Machine Learning Approach for Predicting Cyberattacks Using a Bayesian Network Model

Sulaiman Al Amro [1,*] and Mafawez Thewiban Alharbi [2]

[1] Department of Computer Science, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia
[2] Unit of Scientific Research, Applied College, Qassim University, Buraydah 51452, Saudi Arabia
Email: samro@qu.edu.sa (S.A.); maft.alharbi@qu.edu.sa (M.T.A.)
*Corresponding author

***Abstract*—The increasing complexity and sophistication of cyber threats requires advanced, proactive strategies for information security. Traditional defensive measures are no longer sufficient, prompting organizations to adopt Artificial Intelligence (AI) and machine learning techniques for threat detection and risk management. In this context, Bayesian Networks (BNs) have emerged as a powerful tool to model probabilistic relationships among key behavioral and environmental factors influencing cybersecurity. This study presents a Bayesian Network-based approach to assess individual cybersecurity awareness and predict potential vulnerabilities based on user behavior. Using survey data collected from 102 participants, the model evaluates user responses to risk-related behaviors—such as the use of multifactor authentication, application sources, update practices, and email handling—to generate a predictive risk score. A weighted scoring system that incorporates interaction terms was employed to account for compound risk scenarios. The model achieved a strong performance with an overall accuracy of 92.6%, recall of 88.9%, precision of 57.1%, and an F1-Score of 0.696. These results demonstrate the proposed model's capability to effectively identify users with high-risk behavior patterns, thereby providing valuable support for organizational cybersecurity awareness programs and decision-making processes.**

***Keywords*—cyberattack, Bayesian network, probabilistic prediction, machine learning**

I. Introduction

In the contemporary computing landscape, information security is of paramount importance. Addressing the continual evolution of threats requires effective measures, which make cybersecurity and risk management indispensable for tasks reliant on data or information. Cybersecurity encompasses a set of procedures, human actions, and technological solutions intended to safeguard electronic information resources. The escalating pace at which cyber attackers can breach defense mechanisms is a cause for concern, increasing concerns about the security of valuable digital assets.

The role of system and network technologies in information technology across diverse applications is critical, and security is of paramount importance for both networks and applications. However, despite the crucial need for network security, a notable deficiency exists in straightforwardly implementable security measures. It is imperative to ensure maximum security during network development [1]. However, ensuring the security of network information using conventional information security technology has proven challenging. Therefore, this study focused on exploring the application of machine learning feature extraction methods in a situational awareness system. To this end, we introduce a feature selection method rooted in a Bayesian network to extract the situational features.

Machine learning research focuses on the automatic acquisition of the ability to discern intricate patterns and make informed decisions using data analysis. A key challenge arises from the complexity of articulating all possible behaviors, where most cyberattacks occur as a result of user behavior, such as disobeying the organization's security rules.

This paper aims to implement an approach for predicting cyberattacks using the Bayesian network technique while simultaneously demonstrating the benefits of the proposed model in this study. The proposition in this study assists cybersecurity managers in organizations to forecast the security score levels, which helps them prevent attacks from occurring. It possesses the capability to pinpoint the most probable sources of vulnerability. The proposed approach in this paper is displayed in a clear and readable graph. The model is capable of managing incomplete data when it is not possible to obtain all the input information.

In recent years, there has been a significant increase in attacks and vulnerabilities in organizations, and a considerable number of individuals lack sufficient knowledge about information security. Researchers and developers are diligently working to develop software to prevent such vulnerabilities. However, their efforts must be augmented by adopting new techniques, such as artificial intelligence, which has the capability to predict emerging vulnerabilities based on historical data. In this research, a Bayesian network model was employed to predict cyberattacks. The proposed model can be used to assess the security knowledge and awareness of employees in organizations.

Manuscript received February 9, 2025; revised March 24, 2025; accepted June 17, 2025; published September 12, 2025.

The success of this research will be assessed based on the following criteria: handling incomplete data, working with time-series data, presenting results through a graphical interface, and using reliable statistical data. In this context, reliable statistical data refers to datasets that are validated for accuracy, consistency, and completeness, ensuring that they can be trusted for analysis and inference. This definition is consistent with the description provided by IBM, which defines data reliability as "the completeness and accuracy of data as a measure of how well it can be counted on to be consistent and free from errors across time and sources" [2]. In this study, data reliability was ensured through a survey instrument that was carefully designed for real-world users. The responses were evaluated for completeness and internal consistency, and basic validation rules were applied to detect anomalies or misleading entries. These criteria were established to evaluate the research's effectiveness.

The remainder of the research is structured as follows: Section II provides a literature review, related work is presented in Section III, Section IV presents the problem statement. In Section V, the research methodology is provided. The implementation is shown in Section VI and the comparative study in Section VII.

II. Literature Review

Information Security (InfoSec) is a multifaceted discipline concerned with protecting information from a range of risks, including unauthorized access, alteration, disclosure, and destruction. As defined in [3–6], InfoSec encompasses the policies, procedures, and technological measures used to mitigate these threats. It is a critical component of organizational risk management that is aimed at preserving the confidentiality, integrity, and availability of information. The increasing digitization of services, combined with growing cyber threats, has made InfoSec a central concern across sectors. Despite significant technological advancements, organizations continue to face persistent and evolving security challenges [7]. The success of an organization's security posture is influenced not only by its technological capabilities but also by its governance structures; thus, it is essential to approach information security from a strategic and managerial perspective [8].

Despite numerous technological advancements, information security concerns continue to pose significant challenges for most organizations [7]. It is important to note that the efficacy of technological solutions informs organizations' information security policies and strategies. Hence, when examining this issue, it is important to employ a managerial lens [8].

Research findings confirm that information security should be prioritized at board level, beyond the sole remit of technical or information officers [9]. This underscores the importance of recognizing information security as a strategic concern that should be embraced across all facets of an organization, rather than being confined to specific technical or information-focused roles.

According to Surianarayanan *et al*. [10], Artificial Intelligence represents the convergence of computer science and cognitive processes. In straightforward terms, intelligence is the computational facet of the global capacity to accomplish objectives in the world. It encompasses the ability to think, imagine, create, memorize, understand, recognize patterns, make choices, adapt to change, and learn from experience. AI is being harnessed to enable computers to emulate human behavior while achieving tasks that are significantly faster than human ability.

Learning is the process of gaining knowledge or a skill within a specific domain, primarily within the context of human beings [10]. In the realm of psychology, diverse and generalized definitions of learning have been proposed. Many of these interpretations conceptualize learning as an alteration in the behavior of an individual, influenced by a particular situation or as a consequence of repeated experiences within that specific context [11].

Learning involves acquiring new knowledge and/or improving an individual's skills. Acquiring new knowledge encompasses various processes, including grasping significant concepts, understanding their meanings, and recognizing their relationships to one another and to the relevant area of study. Skill enhancement, from a biological standpoint, can be understood as the reinforcement of neural connections to perform a specific function [10].

According to Refs. [12–16], AI is revolutionizing cybersecurity in terms of several key developments:

- **AI-Driven Threat Detection and Response**: AI systems are increasingly relied upon to detect and respond to threats in real-time using advanced algorithms to scan for irregularities in digital environments. By employing behavioral analytics, AI can learn normal behavioral patterns and detect deviations that may signal a security threat.
- **Adapting to Polymorphic and Metamorphic Malware**: AI and machine learning techniques are essential to countering advanced malware types that constantly change their code to evade detection. These technologies allow for dynamic defenses that evolve alongside threats.
- **Enhanced Zero-Trust Security**: AI plays a vital role in enhancing zero-trust models, which continuously verify each user and device. This approach minimizes vulnerabilities using AI-driven anomaly detection to adapt to new risks and provide constant protection.
- **Preparing for Quantum Computing**: The rise of quantum computing poses significant challenges for traditional encryption methods. AI is likely to contribute to the development and implementation of quantum-resistant encryption algorithms to secure data against the immense power of quantum machines.
- **Combating Ransomware and Cybercrime-as-a-Service**: AI helps defend against increasingly sophisticated ransomware attacks, particularly with the growing availability of hacking tools through Cybercrime-as-a-Service (CaaS). AI-

powered solutions improve threat detection, backup, and recovery strategies.

