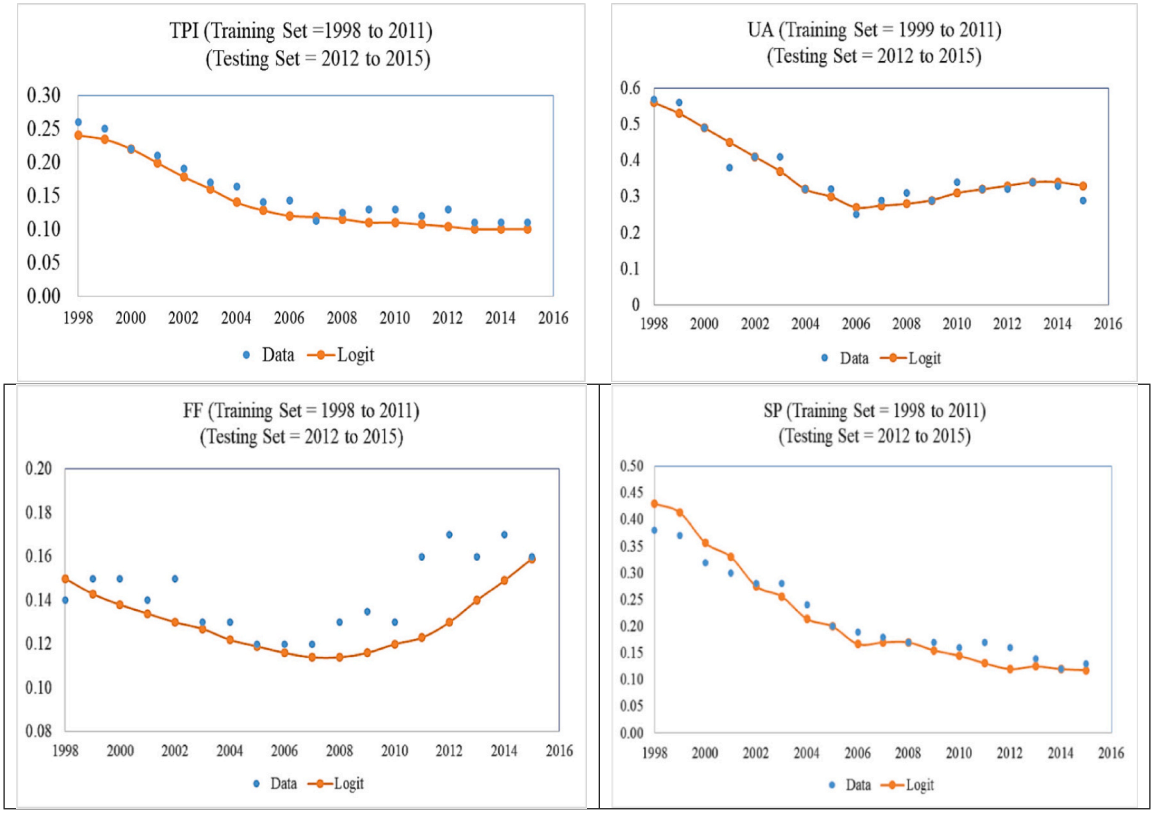


Fig. 4. Attack probabilities from PPA classifier built with training and testing datasets.

Table 13
Performance metrics of classification algorithms (with testing dataset).

Metrics	Bagger DT	NBC	CART	Gini DT	PCA DT	SVM
Accuracy (p_d) **(%)**	94.612	91.765	87.734	91.365	87.887	92.530
Precision (%)	95.363	94.332	84.954	89.722	88.074	93.038
Recall (%)	92.351	86.649	88.073	90.722	84.955	89.901
F-Score (%)	93.833	90.327	86.485	90.219	86.486	91.443

Note: DT = Decision Tree; Bagger DT = Bootstrap aggregating DT; NBC = Naïve Bayes Classifier; CART = Classification and Regression Tree; Gini DT = DT using Top 10 Predictors chosen w.r.t Gini impurity; PCA DT = Principal Component Analysis followed by DT; SVM = Support Vector Machine.

Therefore, we continue our subsequent analysis with the bagger algorithm. Figs. 5(a) and 5(b) illustrate the receiver operating characteristic (ROC) curve for each output class: *phishing* and *genuine* URLs. Next, we fit the posterior probabilities from the ROC curve and classification accuracy generation for the phishing and genuine URLs.

5.1.4. *Modelling the probability of attack p_a*

Fig. 6(a) shows that Beta ~ (4.208, 0.227) best fits the probability of attack (p_a). Similarly, Fig. 6(b) shows that Beta ~ (4.623, 0.286) best fits the probability of attack avoidance$(1 - p_a)$.

5.2. *Modelling the correlated loss E (L)*

Following [50] [66,99], the product-moment of bivariate loss E (L) and joint attack (p_a) with L ~ Exp (λ) and p_a ~ Beta (4.208, 0.227) is:

$$E\left[L^m p_a^n\right] = \{\lambda^m \Gamma(m+a+b)\Gamma(m+n+a)\}/\{\Gamma(a)\Gamma(m+n+a+b)\} \quad (11)$$

Using (11), it follows that $E[L] = 0.958\lambda$ and $Var(L) = 3.952\,\lambda^2$ where m = 1 and *n* = 1; m = 2 and n = 2 respectively. Table 14 shows the first and second moments of the joint loss distribution. Fig. 7 (a) illustrates the joint loss distribution for DoS attacks with a high correlation [i.e., $\rho(p_a, L) = 0.950$]. X-axis represents the variance of the loss distribution, while Y-axis represents the probability of attack. The combined scatter plot represents bivariate distribution with an increased S.D. of 1449 (instead of 729), signified by an extended tail, and a reduced mean of 698 (instead of 729 in case of a univariate loss distribution). Similarly, Fig. 7(b) illustrates the joint loss distribution for FF attacks with an increased S.D. of 7693 (instead of 3870) and a reduced mean of 3707 (instead of 3870). Fig. 7(c) illustrates the joint loss distribution for UA attacks with an increased S.D. of 2902 (instead of 1460) and a reduced mean of 1399 (instead of 1460). (See Table 15.)

