International Journal of Multiphysics
Volume 18, No. 3, 2024
ISSN: 1750-9548

Machine Learning Advancements for Surveillance: Cybersecurity and Behavior Risk Forecasting

[1]Arshad Hashmi, [2]Fahad Alotaibi

[1]Department of Information Systems, Faculty of Computing and Information Technology in Rabigh(FCITR), King Abdulaziz University, P.O. Box 344, Rabigh 21911, Saudi Arabia

[2]Department of Information Systems, Faculty of Computing and Information Technology (FCIT), King Abdulaziz University, Jeddah, Saudi Arabia

Corresponding Author- ahsyed@kau.edu.sa

Abstract: -The increasing integration of surveillance systems in various domains has amplified the importance of ensuring cybersecurity and predicting behavioral risk factors. This paper presents an innovative approach, the "Risk and Anomaly Detection via Deviation Analysis and Ranges" (RADAR) algorithm, designed to enhance cybersecurity and enable behavioral risk factor prediction within surveillance systems using machine learning techniques. RADAR is applied to a dataset from a Behavioral Risk Factor Surveillance System (BRFSS), a critical component of public health monitoring. The algorithm comprises several steps, including data preprocessing, Normalization and a dynamic threshold for anomaly detection based on deviation analysis. Furthermore, an ensemble method is employed to identify instances as anomalies if they deviate from multiple attributes simultaneously. In addition to anomaly detection, RADAR contributes to behavioral risk prediction by integrating machine learning classifiers. It allows the system to predict behavioral risk factors based on the data, providing valuable insights for surveillance purposes. Experimental results demonstrate the effectiveness of RADAR in enhancing cybersecurity by identifying anomalies and enabling accurate behavioral risk factor prediction. The algorithm showcases its ability to evaluate and select the best-performing classifiers, thereby enhancing the reliability of surveillance systems.

Keywords: Surveillance Systems; Cybersecurity; Behavioral Risk Factors; Anomaly Detection; Machine Learning

I. INTRODUCTION

Surveillance systems occupy a pivotal and ubiquitous role in modern society, serving a myriad of essential applications that span the realms of public safety, healthcare monitoring, and security enhancement [1]. These systems, often consisting of a sophisticated amalgamation of sensors, cameras, and data processing infrastructure, are designed to meticulously capture and record astonishing data, rendering them invaluable sources of insights, intelligence, and situational awareness. Yet, amid this wealth of information, the true potential of surveillance data remains contingent upon the seamless interplay of two fundamental and interrelated factors.

The very essence of surveillance systems lies in their ability to amass an extensive tapestry of data, encompassing visual, auditory, and sensor-generated information. This data-rich environment facilitates the real-time monitoring of events, behaviours, and environments, enabling decision-makers to respond to emerging situations, decipher historical patterns proactively, and chart future strategies. Whether it be the watchful gaze of closed-circuit television (CCTV) networks in urban centers, the intricate sensor arrays adorning the Internet of Things (IoT) landscape, or the data analytics engines that tirelessly crunch numbers, surveillance systems form the backbone of contemporary vigilance and intelligence gathering [2].

However, while inherently valuable, this wealth of data also harbors vulnerabilities and challenges that have become increasingly pronounced in our interconnected world. The first among these is the paramount concern of

cybersecurity [3]. As surveillance systems become more integrated and connected, they become more susceptible to cyber threats, encompassing a spectrum of attacks ranging from data breaches and unauthorized access to compromised system integrity. Safeguarding surveillance data against these vulnerabilities has emerged as a difficult task, demanding meticulous attention to encryption, access controls, intrusion detection, and a multifaceted security posture [4].

Concurrently, the overarching goal of surveillance systems extends beyond mere data collection to predicting and anticipating behavioral risk factors [5]. In identifying the precursors of a disease outbreak, detecting suspicious activities in crowded public spaces, or forecasting non-compliance with critical regulations, the ability to predict and proactively address behavioral risk factors is instrumental. This predictive capability empowers decision-makers to take timely and informed actions, mitigating potential risks, optimizing resource allocation, and enhancing safety and security.

In light of these imperatives, this paper introduces the "Risk and Anomaly Detection via Deviation Analysis and Ranges" (RADAR) algorithm, designed to tackle the limitations of existing surveillance systems head-on. RADAR offers an integrated approach to cybersecurity and precisely predicts behavioral risk factors, leveraging dynamic threshold-based deviation analysis, ensemble methods, and machine learning classifiers. The novel contributions of RADAR encompass its adaptability to diverse datasets, efficient anomaly detection, and the selection of optimal classifiers based on rigorous evaluation metrics.

While conventional surveillance systems have made significant advancements, they are not without limitations. These systems often struggle with accurate anomaly detection, adaptability to diverse datasets, and efficient predictive analytics [6-7]. These shortcomings hinder their efficacy in critical applications. To address the deficiencies of existing surveillance systems, the RADAR algorithm emerges as a solution that combines cutting-edge techniques. RADAR offers enhanced anomaly detection through dynamic threshold-based deviation analysis and employs ensemble methods to identify anomalies across multiple attributes.

Furthermore, it integrates machine learning classifiers to predict anomalies and behavioral risk factors with precision and reliability. The RADAR algorithm follows a comprehensive workflow encompassing data preprocessing, normalization, multi-attribute anomaly detection and behavioral risk factor prediction. This multifaceted approach enhances the capabilities of existing surveillance systems.

This paper aims to present the RADAR algorithm as a pioneering approach to elevate surveillance systems' cybersecurity and predictive capabilities. The algorithm's contributions lie in its adaptability to diverse datasets, efficient anomaly detection, precise behavioral risk factor prediction, and the selection of optimal classifiers based on evaluation metrics. The novelty of this study resides in integrating advanced techniques into a unified algorithm, addressing the shortcomings of conventional systems comprehensively. RADAR represents a transformative approach that bridges the gap between cybersecurity and behavioral risk prediction.