These trends highlight the increasing importance of AI in addressing the evolving complexity of cyber threats.

George and Renjith [17] offered an examination of BNs and how they are utilized to evaluate safety and security risks in process industries. The review encompasses an exploration of the typical approaches for constructing BNs, the software tools that facilitate their development, the techniques for validating them, sensitivity analysis methodologies, and the diverse applications of BNs in the context of safety and security assessments.

In the scientific context, the study of algorithms and statistical models employed by computer systems to execute specific tasks is termed "machine learning", and this is considered a form of AI. Machine learning algorithms construct mathematical models based on sample data, which are referred to as training data, enabling them to make predictions or decisions without explicitly programming for the task at hand [18].

A Bayesian Network is used to calculate the probabilistic relationships among different variables of interest. It consists of various components, including factors and vertices, which together form a graph with no cyclic patterns. Each node in the graph is a random variable, and the directional edges between them reflect their relationships. Importantly, all variables in the BN are independent of variables that are not their descendants. When appropriately trained, a BN can serve as a highly accurate classifier capable of precise categorization. For a thorough exploration of BNs, a comprehensive study can be found in references [19–21].

BNs are frequently employed in intrusion detection tasks, often in conjunction with statistical techniques. This approach offers several advantages, as highlighted in [22]. These benefits include the capacity to represent relationships between variables and make predictions about events. Furthermore, the BN can seamlessly integrate prior knowledge and data, thereby facilitating swifter learning and classification processes. In addition, the network is easily updatable when new cases arise.

Chitrakar and Petrović [23] use a machine learning algorithm called k-means clustering, which is a common clustering algorithm in cybersecurity analytics. This type of algorithm has the ability to divide security-related data into groups of similar entities, which, in turn, can support gaining important insights into the known and unknown attack patterns. The k-means ratio aids security analysts in focusing on the data specific to some clusters only for the analysis. This study aims to re-formulate the parallel version of Elkan's k-means using the Triangle Inequality (k-meansTI) algorithm. In addition, this paper also offers a speed comparison of parallel k-meansTI on Spark using the Spark ML k-means clustering algorithm.

A Support Vector Machine (SVM) algorithm study was used in [24] to detect cyberattacks on an AGC system, a single-class attack detection. The research paper uses an SVM-Recursive Feature Elimination (SVM-RFE) method to decrease the feature dimensionality of trial data facts and identify the ideal feature subset. In addition, the time difference is used to develop the data dimensionality and model the dynamic information of the system based on this technique.

Oyinloye *et al*. [25] highlighted broader issues in the artificial intelligence domain, specifically the vulnerability of neural networks to malicious attacks. An Artificial Neural Network (ANN) is used to address data categorization problems caused by unbalanced data. This research presents a revolutionary intrusion detection method that focuses on the ever-evolving nature of cybersecurity threats. The present research recommends employing a novel ANN model to improve the resilience of IDS compared to adversarial attacks, which are becoming increasingly common.

Souza *et al*. [26] proposed a fuzzy neural network model to detect anomalies during cyberattacks. The model combines neural networks with artificial neurons using the leaky rectified linear unit function and fuzzy logic based on the ANFIS technique. The network is trained using an extreme learning machine to keep it simple. The proposed model aims to not only detect anomalies but also generate a set of fuzzy rules to serve as an interpretable knowledge base on potential attacks. This study uses a dataset from the 1998 DARPA Intrusion Detection Evaluation Program, which simulates various attacks in a military network environment. However, the proposed model relies on the DARPA 1998 Intrusion Detection dataset, which may not fully represent modern cyberattack patterns. More recent datasets can provide a better basis for evaluating model performance. In addition, this study does not compare the fuzzy neural network approach to other popular anomaly detection techniques like support vector machines, isolation forests, and deep learning models. Comparing the accuracy and interpretability across different models provides a better sense of the relative strengths of the fuzzy neural network.

The aim of Xue *et al*. [27] was to develop a predictive model for assessing the safety status of Wireless Networks (WNs) using fuzzy logic. This paper attempts to address the limitations of existing methods that only analyze current network security statuses without predicting future trends. The proposed model aims to provide a real-time evaluation of network security, enabling administrators to perceive and understand the current status and potential future developments. By introducing fuzzy logic into the safety status prediction process, the authors aim to better handle the uncertainties and complexities associated with wireless network environments. Fuzzy logic allows for a more nuanced assessment of the safety status, thereby accommodating the inherent vagueness in network conditions. Overall, this paper aims to provide a robust framework for predicting and assessing the safety status of wireless networks, thereby improving the capabilities of network administrators in mitigating risks and effectively responding to potential threats. The paper identifies a few limitations in using fuzzy logic to predict wireless network security status. For example, it considers the complexity of fuzzy rule generation, subjectivity in membership function definition, computational overhead, adaptability to dynamic network changes, and validation and tuning.

III. Research Methodology

The methodology of this research starts with system architecture, followed by data gathering, next applying machine learning algorithms are applied, and the last section is the implementation. A more detailed explanation of the methodology, particularly the rationale behind the use of machine learning algorithms called "Bayesian Networks" and its applicability are provided for the following reasons:

- BN has the ability to implement time-series data; thus, previous and current data are used.
- BN should be used when there is reliable statistical data; as a result, this research gathers user data based on a reliable survey.
- BN can be used in many domains, such as the medical system and cybersecurity.
- One advantage of BN is its ability to handle incomplete data.
- Graphical interface (M)
- A BN is a directed graph model, namely

A. Model Justification

Although various machine learning algorithms, such as Artificial Neural Networks, Support Vector Machines, and Random Forests, are widely used in cybersecurity, they present notable limitations in the context of this study. ANNs and random forests are typically considered "black-box" models, making it difficult to interpret how input features influence predictions—a critical drawback when dealing with human behavior-based risk assessment [28, 29]. SVMs, on the other hand, require clean, high-dimensional datasets and lack probabilistic reasoning, which makes them less effective in uncertain environments common in behavioral data [30, 31]. In contrast, Bayesian Networks offer transparency, interpretability, and the ability to integrate domain knowledge and incomplete data [17, 32]. These features make them a more suitable choice for our behavior-driven, survey-based cybersecurity model.

The selection of the BN model in this research is based on its distinct advantages over other machine learning approaches, particularly in the cybersecurity context. Unlike black-box models such as ANNs or ensemble methods like Random Forests, BNs offer transparent reasoning and an interpretable structure, which are essential when evaluating human behavior and decision-making in security settings.

A key strength of BNs is their ability to handle uncertainty and incomplete data. This is particularly relevant in real-world cybersecurity applications, where user input data may be missing, ambiguous, or delayed. In contrast to models such as SVMs, which typically require clean and complete datasets, BNs can still generate valid probabilistic predictions under uncertainty.

Furthermore, Bayesian networks support the probabilistic modeling of relationships among multiple variables, making them ideal for capturing causal links between different behaviors and potential security risks. This capability allows for a context-aware prediction mechanism, where past and present user behaviors can be used to assess risk levels in a structured and explainable way.

Finally, BNs are well suited for integration with survey-based or expert-derived data because they can combine domain knowledge with empirical evidence through Conditional Probability Tables (CPTs). Collectively, these features make the BN model highly appropriate for predicting user vulnerability to cyberattacks based on behavioral indicators.