5.3. *Mitigation through Cyber Insurance: sensitivity analysis to determine I and CT*

We find that the means of bivariate distribution remain almost unchanged for each case while the variance increases four-fold. Therefore, our revised model is a better estimator of the risk arising from cyber-attacks, which is similar to the *risk of ruin* in modelling insurance and financial risks [90]. Fig. 8 shows that a *risk-averse* firm needs to pay a much higher premium than a *constant-risk*, and a *risk-neutral* firm does. For example, in the (Hi, Hi) zone, when the probability of an attack is 0.90, payable premiums are 816, 709, and 675 for RA, CR, and RN firms with the payable premiums in the ratios of 1.209: 1.050: 1.000.

Next, in the (Hi, Lo) zone, for the probability of an attack 0.55, the payable premiums are 466, 433, and 412 for RA, CR, and RN firms in the ratios of 1.131: 1.051: 1.00. And, in the (Lo, Lo) zone, if the probability of an attack is 0.30, the premiums are 240, 224, and 217 for RA, CR, and RN firms, in the ratios of 1.106: 1.032: 1.00. Further, when the probability of an attack is very high (i.e., 0.90), a third-party cyber-risk insurer may not be keen to cover the risks. Therefore, a firm must first reduce it by investing in security technologies and buying cyber insurance [8; 64].

Fig. 8(a) also shows that $\Delta CT/\Delta p_a$ for RA: CR: RN firms is in the ratio

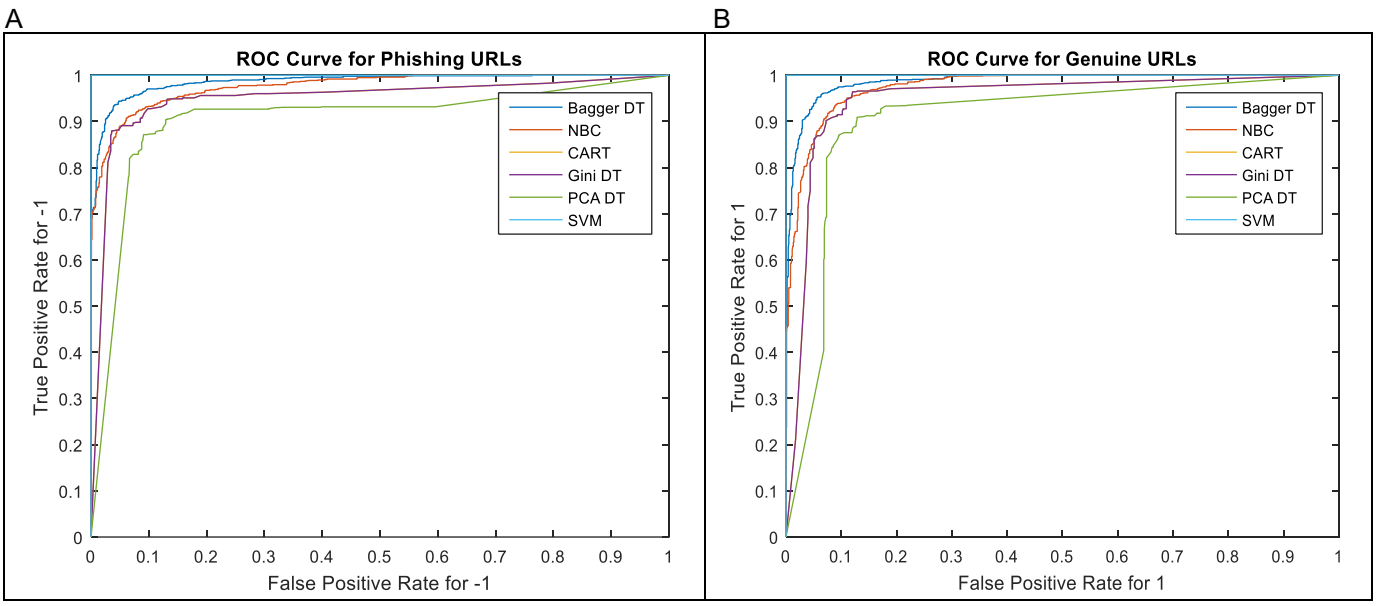


Fig. 5. (a) – ROC for phishing URLs.
(b) - ROC for genuine URLs.

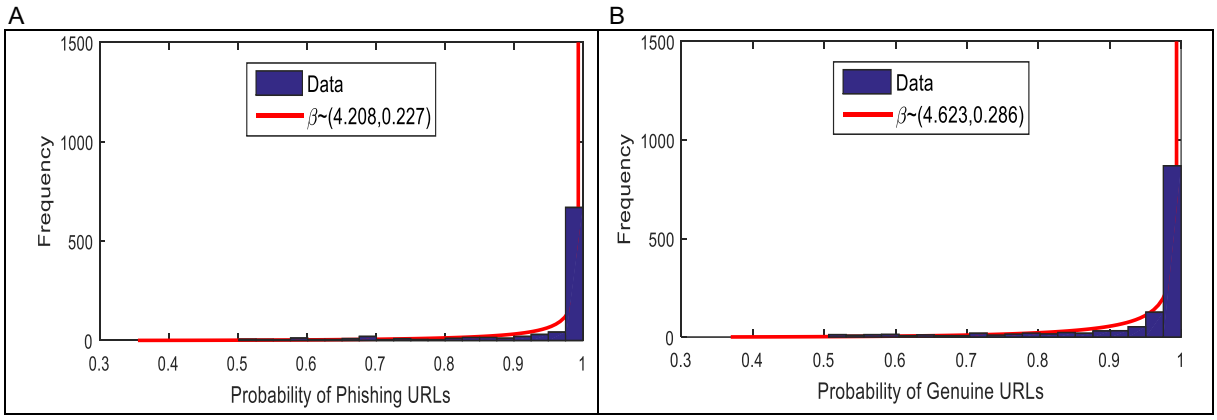


Fig. 6. (a) – Beta distribution fit for the probability of attack (p_a).
(b) – Beta distribution fit for the probability of attack avoidance (1-p_a).