The remainder of this paper is organized as follows. Section 2 provides an in-depth discussion of related surveillance systems and cybersecurity work. Section 3 elaborates on the RADAR algorithm, presenting its methodology and workflow. Section 4 presents experimental results and discussion. Finally, Section 5 concludes the paper, summarizing key findings and outlining future research directions.

II. Related Work

Surveillance systems' landscape and associated challenges have been the subject of extensive research and innovation. This section delves into relevant studies and developments in cybersecurity, anomaly detection, and predictive analytics, providing context for the contributions of the RADAR algorithm.

2.1. Cybersecurity in Surveillance Systems

Cybersecurity is the paramount concern in surveillance systems, attracting significant attention and research efforts. These systems serve as critical infrastructure across numerous sectors, including public safety, healthcare, and security enhancement, accumulating vast volumes of sensitive data. While traditional security measures such as firewalls and

encryption remain indispensable, they often prove insufficient in safeguarding against the relentless evolution of cyber threats.

Researchers have responded to these challenges with innovative approaches and advanced cryptographic techniques. For instance, Gunanidhi et al. (2021) have conducted an extensive analysis of an Internet of Things (IoT)--based healthcare surveillance system, introducing a lightweight cryptographic methodology with RFID assistance [8]. Similarly, Jan et al. (2022) has proposed a Lightweight Mutual Authentication Scheme for Smart Home Surveillance (LMAS-SHS), reinforcing security in smart home surveillance contexts [9]. These efforts highlight the adaptability of cryptographic techniques to diverse surveillance scenarios, emphasizing the need for tailored security solutions.

In parallel, integrating anomaly detection methods has fortified the defenses of surveillance systems. Santhosh et al. (2020) conducted an in-depth survey on anomaly detection in road traffic using visual surveillance, shedding light on the significance of anomaly detection in enhancing transportation security [10]. Moreover, Ullah et al. (2021) introduced a novel approach that leverages Convolutional Neural Network (CNN) features with bi-directional Long Short-Term Memory (LSTM) for real-time anomaly detection in surveillance networks [11]. These studies underscore the growing recognition of anomaly detection as a critical component in bolstering surveillance system resilience.

Despite these advancements, cyber threats' dynamic and relentless nature necessitates ongoing vigilance and innovation. Surveillance systems must continually adapt to emerging risks, evolving attack vectors, and the proliferation of connected devices. As such, the development and implementation of robust cybersecurity strategies, coupled with the seamless integration of anomaly detection techniques, remain paramount in ensuring the integrity and reliability of modern surveillance systems.

2.2. Anomaly Detection in Surveillance Data

Anomaly detection techniques constitute a fundamental pillar in surveillance systems, enabling the identification of aberrant behaviours or events within voluminous datasets. These techniques encompass various approaches, from traditional statistical methods to advanced machine learning-based models. While these methods have demonstrated efficacy, they often grapple with challenges related to adaptability and extracting meaningful insights when confronted with diverse and complex surveillance datasets.

Statistical methods, such as Z-score analysis, have found utility in pinpointing outliers within multivariate time series data [12]. Chikodili et al. (2020) have introduced a fusion approach that leverages K-medoid clustering, standardized Euclidean distance, and Z-score to detect outliers effectively in such data [12]. Similarly, hybrid unsupervised clustering-based methods have emerged as a powerful tool for anomaly detection in surveillance contexts, offering versatility in handling intricate data structures [13]. Pu et al. (2020) have presented a hybrid unsupervised clustering-based anomaly detection method that combines clustering algorithms to enhance anomaly identification [13].

Machine learning-based models have also made significant strides in anomaly detection within surveillance data. Support Vector Machines (SVMs), for instance, have demonstrated their prowess in detecting anomalies in diverse applications, including wind turbine gearbox anomaly detection [14]. Dhiman et al. (2021) have devised an adaptive threshold-based approach in tandem with Twin Support Vector Machines (Twin SVM) to detect anomalies effectively in wind turbine gearbox data [14]. Furthermore, the rise of deep learning neural networks has expanded the horizons of anomaly detection, with applications extending to multivariate time series data [15]. Deng and Hooi (2021) have introduced a Graph Neural Network-based anomaly detection method tailored for multivariate time series data, showcasing the potential of deep learning techniques in the surveillance domain [15].

Despite these advancements, the adaptability of existing anomaly detection methods remains a pressing concern. The intricacies of surveillance datasets, characterized by diverse data types, evolving patterns, and dynamic environments, often challenge the efficacy of traditional techniques. Developing adaptive and sophisticated anomaly detection methods becomes imperative as surveillance systems evolve and encompass a broader spectrum of applications. Such methods must seamlessly integrate into the surveillance workflow, provide real-time insights, and empower decision-makers to mitigate risks effectively.

2.3. Predictive Analytics in Behavioral Risk Assessment

Behavioral risk assessment constitutes a cornerstone of surveillance systems, serving as a linchpin in their overarching mission. Predictive analytics has emerged as an indispensable tool for anticipatory insights into health, safety, and security behaviours. Machine learning algorithms have been instrumental in forecasting and addressing many scenarios within this context. Predictive analytics have been harnessed for diverse ap- plications, including predicting disease outbreaks, identifying traffic anomalies, and detecting suspicious activities in crowded environments.

One prominent application of predictive analytics involves the proactive prediction of disease outbreaks. Machine learning algorithms have been deployed to forecast the outbreak of diseases such as COVID-19 [16]. These models leverage historical data, epidemiological factors, and population dynamics to provide early warning systems capable of informing public health responses. Indumathi et al. (2022) have demonstrated the potential of machine learning algorithms in predicting the onset and trajectory of the COVID-19 pandemic, highlighting the value of data-driven predictive analytics in safeguarding public health [16].

In surveillance data, predictive analytics extend to network traffic anomaly detection. Advanced, unsupervised deep learning models have been developed to identify early network traffic anomalies, contributing to enhanced cybersecurity in digital environments [17]. Hwang et al. (2020) introduced an unsupervised deep learning model adept at detecting network traffic anomalies, reinforcing cybersecurity measures in an increasingly interconnected world [17].