A BN can use mathematical language to extract relationships between variables in a clear system and can build combined probability models with certain variables, where each individual variable is represented by a node in a chart. The straight dependencies between variables are denoted by the directed edges among the corresponding nodes (Fig. 1). In most cases of BNs, the directed edges among variables play a natural and simple role as graphical representations of causal relationships [33].

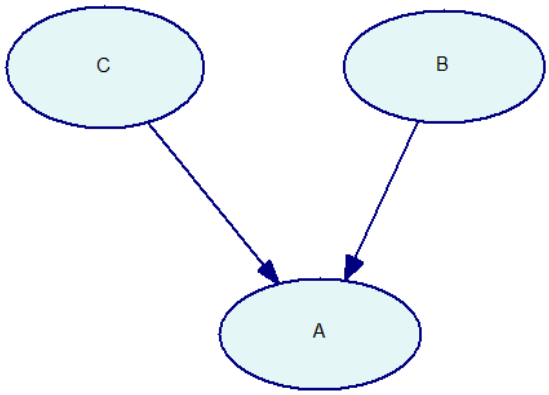

Fig. 1. Bayesian network structure.

B. System Architecture

The proposed system has the ability to be aware of user behavior. An overview of the system architecture is presented in Fig. 2, which consists of four elements: user interface, background, artificial intelligent system, and outside source.

User interface: The user-interface interacts between users and the proposed system; the main objective of this element is to allow the user to manage their profile and present the system predication. In addition, it can be implemented as a mobile application.

Background: This element contains user information, which is described as static information—for example, name, birth data, address, etc.

Outside source: The main task of this element is to provide and supply an artificial intelligent system element with outside data in the form of a virtual sensor.

Artificial Intelligence System (AIS): The main job of this element is to filter outside sources and includes four elements: data collection, predication, history, and results.

- Data collection: The main task of the element is to gather data from outside world which is "outside source."
- History: The purpose of the element is to store high-level data, which have already been delivered

to the user in order to prevent duplication of the data.
- Result: This element is responsible for creating messages for the end user.
- Prediction: The objective of this element is to make predications to the user based on the data delivered from the data collection and history.
- Data Processing: This component prepares the collected data by cleaning, transforming, and structuring the data to ensure accurate and reliable input for the prediction model.

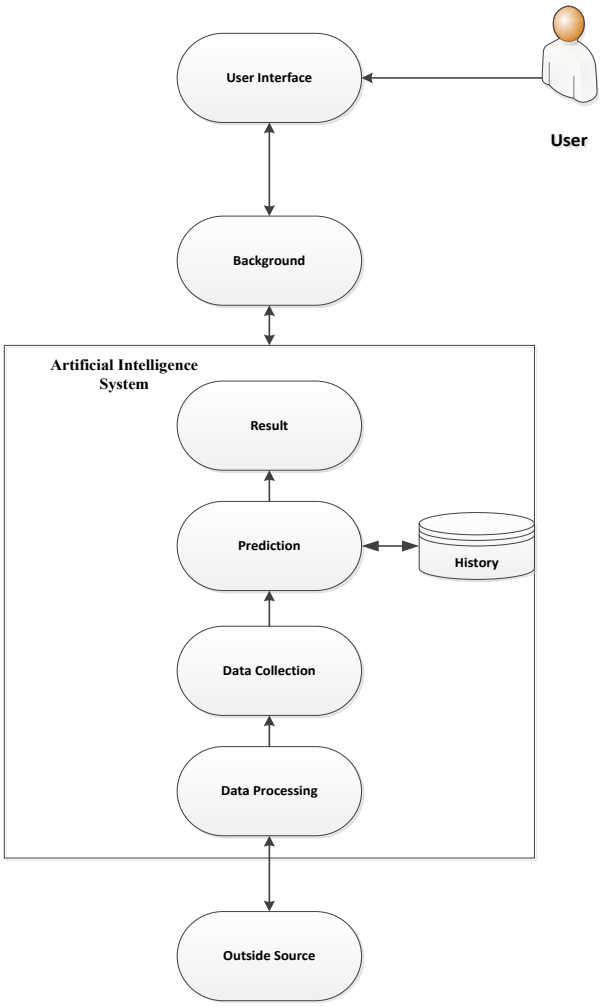

Fig. 2. Proposed system architecture.

All elements in the AIS were out of scope in this research except for the predication element. It would take too much time to implement the fully proposed system, as presented in Fig. 1. However, the most important and challenging element in this propodeal is the predication element (which is the toughest part of the AIS). Therefore, the focus of this research is the implementation of the predication element.

1) Mathematical model

This section defines the mathematical framework for processing user data, collecting external and historical information, reasoning with the data, and presenting the results. The model follows a sequential process.

a) User data interaction with the interface

The user data are mapped to the interface representation via the function F, which represents the mapping between the user inputs and the interface representation as follows:

$$UI = F(U). \tag{1}$$

where:
- U represents the raw user input.
- UI is the processed user-interface data.
- F denotes the function governing the user-interface interaction.

b) Processing data from user interface

Once the user data reach the interface, they undergo an initial transformation using the function g, which processes and structures the received data:

$$B = g(UI). \tag{2}$$

where:
- B is the processed data ready for further analysis.
- g is the transformation function applied to the interface data.

c) Data collection from external and historical sources

To enrich the model, data are gathered from external sources and historical records using the following extraction functions:

$$D = h(O). \tag{3}$$

$$H = h'(D). \tag{4}$$

where:
- D represents the collected data from external sources O,
- H denotes the historical data extracted based on D,
- h is the function responsible for extracting data from the sources, and
- h' is the function that updates historical data.

d) Reasoning and prediction mechanism

The proposed system integrates external and historical data to perform reasoning and generate predictions:

$$P = f(D, H). \tag{5}$$

where:
- P represents the predicted or inferred knowledge.
- f is the function modeling the reasoning process, where external data D and historical knowledge, H.

e) Result extraction and visualization

In this context, visualization plays a critical role in translating complex probabilistic outputs into accessible and actionable format for decision-makers. By representing prediction results graphically—such as risk scores, probability distributions, and status indicators—the system enables users, including nontechnical stakeholders, to quickly identify potential security threats. This facilitates timely interventions, informed policy decisions,

and more intuitive risk communication within an organization.

To present the final result to the user, transformation function K converts the processed data into a visual representation as follows:

$$R = K(P). \tag{6}$$

where:
- R is the final result displayed to the user.
- K is the function that converts the prediction P into a user-friendly format.

2) Conceptual framework

The proposed model follows a structured pipeline, which is divided into four key phases (Fig. 3):

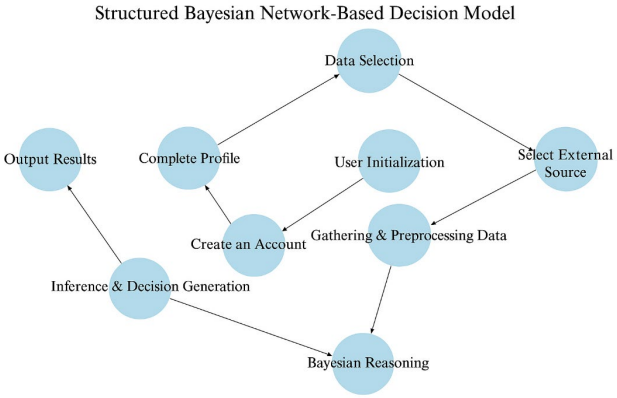

Fig. 3. Structured Bayesian Network-Based Decision Model.

a) Phase 1: User initialization

Account creation: Users register within the system to personalize data access and ensure secure authentication.