Table 14
Descriptives for bivariate loss generated from Eq.(11).

Mean	698	3707	647	5623	1399
Std. Dev.	1449	7693	1342	11,667	2902

of 1.177:1.011:1.000 for A1 ➔A2; B1 ➔B2 and C1 ➔ C2, respectively. Similarly, $\Delta CT/\Delta p_a$ for RA: CR: RN is in the ratio of 1.048:1.038:1.000 for $A_2 \rightarrow A_3$, $B_2 \rightarrow B_3$ and $C_2 \rightarrow C_3$, respectively. So, it becomes reasonable for an RA firm in the (Hi, Hi) zone to move to (Hi, Lo) and then to the (Lo, Lo) zone by investing in security technologies such that its {loss, premium} are within permissible limits to procure CI in the next stage.

Fig. 8(b) then compares the incremental investment in IT security vis-à-vis loss in the (Hi, Hi) zone (at $p_a = 0.90$); $\Delta CT/\Delta p_a$ for RA, CR, and RN is in the ratio of 1.307: 1.087: 1.000. Fig. 8(c) compares the incremental investment in IT security in the (Hi, Lo) zone (at $p_a = 0.55$), $\Delta CT/\Delta p_a$ for RA, CR, and RN is in the ratio of 1.195: 1.053: 1.000. Finally, it must bring down the probability from the (Hi, Lo) zone to the (Lo, Lo) zone so that the insurance company can offer coverage. Fig. 8(d) compares the incremental investment in IT security in a firm's (Lo, Lo) zone (at $p_a = 0.30$), $\Delta CT/\Delta p_a$ for RA, CR, and RN is in the ratio of 1.110: 1.029: 1.000. So, it becomes prudent for an RA firm in the (Hi, Hi) zone to immediately reduce its propensity of being attacked by investing in security technologies.

6. Implications and contributions

This study examines correlated loss suffered by a firm from attacks launched by expert phishers [1] by applying a decision-theoretic approach for cyber-risk assessment and mitigation. Our analysis also applied the Archimedean Copula [50] to compute the bivariate loss [66] while recommending stage-wise technology investment and cyber-insurance premiums for firms [51,57].

6.1. Contributions to IS research

Our study offers three important academic contributions to the relevant IS literature, particularly to the growing body of studies on *explainable AI*, *cyber-risk management* and *the dark side of IS/IT*. First, our framework built an anti-phishing filter using data from malicious URLs with unique and interpretable predictors. Our findings are also

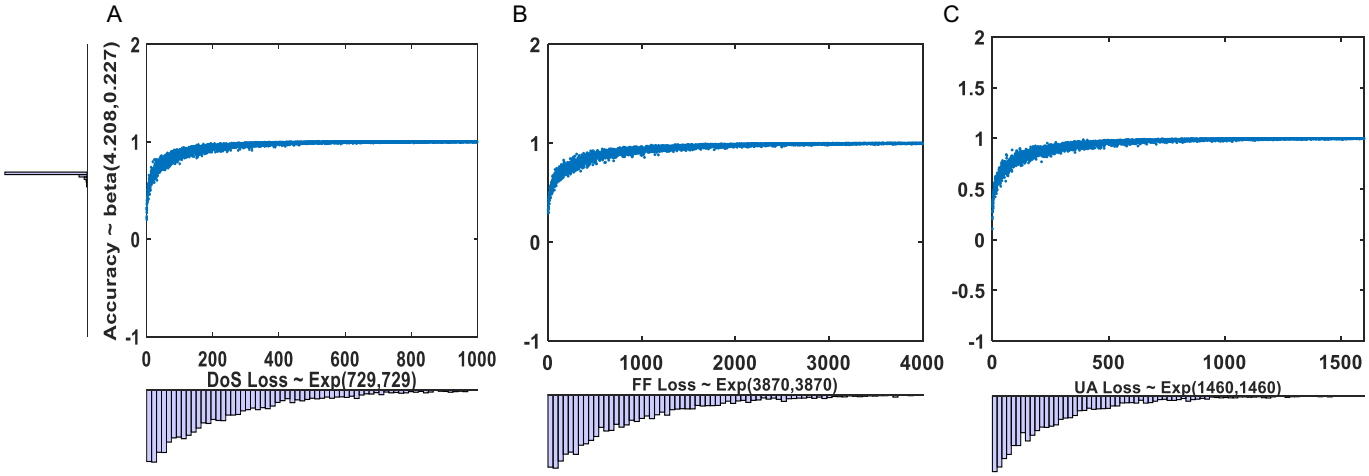


Fig. 7. **(a)** – Joint loss for Dos.
(b) – Joint loss for FF.
(c) – Joint loss for UA.

Table 15
Optimal insurance premiums for different risk attitudes of firm.