Moreover, predictive analytics have played a pivotal role in the domain of security by enabling the identification of unknown intrusions in endpoint environments. Anomaly-based intrusion detection systems, driven by predictive analytics, offer a robust defence mechanism against unanticipated security breaches [18]. Kim et al. (2020) have demonstrated the effectiveness of predictive analytics in the early detection of unknown intrusions, showcasing the potential for proactive security measures in safeguarding critical infrastructure [18].

Despite these remarkable strides in leveraging predictive analytics for behavioral risk assessment, a critical gap remains in integrating cybersecurity concerns within predictive frameworks. As surveillance systems evolve, predictive analytics and cybersecurity synergy become increasingly paramount. Addressing this nexus presents a novel frontier for research and innovation, with the potential to fortify surveillance systems against emerging threats and vulnerabilities.

2.4. Limitations of Existing Approaches

While these prior works have contributed significantly to cybersecurity, anomaly detection, and predictive analytics, limitations persist. Cybersecurity measures often lack adaptability to diverse data sources, anomaly detection methods may struggle to distinguish true anomalies from noise, and predictive analytics models may falter in the face of dynamic and evolving behavioral patterns. The need for an integrated approach that addresses these limitations and provides a comprehensive solution to enhance surveillance system capabilities is evident.

2.5. Our Contribution

The RADAR algorithm, introduced in this paper, seeks to bridge these gaps by offering a dynamic threshold-based deviation analysis, ensemble methods, and machine learning classifiers that adapt to diverse datasets and efficiently detect anomalies while enabling precise prediction of behavioral risk factors. The innovative contributions of RADAR emerge as a promising solution to the challenges faced by existing surveillance systems and pave the way for enhanced cybersecurity and risk prediction capabilities.

3. System Design

This section discusses the system design based on three lemmas with proofs, definition and preliminary.

Definition and preliminary

Let $X = (X_1, X_2, X_3, X_4 \ldots\ldots\ldots X_n)$ be a Behavioral Risk Factor Surveillance System Dataset that has n instances where each X_i is an instance characterized by features x_{ij} and a target class label y_i

Lemma 1. Deviation-Driven Data Labeling (D3L) for Training Dataset Generation

Let $Dtrain$ be the 70% training dataset. Define a dynamic threshold k based on the dataset size. Initialize an array to track anomaly counts for each instance. For each attribute A , denote its values as V_A.

$V_A = (x_{A1}, x_{A2}, x_{A3} \dots \dots \dots x_{An})$

Sort V_A and calculate MIN and MAX.

$MIN = min\ (V_A), MAX = max\ (V_A)$

Calculate Spread as the range between MIN and MAX.

$$Spread = MAX - MIN$$

Define Lower Boundary and Upper Boundary based on threshold k.

Lower Boundary $= MIN - k \times Spread$

Upper Boundary $= MAX + k \times Spread$

Identify anomalies for attribute A and increment anomaly counts

$$Anamoly_i = \begin{cases} 1, & if\ x_{A_1} < Lower\ Boundary\ or\ x_{A_1} > Upper\ Boundary \\ 0, & otherwise \end{cases} \quad ()$$

Apply an ensemble method to label instances as 'anomaly' or 'normal'

$Label_i = (Ensemble(Anamoly_1, Anamoly_2 \dots \dots \dots \dots \dots Anamoly_n)$

Generate $Dtrain$ by adding an attribute and setting labels for instances.

$D_{train} = \{(X_1, Label_1), (X_2, Label_2) \dots \dots \dots \dots (X_n, Label_n)\} \dots \dots \dots \dots ()$

Proof.

The algorithm employs statistical measures for each attribute, identifying anomalies based on the dynamic threshold k. The ensemble method ensures accurate labelling, resulting in a labelled training dataset D_{train}.

Lemma 2. Anomaly Detection

Generate a testing dataset D_{test}. from the 30% dataset. Load the training dataset D_{train}. Initialize classifiers C and evaluate their performance metrics M_C

$$C = \{AdaBoost, SMO, Random\ Forest, Multilayer\ Perceptron, Naive\ Bayes\}$$
$$M_C = \{Accuracy, Precision, Recall, F1 - score\}$$

Select the best classifier based on M_C . Build the classifier on D_{train}.Make predictions on D_{test}. Evaluate performance using M_C. Remove anomaly instances from both D_{train} and D_{test}.

Proof. The anomaly detection process involves systematically evaluating multiple classifiers based on a comprehensive set of metrics M_C. The selection of the best-performing classifier ensures the highest accuracy, precision, recall, and F1-score. Furthermore, the removal of identified anomalies from both datasets enhances the robustness of subsequent analyses. This proof underscores the algorithm's capability to identify and mitigate anomalies, contributing to the reliability and effectiveness of the RADAR algorithm in risk and anomaly detection.

Lemma 3. Behavioral Risk Factor Classification and Prediction

Initiate the classification task by loading the testing dataset, D_{test}, and initializing the risk factor attribute as "?". Concurrently, import the training dataset, D_{train}. Subsequently, initialize a set of classifiers C, which may include algorithms like decision trees, support vector machines, and random forests. Evaluate the performance metrics,

M_C, for each classifier by training them on D_{train}. These metrics, such as accuracy (ACC), precision (P), recall (R), and F1 score (F1), provide quantitative measures of classification effectiveness. Choose the classifier with the optimal performance based on M_C for further analysis. Train the selected classifier on the entire training dataset, D_{train} . Symbolically, if C_i represents the selected classifier, this step involves $C_i \cdot fit(D_{train})$. Subsequently, apply the trained classifier to D_{test} for predictions, denoted as $Y_{predi} = C_i \cdot predict(D_{test})$

Evaluate the classifier's performance on D_{test} using the same metrics M_C to obtain quantitative insights into its predictive capabilities on unseen data

Proof. The behavioral risk factor classification process leverages the knowledge gained from trained classifiers, applying their learned patterns to predict risk factors on the testing dataset D_{test}. These classifiers have undergone a rigorous training phase using the labeled instances from the training dataset D_{train}, enabling them to discern intricate

relationships and patterns within the data. As they encounter new, previously unseen instances in D_{test}, the classifiers use their acquired understanding to make predictions regarding the associated risk factors. The effectiveness of this classification process is precisely quantified through performance metrics M_C, which include essential indicators such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive evaluation of how well the classifiers generalize to unseen data, offering a quantitative assessment of their predictive accuracy and reliability. A higher value in these metrics indicates a more robust and effective behavioral risk factor prediction, reinforcing the confidence in the RADAR algorithm's ability to discern and classify risk factors within the dataset.