Profile completion: Users provide relevant contextual details (e.g., preferences, domain-specific attributes), allowing for adaptive reasoning.

b) Phase 2: Data selection and preprocessing

The Data Selection block enables the system to identify and select relevant input data sources based on the context of the application, such as cybersecurity logs, historical user behavior, and survey responses. This step ensures that only meaningful and domain-specific data are passed forward for processing, enhancing both efficiency and accuracy. It also allows the model to adapt to different environments by dynamically selecting appropriate data inputs.

External data source selection: This system enables users to select structured and unstructured data sources (e.g., clinical databases, financial reports, cybersecurity logs).

Data gathering and preprocessing: Raw data are collected, cleaned, and standardized to ensure consistent feature representation.

c) Phase 3: Bayesian inference and probabilistic reasoning

Bayesian network construction: This system formulates CPTs and defines the probabilistic dependencies between variables.

Inference and decision generation: Bayesian inference is applied using methods such as Markov Chain Monte Carlo or Expectation-Maximization to compute probability distributions.

d) Phase 4: Result output and decision support

Final decision generation: The system provides risk scores, classifications, or predictive insights, allowing users to make informed decisions.

Interpretability & visualization: Results are presented through graphical models, probability distributions, or decision recommendations.

To enhance the system's robustness, basic error-handling mechanisms were incorporated as precautionary measures during the data input and processing stages. This decision was motivated by the potential for incomplete or inconsistent responses observed in the survey data, which could affect the prediction accuracy if not properly handled. The model accounts for uncertainty using the probabilistic nature of Bayesian Networks, which can still function with incomplete or uncertain data. Additionally, input validation steps such as range checks, format verification, and consistency rules are incorporated to detect and mitigate erroneous or misleading survey responses. Future iterations of the system will include more advanced anomaly detection techniques to flag inconsistent or suspicious user input before it influences predictions.

C. Data Gathering

The second stage of the methodology involves conducting a survey. This study utilized real-world survey data collected from 102 users in Qassim, Saudi Arabia, making the findings practically relevant and grounded in actual behavioral insights. The survey involved distributing a Google Forms [34] link and assessing the factors influencing the security levels in the utilization of devices and applications among users via social networking sites.

An open questionnaire consisting of seven questions was designed to identify the most important factors affecting the level of security among users of devices and applications. First, the participants were asked whether they had used binary verification in social media applications. This was followed by a question about downloading programs from outside of official stores, as opening links that arrive via email without confirming the original sender is a concern in this research. Then, a question was raised regarding downloading software updates and periodically updating systems. The use of free public networks makes it possible to hack into any system. Thus, this question was included in the questionnaire. Finally, the participants were asked whether they had been previously hacked.

D. Applying the Bayesian Network Algorithm

This stage will explore more on applying machine learning algorithms, as there are many machine learning algorithms, such as k-means neural networks, random forest, fuzzy logic, probabilistic logic, and Bayesian networks. This research will use BN for the following

reasons. The proposed BN can handle time-series data, thereby allowing for the utilization of both past and present data. In addition, a BN should be employed when reliable statistical data are available; consequently, this research collects user data through a trustworthy survey. Furthermore, a BN can be applied in various fields, including healthcare and cybersecurity. One of the benefits of BN is its ability to handle incomplete data. In addition, a BN is a model represented as a directed graph.

A BN is a graphical model that represents probability relationships according to a number of variables. The name refers to Thomas Bayes (1702–1761), a founder of probability theory. The first BN was the Probabilistic Expert System in the 1990s, which was inspired by the work of Judea Pearl in 1988, another founder of BN [35].

The BN theory is based on Bayes' Theorem, as presented in Eq. (7):

$$P(h/e) = \frac{P(h/e)p(h)}{p(e)}. \quad (7)$$

where:
$P(h/e)$ denotes the probability of the hypothesis(h), given the input(e).
$p(e/h)$ is the probability of input(e) given the hypothesis(h).
$p(e)$ is the prior probability of the input(e).

There are many tools that can be used to implement BN, some of which are free, while others are commercial. The examples of these tools are MATLAB, Bayesian Lab, and GeNIe. For this research, the GeNIe tool was chosen because it is free to download, fast, and well documented. The GeNIe tool is an expansion environment for structuring graphic decision-notional models. GeNIe was created and developed by the Decision Systems Laboratory at the University of Pittsburgh. Its name originates from the name Graphical Network Interface, which was originally developed to be a major teaching and research tool in academic environments [33].

Many researchers around the world have used GeNIe. Because of its reliability and versatility, this tool has become commonplace and is the de facto standard in academia. Various military, government, and commercial users also use it.

GeNIe is designed to be suitable for Windows operating systems; however, it can also run on macOS (formerly OS X) and Linux. The models developed by GeNIe can be embedded in any application and run on any computing platform. To implement BN (GeNIe), four steps are required. Identify the factors, draw the BN model, parametrize the model, and apply the inference algorithm.

1) Step one: Identify the factors

The six most important factors affecting the level of security among device users have been identified, and they are: using multifactor authentication, using unofficial stores, opening email links without confirming the sender, software updates, system updates, and using free public Wi-Fi. Table I presents these factors and their states, and lists the factor abbreviations.

TABLE I. FACTORS WITH THEIR STATES

No	Factor	Factor abbreviation	State 1	State 1
1	Using Multifactor Authentication	MF	Yes	No
2	Use of Unofficial Stores	OS	Yes	No
3	Opening Email Links Without Confirming the Sender	WC	Yes	No
4	Software Updates	SU	Yes	No
5	System Updates	US	Yes	No
6	Free Public Wi-Fi	PW	Yes	No

2) Step two: Integrate the BN structure into the proposed model

In the previous section, the main factors were identified. The second step was to integrate the BN structure into our proposed model, which revealed the relationship between these factors. The result of fusing these factors is presented in Fig. 4.

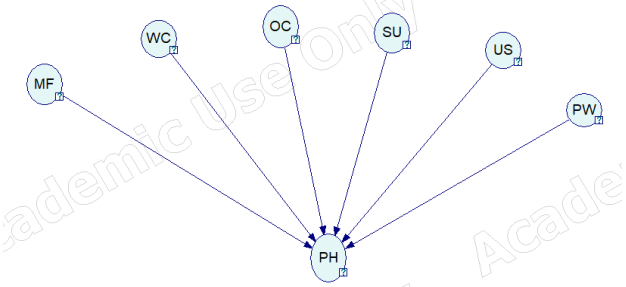

Fig. 4. Proposed model for cyberattack detection.

3) Step three: Parametrize the model

To feed the BN model with data, a survey was created and distributed to 102 users using Google Survey. The survey questions and possible answers are presented below:

1- Do you use multifactor verification in social media applications?
   - ✓ Yes
   - ✓ No
2- Do you install programs outside official stores?
   - ✓ Yes
   - ✓ No
3- Do you open links that arrive in an email without confirming that the original sender has been confirmed?
   - ✓ Yes
   - ✓ No
4- Do you frequently download software updates?
   - ✓ Yes
   - ✓ No
5- Do you periodically update your system?
   - ✓ Yes
   - ✓ No
6- Do you use free public Wi-Fi?
   - ✓ Yes
   - ✓ No

7- Have you been hacked before?
- ✓ Yes
- ✓ No

Once the survey had received sufficient responses, an Excel® file was used to save the survey data. To import the Excel file into the chosen tool (GeNIe), the file must be converted to a text file and saved in Notepad.

Fig. 5 shows the step of matching the survey results (data file) with the proposed BN model. GeNIe is able to classify the name of each column and match it with the data file.