Risk level	U(x)	Upper Limit of I	$\Delta CT/\Delta p_a$
RA	$\begin{cases} x, \text{insured}, \\ E(x) - 0.5k_1 Var(x) \text{ uninsured}, \end{cases}$	$0.5\, p_{ph}(1-p_d)\, L(Lk_1+2)$	$(-2L)$
CR	$1 - e^{-x/k_2}$	$k_2 log_e\left\{1 - p_{ph}(1-p_d)\left(1-e^{-L/k_2}\right)\right\}$	$\left(e^{-L}\right)$
RN	x	$p_{ph}(1-p_d)\, L$	$(-L)$

Note: $k_1 = 0.0000465$; $k_2 = 100000$

congruent with the relevant body of literature on anti-phishing techniques [39,44,61,62,89,100], however, with more interpretability and explainability. Additionally, it identified the significant predictors using principal component analysis followed by variable importance schemes. These steps could offer a simplified yet actionable lens for cyber-risk management. Next, we applied a Beta distribution to explore a machine-learning model and the classifier's accuracy [24], an approach similar to IDS analysis [56,60] but experimentally more efficient to *exponential* [99] or *uniform* [7] distributions.

Second, our study demonstrated that undetected URLs lead to residual risk, which in turn generates a correlated loss, a unique finding missing from extant literature [12,42,48,53]. In addition, the framework estimated the per-unit ratio of investments in technologies versus insurance premiums payable to reduce the risk of future attacks. Our study adds to the existing literature on cyber-risk management [30,63,64]. This finding is by Gordon and Loeb [29], while we additionally allowed supplemental insurance and IT security for cyber-risk mitigation[15] and, therefore, added to the nascent literature [46,58,103]. Then, our framework extended past studies that had applied Clayton Copula [36] and Gaussian Copula [63] to model cyber-risks. At the same time, we applied Archimedean Copula to build the joint distribution for calculating the expected loss.

Third, our study also contributed to the literature on darknet and hacker forums. Our framework computed the probability of an expert phisher among the prevalent attacker population [8], using a simplified role-based classification into "expert" and "novice" with prediction accuracy, precision and recall scores in a two-class problem instead of a multi-class one [76]. In this manner, our findings align with past studies on darknet and hacker forums [5,6,8,32,43,74,75,76,106]. However, we simplified the classification problem by reducing it from a multi-class scenario into a two-class one, thereby reducing the complexity in understanding and bringing more explainability to the models.

6.2. Practice implications

Our study offers two actionable recommendations for managers. First, our framework estimates the per-unit change in "technology investment" versus "reduction in the probability of attack" and complementary insurance. This decision will help firms plan for cyber-risk mitigation and guide the insuring agent to design realistic insurance products contrary to the belief that cyber-insurance markets are ineffective [3,4,52]. Next, our framework is a more accurate estimator of residual cyber-risks, such that a risk-averse firm needs to pay a much higher premium than a constant-risk, and a risk-neutral firm does. For instance, firms in the (Hi, Hi) zone with a very high probability of attack, payable premiums are in the following order $premium_{RA} \gg premium_{CR} \gg premium_{RN}$. Further, when the probability of an attack is very high, a third-party cyber-risk insurer may not be keen to cover the risks. Therefore, a firm must first reduce it by investing in security technologies, and then buy cyber insurance. Next, firms in the (Hi, Lo) zone with a medium probability of attack, payable premiums are in the following order: $premium_{RA} > premium_{CR} > premium_{RN}$. Here, a firm with moderate propensity of attack might think of investing relatively lesser than its peer from the (Hi, Hi) zone. So, it becomes reasonable for an RA firm in the (Hi, Hi) zone to move to (Hi, Lo) and then to the (Lo, Lo) zone by investing in security technologies such that its {loss, premium} are within permissible limits to procure CI in the next stage.

Second, based on the computation of the probability of different cyber incidents arising from phishing attacks on a firm, managers can now implement stronger access controls and take legal action over time [35], reducing the chances of future attacks. Sometimes, a suitable anti-

[15] PwC (2021). Are insurers adequately balancing cyber risk and opportunity?