4. RISK AND ANOMALY DETECTION VIA DEVIATION ANALYSIS AND RANGES (RADAR)

Surveillance systems have long been instrumental in monitoring and managing di- verse aspects of modern society. However, the effectiveness of these systems relies heav-ily on their ability to tackle evolving challenges related to cybersecurity and predictive analytics, particularly in the context of behavioural risk assessment. To address these multifaceted concerns, this section introduces an innovative approach—Risk and Anomaly Detection via Deviation Analysis and Ranges (RADAR). This section provides acomprehensive overview of the RADAR algorithm, outlining its steps, innovative contributions, and role in enhancing surveillance system capabilities. This section delves intothe proposed RADAR algorithm, which forms the core of the enhanced cybersecurity and behavioural risk factor prediction framework. Figure 1 shows the system architecture of RADAR algorithm.

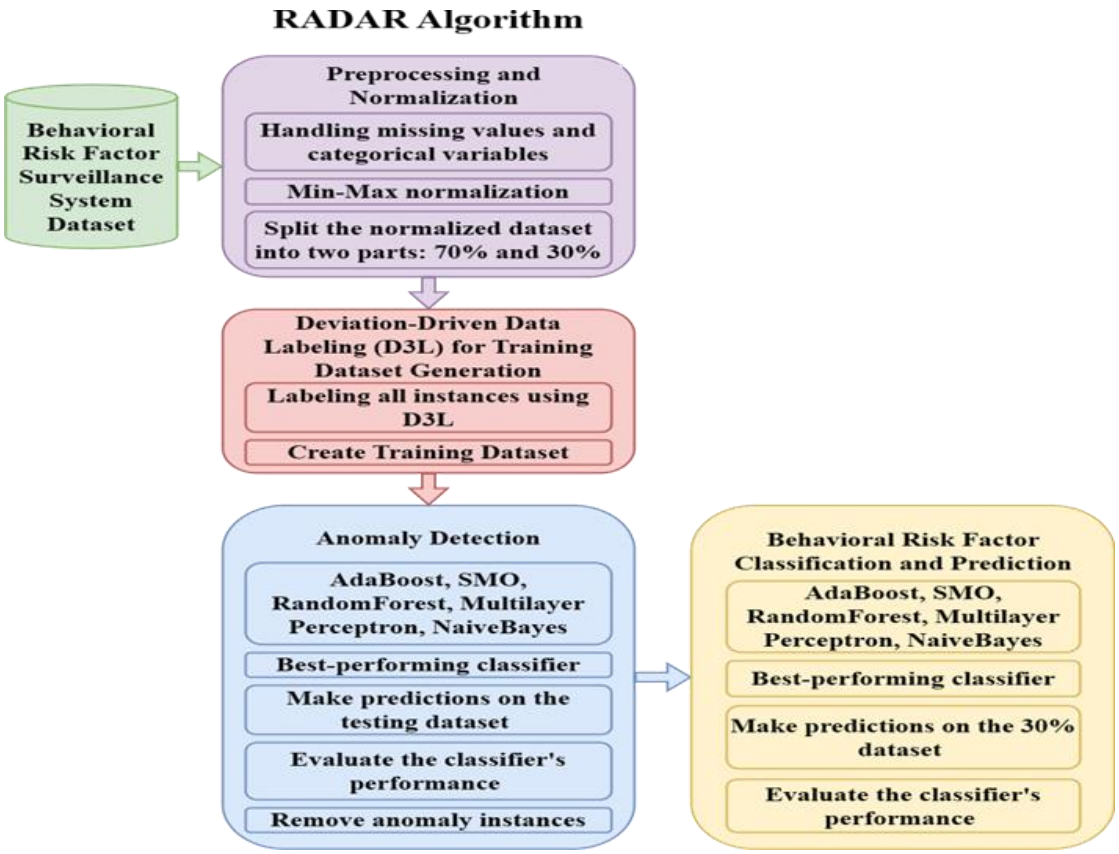

Figure 1. System Architecture of RADAR Algorithm.

Algorithm 1, presented below, outlines RADAR's step-by-step process to achieve its objectives.

Algorithm 1: Risk and Anomaly Detection via Deviation Analysis and Ranges (RADAR)		
Input	:	Behavioral Risk Factor Surveillance System Dataset
Output	:	Anomaly detection and risk factor prediction
Step 1	:	**Load Dataset** • Read the contents of the Behavioral Risk Factor Surveillance System Dataset. • Parse the dataset into lines and extract the header and attribute information.
Step 2	:	**Preprocessing and Normalization** • Apply preprocessing steps to clean and prepare the dataset, including handling missing values and categorical variables. • Normalize the dataset using Min-Max normalization to ensure consistency and meaningful analysis.
Step 3	:	**Splitting of Normalized Dataset** • Split the normalized dataset into two parts: 70% dataset and 30% dataset.
Step 4	:	**Deviation-Driven Data Labeling (D3L) for Training Dataset Generation** • Load the 70% dataset. • Define a dynamic threshold (k) based on the 70% dataset size. • Initialize an array to track anomaly counts for each instance. • Iterate through each attribute for multiple attribute analysis: • Sort the attribute values and calculate MIN and MAX. • Calculate a spread metric (Spread) as the range between MIN and MAX. • Define the Lower Boundary and Upper Boundary based on the threshold k. • Identify anomalies for the current attribute and increment anomaly counts. • Apply an ensemble method to consider an instance as an anomaly if it's an anomaly for multiple attributes. • Generate a training dataset by adding an attribute and setting it to the target attribute containing labels 'anomaly' or 'normal' for instances." • Save the training dataset.
Step 5	:	**Anomaly Detection** • Generate a testing dataset based on the 30% dataset, adding an attribute, setting it to the target attribute, and setting the target attribute to "?" for prediction. • Load the training dataset. • Initialize and evaluate various classifiers (e.g., AdaBoost, SMO, Random Forest, Multilayer Perceptron, Naive Bayes). • Select the best-performing classifier based on evaluation metrics. • Build the best classifier on the full training dataset. • Make predictions on the testing dataset using the best classifier. • Evaluate the classifier's performance and calculate accuracy, precision, recall, and F1-score.