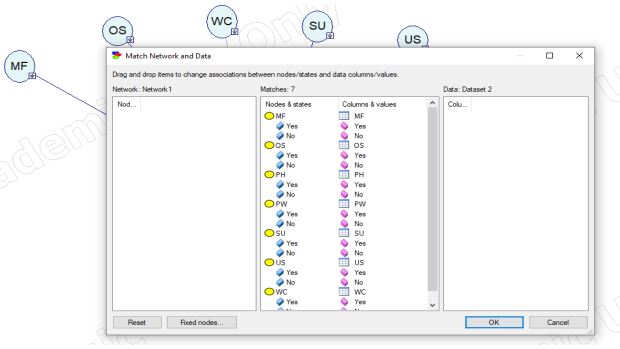

Fig. 5. The data survey was matched to the BN model.

The survey data were inserted into GeNIe, and the proposed BN is shown in Fig. 6 below (data open in GeNIe).

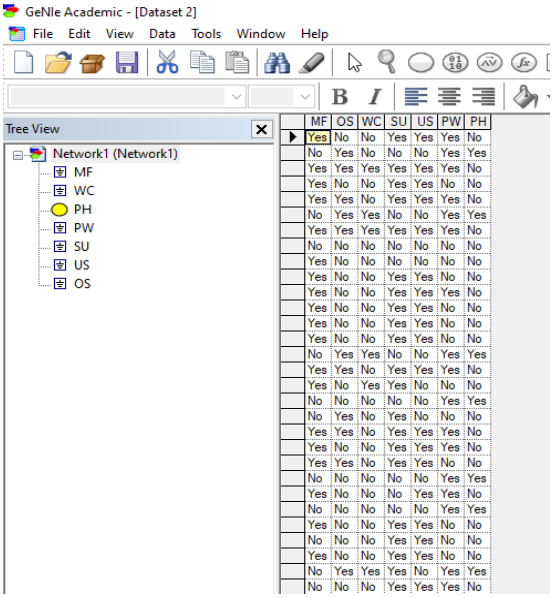

Fig. 6. Data are open in GeNIe.

Finally, parameter learning must be applied to the data. This means that feeding the BN model with the survey data completes the parameter learning, as shown in Fig. 7.

At this point, the BN model was parametrized by inserting the survey data into the proposed model; in other words, the model is learning. For example, the node MF has learned from these data, as shown in Fig. 8, where the number of users who use multifactor verification in social media applications is 0.62376238 and the others are 0.37623762.

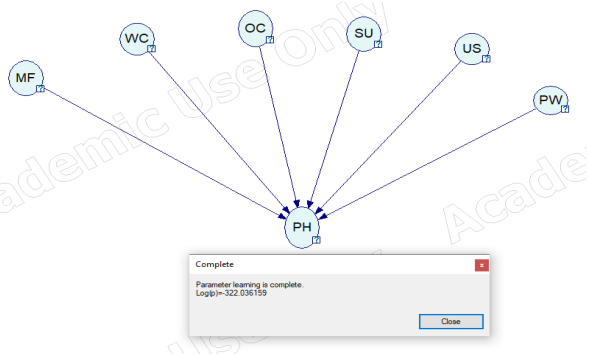

Fig. 7. Applying parameter learning.

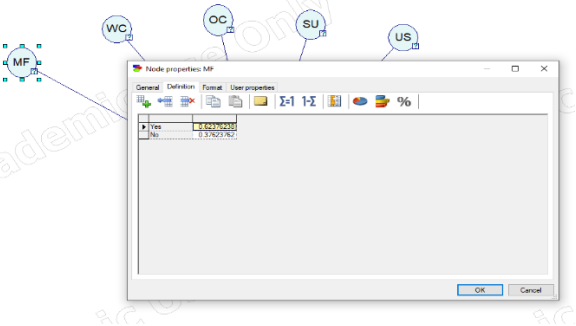

Fig. 8. Learning node.

4) Step four: Apply the inference algorithm

There are many algorithms for inference in BNs, and the selection of algorithms is based on the network structure. The inferences algorithm is already being implemented in BN software (e.g., GeNIe) and consists of two types: exact and approximate algorithms. The exact inference involves clustering and polytopes. Thus, the BN mode proposed for this research will use polytope because the model is a directed acyclic graph, and a single node provides a unique path to every other node.

IV. Implementation

In the proposed model, inference entails determining the conditional probability of a selected variable given that other variables are instantiated to particular values. The evidence for each factor must be identified in order to predict a user attack. To demonstrate the model's predictive capabilities, this section presents two detailed use cases with varying user behavior scenarios. The model's inference engine calculates the likelihood of cyberattack risk, and the outcomes are supported by empirical survey data.

The model's prediction capabilities are closely aligned with real-world cyberattack vectors, such as phishing and malware propagation. For instance, behavioral traits such as opening suspicious email links or neglecting updates, are commonly exploited by phishing attacks and malware. By quantifying these behaviors and correlating them with known risks, the model provides actionable insights that directly map to prevalent threat scenarios in organizational environments. Integrating this model with existing intrusion detection systems can enhance early-warning capabilities against such attacks.

A. Case One: Very High Attack

In this case, the evidence accords with Table II as follows:

TABLE II. EVIDENCES—FIRST CASE

No.	Factor	State
1	Using multifactor authentication	No
2	Use of unofficial stores	Yes
3	Opening email links without confirming the sender	Yes
4	Software updates	No
5	System updates	No
5	Free public Wi-Fi	Yes

BN used previous data alongside current evidence to predict the future. The previous data were based on the survey results and current evidence for each factor as observed by the user. An example of evidence observation is presented in Fig. 9:

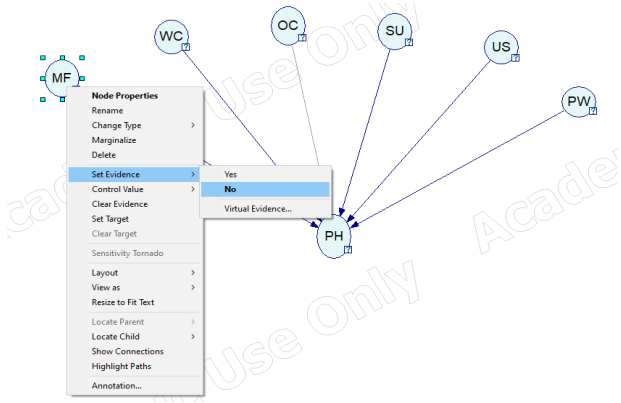

Fig. 9. Evidence observation.

Next, an algorithm must be selected. There are several BN algorithms; the polytope algorithm was chosen for this BN model because the model presented in this research has only one path (Fig. 10).

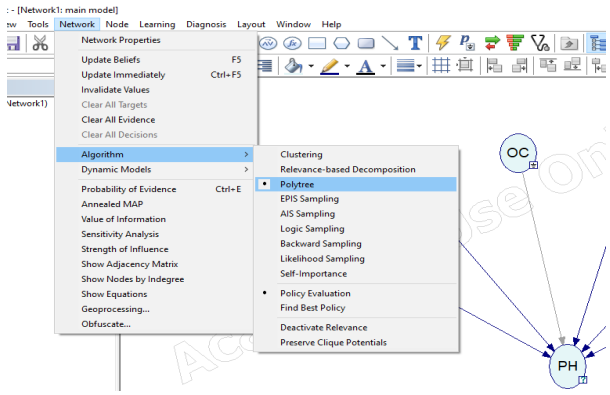

Fig. 10. Polytree algorithm selection.

Based on the evidence given in Table II and the previous data (survey), as well as the selected polytope algorithm, the BN model was executed, as shown in Fig. 11.

The outcome of the inference is shown in Fig. 12. The prediction of a user attack is very high at 0.93; this means that the input evidence has affected the result of this model and further action must be taken for this user, as an attacker could gain access to this user account and damage the PC and its documents. The following paragraphs will explain the reasons for the result obtained in case 1.