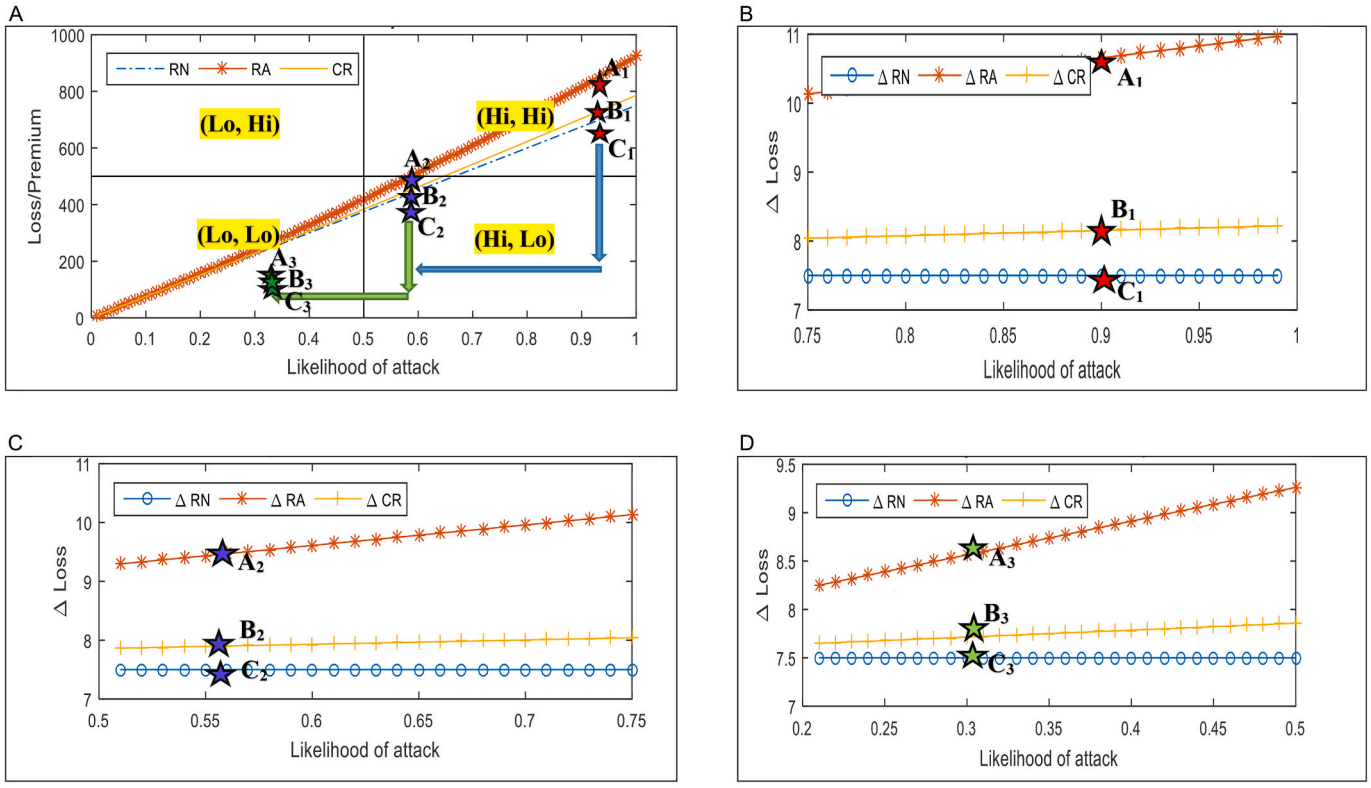


Fig. 8. (c) – Sensitivity plot for $\Delta CT/\Delta p_a$ [Hi, Lo].
Note: A2 (0.55, 466) B2 (0.55, 433) C2 (0.55, 412).
(d) – Sensitivity plot for $\Delta CT/\Delta p_a$ [Lo, Lo].
Note: A3 (0.30, 240) B3 (0.30, 224) C3 (0.30, 217).

spam and anti-phishing policy could improve cybersecurity risks and reduce attacks [45]. In addition, the *organisational use of perimeter security* and *regulatory steps,* including frequent auditing and reporting intrusion(s) to enforcement agencies and legal counsels [25], could reduce cyber-attacks. Additionally, our framework proposed novel cyber-intelligence for a firm and a cyber-insurer by mining phisher messages with the help of keyword-based TF-IDF [8] and sentiment scores to identify potentially harmful actors present in the phisher CoP [8]. Therefore, firms can now apply machine-learning models to examine incumbent hacker forums and their messages for better decision-making, risk mitigation and allocation of IT security budgets, depending on the type of attack and the attackers involved.

In essence, we built a comprehensive cyber-risk assessment and mitigation framework to identify (i) expert-novice phishers, (ii) factors responsible for phishing attacks, (iii) important features for phishing URLs, and (iv) investment mix of cyber-insurance and IT security technology to mitigate cyber-risks at different attack levels. These managerial contributions are unique to our study and have not been reported in the cyber insurance domain.

7. Conclusion

This study introduces an XAI-based ML framework to examine cyber-risk management and then propose remedies in the form of complementary investment in IT security or cyber-insurance or both. In the process, we endeavoured to apply traditional ML algorithms and build a framework that satisfied the following two tenets of XAI as outlined by DARPA[16]: (i) *explainable model*: build more understandable models, yet preserving an elevated level of accuracy, and (ii) *explainable interface*: allow users to comprehend, and effectively manage the AI system. In addition, we purposefully applied traditional ML (restricting neural networks or deep-learning) models [47] to maintain the explainability of our framework and achieve a model-agnostic global explanation [71].

To achieve these objectives, we compute the probability of phishing attacks launched by expert phishers that can adversely affect a business organization. We applied a binary classifier that replicated the anti-phishing technology (or filters) that firms typically install to block phishing URLs. Then, we proposed utility-based IT risk mitigation strategies, such as third-party cyber insurance cybersecurity investments or both, based on the firm's needs. Finally, through scenario analyses, we demonstrated that a third-party insurer would be eager to cover the phishing risk when the probability of an attack was reduced to a reasonably low level by applying relevant security technologies.

This study has a few limitations and thus allows scope for future extension. First, future academic work can focus on text mining with non-English forums to identify top hackers [22]. Second, scholars can explore logarithmic and hyperbolic utility functions for building the decision-theoretic models and then compute the insurance premiums.

Author statement

Conceptualization, Investigation, Statistical Analysis, Data Collection, Methodology, Validation, Formal Analysis, Writing – Original Draft, Writing – Review and Editing, Funding, Supervision

Baidyanath Biswas (Conceptualization, Writing - Original Draft and Editing**)**

Arunabha Mukhopadhyay (Conceptualization, Supervision, Writing - Original Draft, Reviewing and Editing**)**

Ajay Kumar (Conceptualization, Supervision, Writing – Editing and Revision)

Dursun Delen (Proofreading, Editing and Revision)

[16] Explainable Artificial Intelligence (XAI)