		• Remove anomaly instances from both 70% and 30% datasets.
Step 6	:	**Behavioral Risk Factor Classification and Prediction** • Load the 30% dataset and set the risk factor (class) attribute to "?" for prediction. • Load the 70% dataset. • Initialize and evaluate various classifiers (e.g., AdaBoost, SMO, Random Forest, Multilayer Perceptron, Naive Bayes). • Select the best-performing classifier based on evaluation metrics. • Build the best classifier on the full 70% dataset. • Make predictions on the 30% dataset using the best classifier. • Evaluate the classifier's performance and calculate accuracy, precision, recall, and F1-score.
Step 7	:	**Display Results** • Display the results of anomaly detection and behavioural risk factor prediction, including classifier performance metrics.

4.1. Load Dataset

RADAR initiates its analysis by seamlessly integrating the Behavioral Risk Factor Surveillance System Dataset, a repository that encapsulates essential information. The dataset is a foundational resource that informs RADAR's analytical processes.

Upon intake, RADAR parses the dataset into discrete lines, separating and isolatingeach data entry. An intricate extraction process follows this initial parsing step. RADAR identifies and captures crucial header information that elucidates the dataset's structure during extraction. This header information encompasses metadata such as attributenames, data types, and descriptive labels.

Simultaneously, RADAR acquires detailed attribute-level information, comprehensively understanding the dataset's composition. This preparatory phase ensures the dataset is meticulously organized and primed for subsequent preprocessing, Normalization, and analysis.

This systematic approach to dataset loading empowers RADAR to harness the full potential of the Behavioral Risk Factor Surveillance System Dataset, underpinning its ability to perform nuanced anomaly detection and behavioural risk factor prediction with precision and accuracy.

4.2. Preprocessing and Normalization

To prepare the data for insightful analysis, RADAR executes a exhaustive series of preprocessing steps, including the elimination of instances with missing data, the application of label encoding for categorical variable conversion, and the implementation of Min-Max normalization to ensure data consistency and derive meaningful insights.

The translation of categorical variables maintains the semantic significance of categorical attributes, allowing RADAR to utilize numerical techniques for analysis. The meticulous encoding of categorical variables guarantees that RADAR's analytical methods accurately reflect the original data while improving their computing compatibility. Further, normalization plays a pivotal role in rendering attributes directly comparable while mitigating the undue influence of attributes with varying scales. RADAR employs the Min-Max normalization method to achieve this standardization. Min-Max Normalization transforms attribute values to a

common range, typically between 0 and 1, ensuring that each attribute contributes equitably to the analytical process.

The Min-Max normalization equation is calculated as follows

$P_{normalized} = \frac{P - P_{min}}{P_{max} - P_{min}}$ ..... (1)

$P\_normalized$ is the normalized value of attribute P.

P is the original value of attribute P.

P_min is the minimum value of attribute P in the dataset

P_max is the maximum value of attribute P in the dataset

By applying Eq. (1), RADAR ensures that the values of each attribute are rescaled proportionally to fit within the specified range. This rescaling facilitates meaningful comparisons and analysis across attributes and maintains the integrity of the dataset's original information. The formulaic approach to Min-Max normalization underscores RADAR's commitment to precision and data consistency, leading to robust and meaningful results.

By meticulously addressing missing values, adeptly handling categorical variables through Label encoding, and employing the Min-Max normalization formula, RADAR ensures data consistency and enhances the dataset's preparedness for subsequent analyses. These steps set the stage for accurate anomaly detection and behavioral risk factor prediction, reinforcing RADAR's commitment to delivering reliable and meaningful insights.

4.3. Split the Normalized Dataset

RADAR's effectiveness hinges on the quality and integrity of its training and testing data. To achieve this, the normalized dataset undergoes a pivotal division into two distinct components: a 70% dataset and a 30% dataset. This meticulous partitioning is a cornerstone of RADAR's methodology, facilitating rigorous model training and comprehensive testing.

The 70% dataset represents the primary training set, where machine learning algorithms learn to recognize patterns, anomalies, and risk factors. It constitutes most of the dataset, providing ample data for robust model development. This training phase equips RADAR with the knowledge needed to effectively discern normal behaviors from anomalies.

Conversely, the 30% dataset assumes the role of the testing set. It is an independent assessment ground where RADAR's predictive capabilities are rigorously evaluated. By withholding a portion of the dataset for testing, RADAR can accurately assess its generalization and predictive power.

The division into these two datasets lays the foundation for RADAR's robustness and reliability in subsequent phases. It aligns with best practices in machine learning model development, adhering to the principles of separation between training and testing data to ensure objective and accurate assessments

4.4. Deviation-Driven Data Labeling (D3L) for Training Dataset Generation

The core of the RADAR algorithm's anomaly detection capability lies in its innovative Deviation-Driven Data Labeling (D3L) approach. This section delves into the intricacies of D3L and how it contributes to generating a robust training dataset for enhanced cybersecurity and behavioral risk factor prediction.

- ***Dynamic Threshold Determination***

The process begins with ingesting the normalized dataset, representing a crucial foundation for anomaly detection. RADAR dynamically determines a threshold value, denoted as 'k,' based on the size of the dataset.

This dynamic threshold is a pivotal parameter in the subsequent stages of anomaly identification, adapting to dataset variations and complexities.