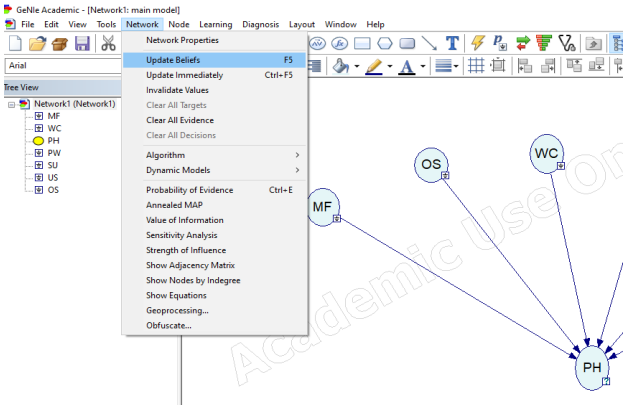

Fig. 11. Inference execution.

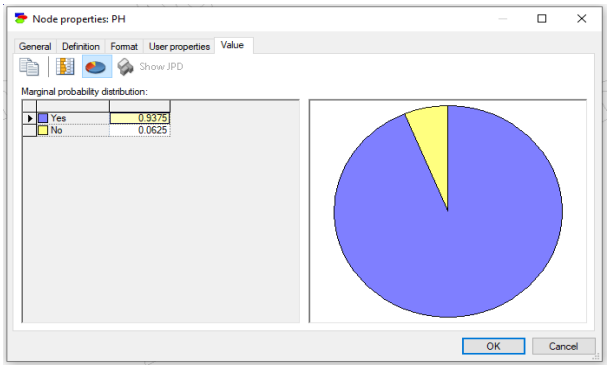

Fig. 12. Very high attack.

The user is using safe applications because the official store, which is normally filtered, is used. In addition, the user cannot open an email link without confirming the sender, which prevents the user from committing external attacks. On the other hand, the user did not use multifactor authentication, which made their username and password guessable. Therefore, the system in this case is vulnerable. Furthermore, as shown in Table II (factor four), users do not regularly update their software or applications. This leads to more vulnerability in the system. In the fifth factor, the user is not usually updating the system. This also results in increased system vulnerability. One risk associated with public Wi-Fi use is its potential lack of security, which makes it vulnerable to attacks. Hackers can exploit this vulnerability to steal personal information or covertly install malicious software on users' devices. In this case, the user confirms that he/she is using a public free Wi-Fi network, which can be harmful to the organization.

The result extracted from Table II indicates that the user is weak in security and may damage the security architecture in the organization. Therefore, these six factors will help the administration judge whether employees are aware of security threats and vulnerability or not. The help of BN is significant in determining the prediction of user attacks.

B. Case Two: Very Low Attack

In this case, the evidence accords with Table III as follows:

TABLE III. EVIDENCES—SECOND CASE

NO	Factor	State
1	Using multifactor authentication	Yes
2	Use of unofficial stores	No
3	Opening email links without confirming the sender	No
4	Software updates	Yes
5	System updates	Yes
6	Free public Wi-Fi	No

Once the evidence has been observed (Table III) and the algorithm has been selected, it is time to execute the model. The inference result for the second case is presented in Fig. 13. The prediction of a user attack was very low at 0.970, which means that the input evidence affected the result of this model. The user was safe; thus, no further action was required.

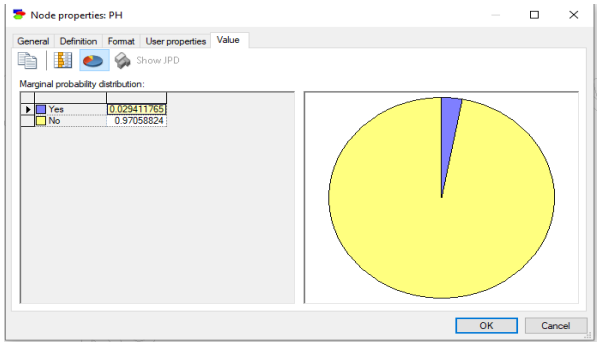

Fig. 13. Very low attack.

The user uses multifactor authentication in the first factor. This will add more security to the system due to the fact that multifactor authentication provides robust security measures that greatly enhance the protection against malicious actors and unauthorized access attempts. The user in the second factor of the second case uses only official stores, which enhances security. Moreover, users will not open email links without confirming the sender, which further strengthens the security of the organization.

As is well known, software updates introduce new and enhanced features while resolving existing issues such as bugs and crashes. Similarly, antivirus updates are continually improved by developers to protect users from emerging viruses and malware threats. In the second case, the user regularly uses software updates. In addition, users periodically attempt to update the system, which might enhance a system's performance, security, and stability. As discussed earlier, using public free Wi-Fi can be detrimental to an organization. However, the user in the second case did not use an open Wi-Fi connection, which also enhanced the security of the organization.

As explained previously, case 2 indicates that users are more aware of security threats and vulnerabilities than the one in case 1. Thus, the prediction of a user attack is low, which is explained next. Invariably, the help of the six factors and Bayesian networks will increase the ability of users to adjudicate whether or not a user has experience with security risks. As a result, this approach can be applied to both old and new employees to assist establishments with security matters.

C. Different Cases

More tests were performed for the three other cases to demonstrate different scenarios for the proposed mode. Fig. 14 presents middle attach. A high attack is shown in Fig. 15, and a low attack is presented in Fig. 16.

The inference outcomes for this case are presented in Fig. 14. The prediction for a user attack is at a medium level of 0.5, which indicates that the input evidence has influenced the model's outcome, which suggests that the user is not completely secure. Consequently, this user requires improvements in cybersecurity measures, although to a lesser extent than in case 1.

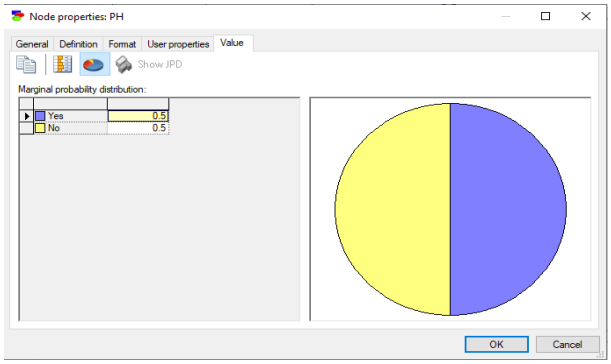

Fig. 14. Middle attack.

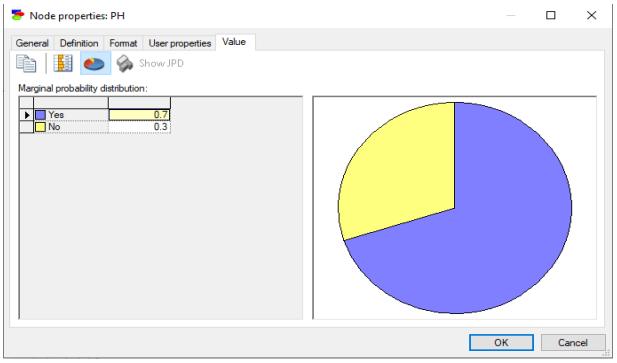

Fig. 15. High attack.

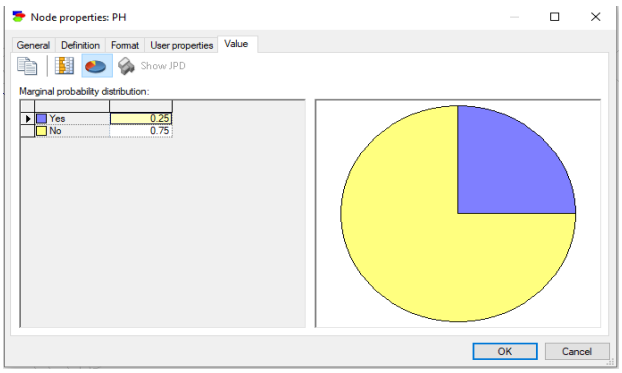

Fig. 16. Low attack.