Declaration of Competing Interest

On behalf of all co-authors, I (Ajay Kumar) hereby declare that the disclosed information is correct and that no other situation of real, potential, or apparent conflict of interest is known to me.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] G.A. Akerlof, R.J. Shiller, Phishing for Phools: The Economics of Manipulation and Deception, Princeton University Press, 2015.
[2] M. Babagoli, M.P. Aghababa, V. Solouk, Heuristic nonlinear regression strategy for detecting phishing websites, Soft. Comput. 23 (12) (2019) 4315–4327.
[3] T. Bandyopadhyay, V.S. Mookerjee, R.C. Rao, Why IT managers don't go for cyber-insurance products, Commun. ACM 52 (11) (2009) 68–73.
[4] T. Bandyopadhyay, V. Mookerjee, A model to analyze the challenge of using cyber insurance, Inf. Syst. Front. (2017) 1–25.
[5] V. Benjamin, H. Chen, Securing cyberspace: identifying key actors in hacker communities, in: 2012 IEEE international conference on intelligence and security informatics, 2012, pp. 24–29.
[6] V. Benjamin, J.S. Valacich, H. Chen, DICE-E: a framework for conducting Darknet identification, collection, evaluation with ethics, MIS Q. 43 (1) (2019).
[7] A. Bensoussan, V. Mookerjee, W.T. Yue, Managing information system security under continuous and abrupt deterioration, Prod. Oper. Manag. 29 (8) (2020) 1894–1917.
[8] B. Biswas, A. Mukhopadhyay, S. Bhattacharjee, A. Kumar, D. Delen, A text-mining based cyber-risk assessment and mitigation framework for critical analysis of online hacker forums, Decis. Support. Syst. 113651 (2021).
[9] R. Böhme, Cyber-insurance revisited, in: Proceedings of Workshop on the Economics of Information Security (WEIS), 2005.
[10] R. Böhme, G. Kataria, Models and measures for correlation in cyber-insurance, in: *Proceedings of Workshop on the Economics of Information Security* (WEIS), 2006.
[12] I. Bose, A.C.M. Leung, Do phishing alerts impact global corporations? A firm value analysis, Decis. Support. Syst. 64 (3) (2014) 67–78.
[13] A.S. Bozkir, F.C. Dalgic, M. Aydos, GramBeddings: a new neural network for URL based identification of phishing web pages through N-gram Embeddings, Comput. Secur. 124 (2023) 102964.
[14] H. Cavusoglu, S. Raghunathan, Configuration of detection software: a comparison of decision and game theory approaches, Decis. Anal. 1 (3) (2004) 131–148.
[17] L. Chen, A. Baird, D. Straub, A linguistic signaling model of social support exchange in online health communities, Decis. Support. Syst. 130 (2020) 113233.
[19] I. Cordón, S. García, A. Fernández, F. Herrera, Imbalance: oversampling algorithms for imbalanced classification in R, Knowl.-Based Syst. 161 (2018) 329–341.
[22] M. Ebrahimi, Y. Chai, S. Samtani, H. Chen, Cross-lingual cybersecurity analytics in the international dark web with adversarial deep representation learning, MIS Q. 46 (2) (2022).
[24] S. Ferrari, F. Cribari-Neto, Beta regression for modelling rates and proportions, J. Appl. Stat. 31 (7) (2004) 799–815.
[25] E.A. Fischer, Federal Laws Relating to Cybersecurity: Overview and Discussion of Proposed Revisions, 2013 s.
[29] L.A. Gordon, M.P. Loeb, The economics of information security investment, ACM Trans. Inf. Syst. Secur. 5 (4) (2002) 438–457.
[30] L.A. Gordon, M.P. Loeb, T. Sohail, A framework for using insurance for cyber-risk management, Commun. ACM 46 (3) (2003) 81–85.
[31] L.A. Gordon, M.P. Loeb, W. Lucyshyn, R. Richardson, CSI/FBI Computer Crime and Security Survey, 2009. GoCSI.com.
[32] J. Grisham, S. Samtani, M. Patton, H. Chen, Identifying mobile malware and key threat actors in online hacker forums for proactive cyber threat intelligence, in: 2017 IEEE international conference on intelligence and security informatics (ISI), 2017, July, pp. 13–18.
[33] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G.Z. Yang, XAI—explainable artificial intelligence, Sci. Robot. 4 (37) (2019) eaay7120.
[35] O.A. Hathaway, R. Crootof, P. Levitz, H. Nix, A. Nowlan, W. Perdue, J. Spiegel, The law of cyber-attack, Calif. Law Rev. 100 (4) (2012) 817–885.
[36] H.S. Herath, T.C. Herath, Cyber-insurance: copula pricing framework and implication for risk management, in: WEIS, 2007, June.
[37] Ponemon 2020., Cost of Data Breach: Global Analysis, 2010-2020. Ponemon Institute.
[39] A.K. Jain, B.B. Gupta, Towards detection of phishing websites on client-side using machine learning based approach, Telecommun. Syst. 68 (2018) 687–700.
[40] A.K. Jain, B.B. Gupta, A survey of phishing attack techniques, defence mechanisms and open research challenges, Enterprise Inform. Syst. 16 (4) (2022) 527–565.
[42] M. Jakobsson, S. Myers, Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft, John Wiley & Sons, 2007.
[43] S. Jiang, H. Chen, J.F. Nunamaker, D. Zimbra, Analyzing firm-specific social media and market: a stakeholder-based event analysis framework, Decis. Support. Syst. 67 (2014) 30–39.
[44] Y. Joshi, S. Saklikar, D. Das, S. Saha, Phishguard: a browser plugin for protection from phishing, in: *Proceedings of the 2nd IMSAA Conference*, 2008, pp. 1–6.
[45] J. Ju, D. Cho, J.K. Lee, J.H. Ahn, Can It Clean Up Your Inbox? Evidence from South Korean Anti-spam Legislation. Production and Operations Managements, 2021.
[46] M.M. Khalili, P. Naghizadeh, M. Liu, Designing cyber insurance policies: the role of pre-screening and security interdependence, IEEE Trans. Inf. Forensics Secur. 13 (9) (2018) 2226–2239.
[47] B. Kim, J. Park, J. Suh, Transparency and accountability in AI decision support: explaining and visualizing convolutional neural networks for text information, Decis. Support. Syst. 134 (2020) 113302.
[48] C. Konradt, A. Schilling, B. Werners, Phishing: an economic analysis of cybercrime perpetrators, Comput. Secur. 58 (2016) 39–46.
[49] D. Kotz, K. Fu, C. Gunter, A. Rubin, Security for mobile and cloud frontiers in healthcare, Commun. ACM 58 (8) (2015) 21–23.
[50] D. Kundu, R.D. Gupta, Absolute continuous bivariate generalized exponential distribution, Adv. Stat. Anal. 95 (2) (2011) 169–185.
[51] H. Kunreuther, Mitigating disaster losses through insurance, J. Risk Uncertain. 12 (2–3) (1996) 171–187.
[52] A. Laszka, J. Grossklags, Should cyber-insurance providers invest in software security?, in: European Symposium on Research in Computer Security Springer, 2015, pp. 483–502.
[53] A. Laszka, S. Farhang, J. Grossklags, On the economics of ransomware, in: Proceedings of International Conference on Decision and Game Theory for Security, 2017, pp. 397–417.
[54] G.N. Lauer, Acceptance probabilities for sampling plans where the proportion defective has a Beta distribution, J. Qual. Technol. 10 (2) (1978) 52–55.
[55] E.R. Leukfeldt, E.R. Kleemans, W.P. Stol, Cybercriminal networks, social ties and online forums: social ties versus digital ties within phishing and malware networks, Br. J. Criminol. 57 (3) (2016) 704–722.
[56] R. Lippmann, J.W. Haines, D.J. Fried, J. Korba, K. Das, The 1999 DARPA off-line intrusion detection evaluation, Comput. Netw. 34 (4) (2000) 579–595.
[57] P. Majuca, W. Yurcik, J.P. Kesan, The evolution of cyber insurance. http://arxiv.org/ftp/cs/papers/0601/0601020.pdf, 2006.
[58] A. Mazzoccoli, M. Naldi, Robustness of optimal investment decisions in mixed insurance/investment cyber risk management, Risk Anal. 40 (3) (2020) 550–564.
[60] J. McHugh, Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln laboratory, ACM Trans. Inform. System Security (TISSEC) 3 (4) (2000) 262–294.
[61] R.M. Mohammad, F. Thabtah, L. McCluskey, Predicting phishing websites based on self-structuring neural network, Neural Comput. & Applic. 25 (2) (2014) 443–458.
[62] T. Moore, R. Clayton, Examining the impact of website take-down on phishing, in: Proceedings of the Anti-phishing Working Groups 2nd Annual eCrime Researcher Summit, 2007, pp. 1–s.
[63] A. Mukhopadhyay, S. Chatterjee, D. Saha, A. Mahanti, S.K. Sadhukhan, Cyber-risk decision models: to insure IT or not? Decis. Support. Syst. 56 (2013) 11–26.
[64] A. Mukhopadhyay, S. Chatterjee, K.K. Bagchi, P.J. Kirs, G.K. Shukla, Cyber risk assessment and mitigation (CRAM) framework using logit and probit models for cyber insurance, Inf. Syst. Front. (2019) 1–22.
[66] S. Nadarajah, A bivariate distribution with gamma and beta marginals with application to drought data, J. Appl. Stat. 36 (3) (2009) 277–301.
[67] H. Öğüt, S. Raghunathan, N. Menon, S. Raghunathan, Cyber insurance and IT security investment: impact of interdependent risk, in: Proceedings of WEIS, 2005.
[68] H. Öğüt, S. Raghunathan, N. Menon, Cyber security risk management: public policy implications of correlated risk, imperfect ability to prove loss, and observability of self-protection, Risk Anal. 31 (3) (2011) 497–512.
[70] G. Phillips-Wren, M. Daly, F. Burstein, Reconciling business intelligence, analytics and decision support systems: more data, deeper insight, Decis. Support. Syst. 146 (2021) 113560.
[71] A. Rai, Explainable AI: from black box to glass box, J. Acad. Mark. Sci. 48 (2020) 137–141.
[72] H. Rathore, A. Samavedhi, S.K. Sahay, M. Sewak, Towards Adversarially superior malware detection models: an adversary aware proactive approach using adversarial attacks and defenses, Inf. Syst. Front. (2022) 1–21.
[73] O.K. Sahingoz, E. Buber, O. Demir, B. Diri, Machine learning based phishing detection from URLs, Expert Syst. Appl. 117 (2019) 345–357.
[74] S. Samtani, Hacker Web Forum Collection: Hackhound Forum Dataset. http://www.azsecure-data.org/, 2016.
[75] S. Samtani, H. Chen, Using social network analysis to identify key hackers for keylogging tools in hacker forums, in: 2016 IEEE conference on intelligence and security informatics (ISI), 2016, September, pp. 319–321.
[76] S. Samtani, R. Chinn, H. Chen, J.F. Nunamaker Jr., Exploring emerging hacker assets and key hackers for proactive cyber threat intelligence, J. Manag. Inf. Syst. 34 (4) (2017) 1023–1053.
[77] SANS Institute, 2021 SANS Cyber Threat Intelligence (CTI) Survey. https://www.sans.org/reading-room/whitepapers/analyst/2021-cyber-threat-intelligence-cti-survey-40080, 2021.
[78] K. Sharma, A. Mukhopadhyay, Cyber-risk management framework for online gaming firms: an artificial neural network approach, Inf. Syst. Front. (2022) 1–22.
[79] N. Shetty, G. Schwartz, M. Felegyhazi, J. Walrand, Competitive cyber insurance and internet security, in: Proceedings of the Workshop on Economics of Information Security, London, England, 2009.