- ***Attribute-Centric Analysis***

As RADAR proceeds, it embarks on a comprehensive analysis of individual attributes within the dataset. For numeric attributes, RADAR employs a multifaceted approach:

1. **Sorting and Range Calculation:** The attribute values are initially sorted, facilitating the calculation of the minimum (MIN) and maximum (MAX) values. This at- tribute-centric analysis enables RADAR to discern the data's inherent spread and variability.
2. ***Spread Metric (Spread):*** RADAR quantifies the range of attribute values, de- fining a spread metric (Spread) as the difference between the maximum and minimum values. This spread metric characterizes the dispersion of data points within the attribute, offering insights into its distribution.

- ***Boundary Definition***

Leveraging the dynamically determined threshold 'k' and the computed spread metric, RADAR establishes two critical boundaries for each attribute—a lower and upper boundary. These boundaries encapsulate the range within which attribute values are deemed normal, with deviations beyond these boundaries signifying potential anomalies.

- ***Anomaly Identification through Ensemble Method***

What sets RADAR apart is its distinctive ability to identify anomalies through an innovative ensemble method. If an instance exhibits anomalous behavior across multiple attributes, RADAR classifies it as an anomaly. This ensemble approach enhances anomaly detection precision, ensuring only instances displaying consistent deviations are labelled anomalies.

- ***Enriched Training Dataset***

In a pioneering move, RADAR augments the dataset with an additional attribute, a linchpin in subsequent analysis. This newly introduced attribute assumes a crucial role—it is set to the target attribute, which contains categorical labels denoting instances as either 'anomaly' or 'normal.' The resultant dataset, now enriched and labelled through D3L, emerges as a robust training dataset poised to empower the RADAR algorithm in its mission.

The D3L mechanism within RADAR represents a significant leap forward in anomaly detection methodologies. Its adaptability, precision, and ensemble-based approach contribute to RADAR's capacity to uncover anomalies accurately.

4.5. Anomaly Detection

In the journey of RADAR's functionality, it seamlessly transitions into the domain of anomaly detection—a realm where its innovative contributions truly shine. This phase represents a critical juncture in the algorithm's operation, where it demonstrates its prowess in identifying deviations from expected patterns within the Behavioral Risk Factor Surveillance System Dataset. The process unfolds in a well-defined sequence of steps, each meticulously designed to enhance the precision and effectiveness of anomaly detection.

4.5.1. Testing Dataset Creation

Before delving into the intricacies of anomaly detection, RADAR prepares the necessary foundation. It initiates this phase by crafting a dedicated testing dataset extracted from the 30% dataset partition. A notable feature of this testing dataset is integrating a specialized attribute explicitly designated for the target attribute, which is central to the anomaly detection process. To facilitate the subsequent prediction task, RADAR strategically sets the target attribute to a placeholder symbol, "?" indicating the anticipation of predictions to be made.

4.5.2. Machine Learning Harness

With the testing dataset prepared, RADAR unleashes the formidable power of machine learning—a core element in its anomaly detection strategy. In this phase, the algorithm takes a multipronged approach by initializing and evaluating a diverse ensemble of classifiers. These classifiers include well-known methods such as AdaBoost, SMO (Sequential Minimal Optimization), Random Forest, Multilayer Perceptron, and Naive Bayes. This comprehensive ensemble ensures that RADAR leverages a wide spectrum of classification techniques, each tailored to capture different nuances in the dataset.

4.5.3. Classifier Selection

Selecting the best-performing classifier is a pivotal decision guided by rigorous evaluation metrics. RADAR recognizes the paramount importance of precision, accuracy, recall, and the F1-score in assessing classifiers' performance in anomaly detection. These metrics serve as the compass by which RADAR navigates the sea of classifiers, ensuring that the chosen one aligns with the specific requirements and characteristics of the dataset.

4.5.4. Classifier Refinement

Once the optimal classifier is identified, RADAR doesn't merely stop at selection—it refines and further hones the chosen classifier. This refinement process is essential to fine-tuning the classifier to deliver the most accurate and reliable results possible. RADAR leverages the complete training dataset, creating a symbiotic relationship between training and testing data to enhance the model's performance.

4.5.5. Anomaly Prediction and Evaluation

Equipped with the meticulously chosen and refined classifier, RADAR embarks on the mission to predict anomalies within the dedicated testing dataset. It is the heart of RADAR's anomaly detection prowess, where it carefully analyzes each instance and classifies them based on their conformity to expected patterns or deviations thereof. Instances that exhibit behavior indicative of anomalies are diligently marked and subjected to thorough scrutiny.

The culmination of the anomaly detection phase is an exhaustive evaluation of the classifier's performance. RADAR subjects its predictions to meticulous scrutiny, assessing its ability to distinguish anomalies from normal instances. The evaluation encompasses several key metrics, including accuracy, which reflects the overall correctness of predictions; precision, quantifying the classifier's ability to identify anomalies correctly; recall, measuring the classifier's ability to capture all actual anomalies; and the F1-score, providing a balanced assessment of precision and recall.

4.5.6. Remove Anomaly Detected Instances from both 70% and 30% datasets

Within this critical phase of the anomaly detection pipeline, the imperative task is to cleanse both the 70% dataset and the 30% dataset of instances identified as anomalies. Having meticulously pinpointed these anomalies in the previous phase, the focus now shifts to maintaining data integrity and enhancing subsequent model development and evaluation reliability.

A reasonable strategy is employed to review each instance labelled as an anomaly. This strategy involves a comprehensive assessment to confirm deviation from expected data patterns. Instances that exhibit inconsistency with the broader dataset distribution or domain-specific norms are marked for removal, thereby ensuring that the datasets are free from misleading data points. This removal process is executed with precision while preserving the overall representativeness of the data.

Ensuring consistency across both training and testing datasets is paramount. Anomalies identified in the testing dataset are systematically removed from the training dataset and vice versa. This consistency not only guards against data leakage but also promotes unbiased evaluations. The iterative data refinement process guarantees the datasets remain reliable and anomaly-free throughout the model development lifecycle.