The inference results are displayed in Fig. 15. The prediction for a user attack is high at 0.7, which indicates that the input evidence has influenced the model's outcome. Thus, further action is required for this user. This case differs from the situation in case 1 as this user needs less training.

The inference outcome for this case is presented in Fig. 16. The predicted likelihood of a user attack was low at 0.25, indicating that the input evidence influenced the model's outcome. This means that users are not necessarily safe, which suggests that further action is necessary. This case is better than the one in Fig. 15 but worse than the one in Fig. 13.

D. Model Performance with Weighted Scoring and Interaction Terms

A weighted scoring system was implemented to assess the cybersecurity risk based on user behavior. Different importance levels were assigned to six key behavioral factors: absence of multifactor authentication (3 points), use of unofficial stores (2 points), opening suspicious email links (3 points), neglecting software updates (2 points), neglecting system updates (2 points), and using free public Wi-Fi (1 point). In addition, three interaction features capturing compound risks were included: simultaneous absence of multifactor authentication and use of unofficial stores (MF_OS), opening suspicious email links combined with missing software updates (WC_SW), and neglecting system updates combined with use of public Wi-Fi (SY_PW), each weighted with 2 points.

The total risk score was calculated by summing the weighted factors and interaction terms. Users with a total score of 6 or higher were classified as high risk, whereas those below this threshold were classified as low risk.

Table IV presents the confusion matrix comparing the model's predictions to the actual user outcomes in the survey dataset.

TABLE IV. CONFUSION MATRIX FOR THE BAYESIAN NETWORK RISK PREDICTION MODEL

	Actual: Hacked (Positive)	**Actual: Not Hacked (Negative)**	**Total**
Predicted: High-risk	True Positive (TP) = 8	False Positive (FP) = 6	14
Predicted: Low Risk	False Negative (FN) = 1	True Negative (TN) = 87	88
Total	9	93	102

From this confusion matrix, the key performance metrics were calculated, and they are summarized in Table V.

TABLE V. EVALUATION METRICS FOR THE BAYESIAN NETWORK RISK PREDICTION MODEL

Metric	**Value**
Accuracy	0.926
Precision	0.571
Recall	0.889
F1-Score	0.696

The model demonstrated an overall accuracy of 0.926 (92.6%), and it correctly classified the vast majority of users based on their risk profiles. The recall of 0.889 (88.9%) indicates the model's effectiveness in identifying true positive cases of compromised users, which is essential for timely cybersecurity interventions. The precision of 0.571 (57.1%) indicates a balanced approach to minimizing false-positive identifications, and the F1-Score of 0.696 (69.6%) reflects a robust balance between precision and recall.

V. COMPARISONS STUDY

The evaluation of this study will be performed by comparing the results of this research with those of other related studies based on five chosen criteria, as shown in Table VI. Namely, dealing with incomplete data, time-series data, time, graphical interface, and reliable statistical data. In Refs. [23, 24], the system has some limitations due to the fact that they have three missing criteria. Both research lacks the ability to handle incomplete and time-series data. Furthermore, A graphical interface is not available for either of the studies. In Ref. [25], the research cannot deal with incomplete data. In addition, they did not provide a graphical interface. In Refs. [26, 27], no graphical interface is not supported in both researches.

However, the result of this research has proven that the four chosen criteria have been successfully achieved.

- When dealing with incomplete data, the proposed system can still provide a result even when the current data are missing. For example, in case 1 Table II, it can be done without knowing the evidence of the factor "Using free public Wi-Fi."
- BN is used to handle time-series data, although they are primarily designed to represent probabilistic relationships among variables, taking into account the previous and current context of the user.
- BN is well-equipped to handle reliable statistical data, thereby allowing for effective modeling of probabilistic relationships and inference. For this reason, this research conducted a survey to gather reliable data.
- The graphical interface was supported and played a crucial role in this research, as shown in Figs. 11–14.

Table VI presents a comparison between our proposed system to other machine learning-based cybersecurity models (e.g., k-means, SVM, ANN, Fuzzy logic). The evaluation criteria included handling incomplete data, support for time-series analysis, graphical interface availability, and use of reliable statistical data. The proposed model meets all four criteria, which highlights its robustness and applicability.

TABLE VI. COMPARISONS STUDY

No/ Factors	**Dealing with incomplete data**	**Time-series data**	**Graphical interface**	**Reliable statistical data**
22	x	x	x	√
23	x	x	x	√
24	x	√	x	√
25	√	√	x	√
26	√	√	x	√
The proposed system	√	√	√	√

The comparative data in Table VI were compiled based on an in-depth review of the respective methodologies

described in [22–26]. Each study was assessed for its ability to handle incomplete data, time-series processing, graphical representation of results, and reliance on validated statistical data. For example, the k-means-based model in [24] lacked mechanisms for dealing with incomplete data and does not incorporate a visual output interface. Similarly, the SVM approach in [25] does not provide a graphical interface or support for uncertain inputs. These gaps highlight the comparative advantages of the BN model proposed in this paper. The comprehensive structure of our model, coupled with interpretable outputs, enhances its practical utility in organizational settings.

VI. Conclusion

This research aimed to predict potential cyberattacks based on user knowledge about information security. The study focused on a machine learning algorithm that can offer probabilistic predictions, generating a score indicative of a person's level of information security awareness by gathering users' behaviors and importing the data to the machine learning technology. As a result, the model proposed in this paper can predict the level of security in organizations based on a BN. The proposed BN model can be integrated into existing Security Information and Event Management and Identity and Access Management systems. By embedding this model as a plug-in, organizations can leverage real-time awareness scoring to prioritize alerts, trigger additional authentication steps, or block access in high-risk scenarios. However, this study is not without limitations. Most notably, the dataset used for training and evaluation was limited to 102 participants from a single geographical region (Qassim, Saudi Arabia). This relatively small and localized sample may have affected the generalizability of the findings across other populations or organizational contexts. Additionally, the reliance on self-reported survey data introduces the potential for response bias, which could impact model accuracy. These limitations should be considered when interpreting the results and assessing the applicability of the model to broader environments.

Future work building on this research could include the following:

- Add more user behaviors and collect more data to feed the AI algorithms
- Evaluate other different machine learning algorithms, such as Neural Networks K-means and Random Forest
- Explore the rest of the Artificial Intelligence System, such as history, data collection, and results.

List of Abbreviations

Abbreviations	Explanation
BN	Bayesian Network
AI	Artificial Intelligence
CaaS	Cybercrime-as-a-Service
SVM	Support Vector Machines
ANN	Artificial Neural Network
WN	Wireless Network
KSA	Kingdom of Saudi Arabia

Declarations

Ethical considerations and privacy protection: The system design prioritizes ethical responsibility by ensuring that all user data are anonymized before processing. The Bayesian model does not track individual identities but rather assesses general awareness levels based on input responses. Data collection adhered to informed consent procedures, and no personally identifiable information was stored. Future deployment scenarios may adopt federated learning or edge-computing models may be adopted to further enhance data privacy.

Availability of supporting data: The authors confirm that data supporting the findings of this study are available within the article and its supplementary material. Raw data supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Sulaiman Al Amro and Mafawez Thewiban Alharbi equally contributed to all aspects of this work, including the conception and design of the study, data analysis and interpretation, manuscript drafting and revision. Both authors have read and approved the final version of the manuscript.