[81] D. Shin, The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI, Int. J. Human-Comput. Stud. 146 (2021) 102551.
[82] C.A. Siegel, T.R. Sagalow, P. Serritella, Cyber-risk management: technical and insurance controls for enterprise-level security, Security Manag. Pract. (2002) 33–49.
[83] H.R. Skeoch, Expanding the Gordon-Loeb model to cyber-insurance, Comput. Secur. 112 (2022) 102533.
[84] M. Sklar, Fonctions de repartition an dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.
[85] S. Smadi, N. Aslam, L. Zhang, Detection of online phishing email using dynamic evolving neural network based on reinforcement learning, Decis. Support. Syst. 107 (2018) 88–102.
[86] B. Srinidhi, J. Yan, G.K. Tayi, Allocation of resources to cyber-security: the effect of misalignment of interest between managers and investors, Decis. Support. Syst. 75 (2015) 49–62.
[88] F. Tajaddodianfar, J.W. Stokes, A. Gururajan, Texception: a character/word-level deep learning model for phishing URL detection, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, May, pp. 2857–2861.
[89] C.L. Tan, K.L. Chiew, K. Wong, PhishWHO: phishing webpage detection via identity keywords extraction and target domain name finder, Decis. Support. Syst. 88 (2016) 18–27.
[90] Q. Tang, G. Tsitsiashvili, Precise estimates for the ruin probability in finite horizon in a discrete-time model with heavy-tailed insurance and financial risks, Stoch. Process. Appl. 108 (2) (2003) 299–325.
[91] M. Tsikerdekis, S. Zeadally, Online deception in social media, Commun. ACM 57 (9) (2014) 72–80.
[92] G. Varshney, M. Misra, P.K. Atrey, A phish detector using lightweight search features, Comput. Secur. 62 (2016) 213–228.
[94] A. Vishwanath, T. Herath, R. Chen, J. Wang, H.R. Rao, Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model, Decis. Support. Syst. 51 (3) (2011) 576–586.
[98] N. Whelan, Sampling from Archimedean copulas, Quant. Finan. 4 (3) (2004) 339.
[99] R.L. Wolpert, Exponential Families, Duke University, 2000.
[100] G. Xiang, J. Hong, C.P. Rose, L. Cranor, Cantina+: a feature-rich machine learning framework for detecting phishing web sites, ACM Trans. Inform. System Security (TISSEC) 14 (2) (2011) 21.
[101] X. Yang, E.W. Frees, Z. Zhang, A generalized beta copula with applications in modeling multivariate long-tailed data, Insurance: Math. Econ. 49 (2) (2011) 265–284.
[103] D. Young, J. Lopez Jr., M. Rice, B. Ramsey, R. McTasney, A framework for incorporating insurance in critical infrastructure cyber risk strategies, Int. J. Crit. Infrastruct. Prot. 14 (2016) 43–57.
[104] Y. Zhang, J.I. Hong, L.F. Cranor, Cantina: a content-based approach to detecting phishing web sites, in: Proceedings of the 16th International Conference on World Wide Web, 2007, pp. 639–648.
[105] X. Zhang, J. Zhao, Y. LeCun, Character-level convolutional networks for text classification, Adv. Neural Inf. Proces. Syst. 28 (2015).
[106] X. Zhang, A. Tsang, W.T. Yue, M. Chau, The classification of hackers by knowledge exchange behaviors, Inf. Syst. Front. 17 (2015) 1239–1251.
[108] S. Jain, A. Mukhopadhyay, S. Jain, Can Cyber Risk of Health Care Firms be Insured? A Multinomial Logistic Regression Model, Journal of Organizational Computing and Electronic Commerce 0 (0) (2023) 1–29.
[109] M. Tripathi, A. Mukhopadhyay, Does privacy breach affect firm performance? An analysis incorporating event-induced changes and event clustering, Information & Management 59 (8) (2022) 103707.