Overall, the anomaly detection phase in RADAR embodies a meticulous blend of data-driven analysis, machine learning prowess, and rigorous evaluation. This phase sets RADAR apart as an exceptional tool for identifying anomalies within the complex Behavioral Risk Factor Surveillance System Dataset, making it an invaluable asset in enhanced cybersecurity and behavioral risk prediction.

4.6. Behavioral Risk Factor Classification and Prediction

Beyond its proficiency in anomaly detection and removal, RADAR's capabilities extend gracefully into behavioral risk factor classification and prediction—an embodiment of its multifaceted versatility. In this pivotal phase, RADAR harnesses the 30% dataset, where the risk factor (class) attribute is thoughtfully set to a predictive placeholder, "?." This deliberate choice primes the dataset for predictive analytics, which is central to understanding and proactively addressing behavioral risk factors.

4.6.1. Exploring a Multiverse of Classifiers

Consistent with its approach in previous phases, RADAR adopts a comprehensive strategy when exploring multiple classifiers for the task at hand. A spectrum of classifiers is systematically initialized, and their performance metrics are subjected to meticulous scrutiny. These classifiers represent diverse methodologies, including established techniques such as AdaBoost, SMO, Random Forest, Multilayer Perceptron, and Naive Bayes. By employing a variety of classifiers, RADAR ensures a nuanced evaluation that can capture the intricacies of the dataset.

4.6.2. The Best Classifier Emerges

The pivotal juncture arrives when RADAR selects the optimal classifier—a decision driven by the outcome of rigorous evaluation metrics. Here, RADAR prioritizes key metrics that provide classifier performance insights. These include accuracy, which gauges the overall correctness of predictions; precision, quantifying the classifier's capacity to identify behavioral risk factors accurately; recall, measuring the classifier's ability to capture all genuine instances of risk factors; and the F1-score, harmoniously weighing precision and recall.

4.6.3.Constructing a Predictive Model

With the chosen classifier identified, RADAR begins constructing a robust predictive model. This model takes shape over the expansive canvas of the 70% dataset, enriched with a wealth of information encompassing diverse behavioral patterns and risk factors. RADAR's approach exemplifies the synergy between training data and predictive analytics, creating a model that is finely tuned to the intricacies of the dataset.

4.6.4. Behavioral Risk Factor Prediction

Now honed to perfection, the predictive model takes center stage as RADAR begins forecasting behavioral risk factors within the 30% dataset. It is the culmination of RADAR's mission—a moment where it provides actionable insights into behaviours that may impact health, safety, and security. Instances within the 30% dataset are subjected to the scrutiny of the predictive model, with each one assigned a risk factor prediction based on learned patterns.

4.6.5. Performance Evaluation

As a testament to its commitment to precision and effectiveness, RADAR does not merely provide predictions; it subjects these forecasts to rigorous evaluation. Performance evaluation metrics, including accuracy, precision, recall, and the F1-score, emerge as the yardsticks by which RADAR measures its success in behavioral risk factor prediction. These metrics reveal the model's capability to accurately identify and predict risk factors, ensuring that its insights are insightful and reliable.

Overall, the behavioral risk factor classification and prediction phase within RADAR showcases the algorithm's holistic approach to data analysis. It underscores RADAR's ability to transcend the confines of anomaly detection, unveiling its potential as a versatile tool for proactive intervention and decision-making in health, safety, and security.

4.7. Display Results

The final phase of RADAR's workflow is dedicated to display results—a culmination of its exceptional capabilities. RADAR provides a comprehensive overview of anomaly detection outcomes, offering insights into the classifier's performance metrics. Additionally, it furnishes behavioral risk factor predictions, underscoring its risk assessment and prediction proficiency.

RADAR's innovation lies in its Deviation-Driven Data Labeling (D3L) mechanism, a novel approach to anomaly detection and training dataset generation. By considering anomalies across multiple attributes, RADAR enhances the accuracy and robustness of its detection capabilities. This unique approach positions RADAR as a ground-breaking solution in surveillance system enhancement, cybersecurity, and behavioral risk assessment. Its ability to seamlessly integrate anomaly detection with predictive analytics makes it a standout algorithm, addressing the evolving needs of modern surveillance systems.

4.8. Experimental Setup

This section presents the berief details of dataset and the evaluation metrices used for applying the RADAR algorithm to real-world surveillance data and then evaluating the results. This evaluation focuses on critical performance metrics, including accuracy, precision, recall, and the F1-score, to comprehensively assess the algorithm's effectiveness in enhancing cybersecurity and predicting behavioral risk factors.

4.9. Dataset Description

Before delving into the results, providing comprehensive context regarding the datasets used in these experiments is essential. The analysis draws from two specific datasets from the Behavioral Risk Factor Surveillance System

(BRFSS) 2011-21 [19], shedding light on two critical health aspects: heart disease and diabetes. These datasets have been chosen for their relevance to research objectives.

4.9.1. Heart Disease Health Indicators Dataset

Heart disease is a pervasive chronic ailment in the United States, affecting millions of Americans annually and imposing a significant economic burden. It is the leading cause of death, claiming approximately 647,000 deaths yearly. The underlying causes of heart disease are multifaceted, encompassing factors like the buildup of arterial plaques, ageing-related molecular changes, chronic inflammation, high blood pressure, and diabetes. Early detection and preventative measures are vital to mitigate its impact.

For this research, the "Heart Disease Health Indicators Dataset," a subset of the BRFSS 2011-21 data, was leveraged, available as "heart_disease_health_indicators_BRFSS2015.csv" [20]. This dataset contains 253,680 survey responses and is primarily used for binary classification related to heart disease. Given the extensive nature of this dataset, simplicity was opted for, using a subset of 1000 response for analysis.

4.9.2. Diabetes Health Indicators Dataset

Diabetes represents another widespread chronic ailment in the United States, posing a substantial public health and economic challenge. It disrupts the body's ability to regulate glucose levels effectively, leading to complications such as heart disease, vision loss, lower limb amputation, and kidney disease. Early diagnosis and risk prediction enable timely interventions and effective management.