Funding

This research is funded by the Deanship of Graduate Studies and Scientific Research 934 at Qassim University (QU-APC-2025).

Acknowledgment

The researchers would like to thank the Deanship of Graduate Studies and Scientific Research 934 at Qassim University for financial support.

References

[1] A. Yeshmuratova, "Technological methods of ensuring information security in technical systems," *Eurasian Journal of Academic Research*, vol. 2, no. 4, pp. 188–192, 2023.

[2] IBM. What is data reliability? IBM Topics. [Online]. Available: https://www.ibm.com/topics/data-reliability

[3] M. Curry, "ACM SIGMIS database: The DATABASE for advances in information systems," *ACM SIGMIS Database*, vol. 49, pp. 49–66, 2018. https://doi.org/10.1145/3210530.3210535

[4] C. Joshi and U. K. Singh, "Information security risks management framework—A step towards mitigating security risks in university network," *Journal of Information Security and Applications*, vol. 35, pp. 128–137, 2017. https://doi.org/10.1016/j.jisa.2017.06.006

[5] M. Fletcher. (2016). An introduction to information risk. The National Archives Blog. [Online]. Available: https://blog.nationalarchives.gov.uk/introduction-information-risk/

[6] SANS Institute. Cyber Security Resources. [Online]. Available: https://www.sans.org/security-resources/

[7] K. Grant, D. Edgar, A. Sukumar, and M. Meyer, "Risky business: Perceptions of e-business risk by UK Small and Medium sized Enterprises (SMEs)," *International Journal of Information Management*, vol. 34, no. 2, pp. 99–122, 2014. https://doi.org/10.1016/j.ijinfomgt.2013.11.001

[8] Z. A. Soomro, M. H. Shah, and J. Ahmed, "Information security management needs more holistic approach: A literature review," *International Journal of Information Management*, vol. 36, no. 2, pp. 215–225, 2016. https://doi.org/10.1016/j.ijinfomgt.2015.11.009

[9] A. G. Young. Fighting to close the gap: Ernst & Young's 2012 global information security survey. *Scribd*. [Online]. Available: https://www.scribd.com/document/138897526/Ernst-Young-2012-Global-Information-Security-Survey

[10] C. Surianarayanan, J. J. Lawrence, P. R. Chelliah, E. Prakash, and C. Hewage, "Convergence of artificial intelligence and neuroscience towards the diagnosis of neurological disorders-a scoping review," *Sensors*, vol. 23, no. 6, 3062, 2023. https://doi.org/10.3390/s23063062

[11] K. C. A. Khanzode and R. D. Sarode, "Advantages and disadvantages of artificial intelligence and machine learning: A literature review," *International Journal of Library & Information Science (IJLIS)*, vol. 9, no. 1, pp. 30–36, 2020. https://doi.org/10.17605/OSF.IO/GV5T4

[12] A. Yaseen, "AI-driven threat detection and response: A paradigm shift in cybersecurity," *International Journal of Information and Cybersecurity*, vol. 7, no. 12, pp. 25–43, 2023.

[13] K. Brezinski and K. Ferens, "Metamorphic malware and obfuscation: A survey of techniques, variants, and generation kits," *Security and Communication Networks*, vol. 2023, 8227751, 2023. https://doi.org/10.1155/2023/8227751

[14] C. Daah, A. Qureshi, I. Awan, and S. Konur, "Enhancing zero trust models in the financial industry through blockchain integration: A proposed framework," *Electronics*, vol. 13, no. 5, 865, 2024. https://doi.org/10.3390/electronics13050865

[15] S. Singh and D. Kumar, "Enhancing cyber security using quantum computing and artificial intelligence: A review," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 4–11, 2024. https://doi.org/10.48175/ijarsct-18902

[16] A. Mathew, "Cybercrime-as-a-service & AI-enabled threats," *International Journal of Computer Science and Mobile Computing*, vol. 12, no. 1, pp. 28–31, 2023. https://doi.org/10.47760/ijcsmc.2022.v12i01.004

[17] A. George and V. R. Renjith, "Evolution of safety and security risk assessment methodologies towards the use of Bayesian networks in process industries," *Process Safety and Environmental Protection*, vol. 149, pp. 758–775, 2021. https://doi.org/10.1016/j.psep.2021.03.031

[18] P. Fieguth, "An introduction to pattern recognition and machine learning," in *An Introduction to Pattern Recognition and Machine Learning*, pp. 5–28, 2022. https://doi.org/10.1007/978-3-030-95995-1_2

[19] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral Earth observation data: A review," *Remote Sensing*, vol. 12, no. 15, 2495, 2020. https://doi.org/10.3390/rs12152495

[20] S. Gil-Begue, C. Bielza, and P. Larrañaga, "Multi-dimensional Bayesian network classifiers: A survey," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 519–559, 2020. https://doi.org/10.1007/s10462-020-09858-x

[21] K. Kitson, A. C. Constantinou, Z. Guo, Y. Liu, and K. Chobtham, "A survey of Bayesian network structure learning," *Artificial Intelligence Review*, 2023. https://doi.org/10.1007/s10462-022-10351-w

[22] M. Al Lail, A. Garcia, and S. Olivo, "Machine learning for network intrusion detection—A comparative study," *Future Internet*, vol. 15, no. 7, 243, 2023. https://doi.org/10.3390/fi15070243

[23] A. S. Chitrakar and S. Petrović, "Efficient k-means using triangle inequality on Spark for cyber security analytics," in *Proc. ACM Int. Workshop Secur. Priv. Analytics*, 2019, pp. 37–45. https://doi.org/10.1145/3309182.3309187

[24] C. Hong *et al.*, "A single-class attack detection algorithm for smart grid AGC system based on improved support vector machines algorithm," in *Proc. 2024 3rd Int. Conf. Cyber Secur., Artif. Intell. Digit. Econ.*, 2024, pp. 44–50. https://doi.org/10.1145/3672919.3672928

[25] T. S. Oyinloye, M. O. Arowolo, and R. Prasad, "Enhancing cyber threat detection with an improved artificial neural network model," *Data Science and Management*, 2024. https://doi.org/10.1016/j.dsm.2024.05.002

[26] P. V. C. Souza, A. J. Guimarães, T. S. Rezende, V. J. Silva Araujo, and V. S. Araujo, "Detection of anomalies in large-scale cyberattacks using fuzzy neural networks," *AI*, vol. 1, no. 1, pp. 92–116, 2020. https://doi.org/10.3390/ai1010005

[27] X. Xue, Y. Zheng, and C. Lu, "Wireless network safety status prediction based on fuzzy logic," *J. Cyber Secur. Mob*., 2023.

[28] Z. C. Lipton, "The mythos of model interpretability," *Commun. ACM*, vol. 61, no. 10, pp. 36–43, 2018. https://doi.org/10.1145/3233231

[29] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell*., vol. 1, no. 5, pp. 206–215, 2019. https://doi.org/10.1038/s42256-019-0048-x

[30] A. Ben-Hur and J. Weston, "A user's guide to support vector machines," in *Data Mining Techniques for the Life Sciences*, 2010, pp. 223–239.

[31] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv*., vol. 41, no. 3, pp. 1–58, 2009. https://doi.org/10.1145/1541880.1541882

[32] D. Heckerman, "A tutorial on learning with Bayesian networks," *Learning in Graphical Models*, pp. 301–354, 1998.

[33] BayesFusion. GeNIe documentation–BayesFusion documentation. [Online]. Available: https://support.bayesfusion.com/docs

[34] Google Forms: Online form creator | Google Workspace. [Online]. Available: https://www.google.com/forms/about/

[35] M. Scutari and J.-B. Denis, *Bayesian Networks: With Examples in R*, CRC Press, 2021.