Baidyanath Biswas is an Assistant Professor of Business Analytics at the Trinity Business School, Trinity College Dublin, Ireland. He received his Ph.D in Information Systems with a specialisation in cyber-risk management from the Indian Institute of Management Lucknow. His research has appeared in *Decision Support Systems, Journal of Business Research, Computers in Industrial Engineering,* and the *Journal of Enterprise Information Management*. Baidyanath is also associated with top peer-reviewed international conferences, namely, HICSS and ICIS. He has a rich industry experience of nine years working as a Mainframe and DB2 Database Analyst at Infosys and IBM. Currently, Baidyanath serves as the Associate Editor of Electronic Markets (Springer) Global Business Review (Sage) journals.

Arunabha Mukhopadhyay is a Professor of Information Technology & Systems Area at Indian Institute of Management Lucknow (IIM Lucknow). He received his Ph.D. and Post Graduate Diploma in Business Management (PGDBM) from the Indian Institute of Management Calcutta (IIM Calcutta), in the area of Management Information Systems. He was awarded the Infosys scholarship during his Ph.D. He has published in various referred journals and conferences including *Decision Support Systems, Information & Managment, Information Systems Frontier, Journal of Organizational Computing and Electronic Commerce, Journal of Global Information Technology Management (JGITM), JIPS, International Journal of Information Systems and Change Management (IJISCM), Decision, IIMB Review, Hawaii International Conference on System Sciences (HICSS), Americas Conference on Information Systems (AMCIS), Pre-International Conference On Information Systems (ICIS) workshops,* etc. He is a Senior Editor of Journal of Organizational Computing and Electronic Commerce.

Ajay Kumar is an Associate Professor at the EMLYON Business School in France. His research and teaching interests are in data and text mining, decision support systems, business intelligence and enterprise modelling. He has been a Postdoctoral Fellow at the Massachusetts Institute of Technology and Harvard University. He has published several research papers in reputed journals, including *Decision Support Systems, Harvard Business Review, European Journal of Operational Research, European Journal of Information Systems, Production and Operations Management, British Journal of Management*, *Journal of Business Research*, *Technological Forecasting & Social Change*, etc.

Dursun Delen is the holder of Spears and Patterson Endowed Chairs in Business Analytics, Director of Research for the Center for Health Systems Innovation, and Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University. He authored/co-authored 100+ journal and 40+ peer-reviewed conference proceeding articles. His research has appeared in major journals including *Decision Support Systems, Decision Sciences, Production and Operations Management, Communications of the ACM, Computers and Operations Research, Computers in Industry, Artificial Intelligence in Medicine, Expert Systems*, among others. He has recently published ten books/textbooks in the broad area of Business Intelligence and Business Analytics. He is currently serving as the editor-in-chief, senior editor, associate editor, and editorial board member of more than a dozen academic journals.