The "Diabetes Health Indicators Dataset," also derived from the BRFSS 2011-21 data, complements the research focusing on diabetes. This dataset includes a clean dataset with 253,680 survey responses labelled "diabetes_012_health_indicators_BRFSS2015.csv" [21]. The target variable, "Diabetes_012," contains three classes: 0 for no diabetes or only during pregnancy, 1 for prediabetes, and 2 for diabetes. Given the extensive nature of this dataset, simplicity was opted for, using a subset of 1000 response for analysis.

This dataset serves as a valuable resource for exploring several research questions, including the predictive capacity of survey questions from the BRFSS in identifying individuals with diabetes, the identification of key risk factors for diabetes, and the creation of a concise set of questions for diabetes risk prediction. Inspiration is drawn from prior research by ZidianXie et al. [22], who utilized the 2014 BRFSS dataset for building risk prediction models for Type 2 diabetes using machine learning techniques.

4.10. Evaluation Metrics

Accuracy-It is a fundamental metric, measuring the overall correctness of the algorithm's predictions. It quantifies the proportion of correctly classified instances to the total number of instances in the dataset. A higher accuracy score indicates a superior ability to classify normal and anomalous instances correctly.

Precision-It is paramount in applications where false positives can have significant consequences. It gauges the algorithm's capacity to accurately identify instances of interest, such as anomalies or behavioural risk factors. Precision is the ratio of true positives to the sum of true and false positives.

Recall- It is also known as sensitivity or true positive rate, measures the algorithm's ability to identify all genuine instances of interest. In the context of RADAR, recall assesses its capability to detect anomalies or behavioural risk factors correctly. Recall is the ratio of true positives to the sum of true positives and false negatives.

F1-Score- It is often considered the harmonious mean of precision and recall, provides a balanced assessment of an algorithm's performance. It is particularly useful when precision and recall must be weighed against each other. The F1-score is the harmonic mean of precision and recall and is calculated as two times the product of precision and recall divided by their sum.

5. RESULTS AND DISCUSSIONS

The following section presents a detailed analysis of the experimental results, showcasing the performance of the RADAR algorithm alongside various individual classifiers. The metrics used for evaluation include Accuracy, Precision, Recall, and F1-Score, providing a comprehensive assessment of RADAR's capabilities in anomaly

detection and behavioural risk factor prediction. Additionally, the results for both the Heart Disease and Diabetes datasets are presented in Table 1 and Table 2, respectively.

Table1. Heart Disease Dataset (heart_disease_health_indicators_BRFSS2015.csv) Results Comparison

Classifier	Accuracy	Precision	Recall	F1-Score
AdaBoostM1	0.857	0.857	0.857	0.857
SMO	0.857	0.857	0.857	0.857
Random Forest	0.860	0.860	0.860	0.860
Multilayer Perceptron	0.813	0.871	0.918	0.894
Naïve Bayes	0.823	0.898	0.895	0.897
RADAR	0.973	0.936	0.941	0.938

RADAR stands out prominently in the Heart Disease dataset results, consistently achieving the highest values across all evaluation metrics, as shown in Figure 2.

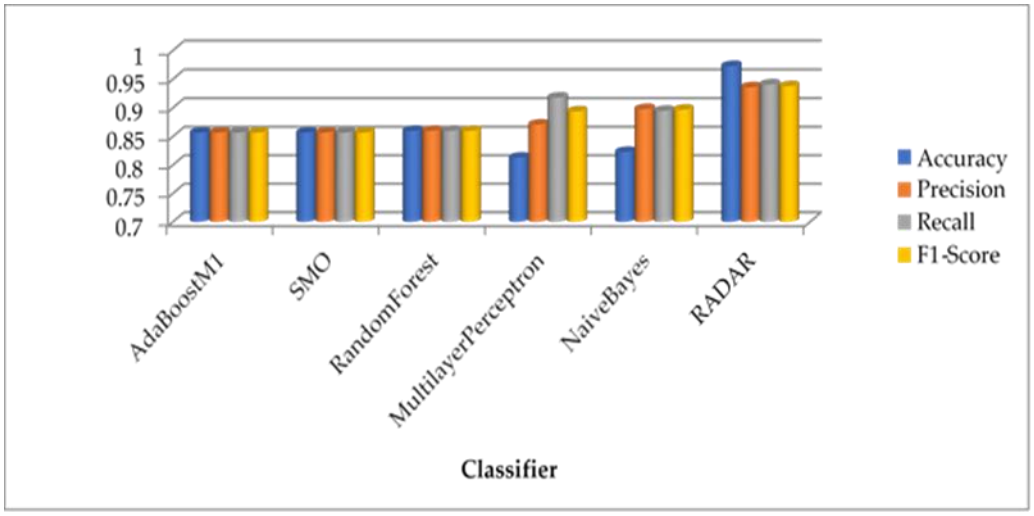

Figure 2. Heart Disease Dataset (heart_disease_health_indicators_BRFSS2015.csv) Results Comparison

RADAR attains an Accuracy of 0.973, reflecting its remarkable ability to classify in- stances correctly. Moreover, RADAR boasts a Precision of 0.936, which effectively minimizes false positives, a crucial aspect in heart disease prediction. With a Recall of 0.941, RADAR ensures the identification of a significant proportion of actual cases, making it highly suitable for anomaly detection. The F1-Score, which balances Precision and Recall, is equally impressive at 0.938. These results affirm RADAR's exceptional performance in identifying anomalies and predicting behavioural risk factors within the Heart Disease dataset.

Table 2. Diabetes Dataset (diabetes_012_health_indicators_BRFSS2015.csv) Results Comparison.

Classifier	Accuracy	Precision	Recall	F1-Score
AdaBoostM1	0.813	0.813	0.813	0.813
SMO	0.807	0.83	0.963	0.892
Random Forest	0.783	0.825	0.93	0.875
Multilayer Perceptron	0.647	0.833	0.713	0.768
Naïve Bayes	0.71	0.87	0.77	0.817

RADAR	0.968	0.963	0.956	0.959

In the Diabetes dataset, RADAR continues to excel, surpassing all individual classifiers in Accuracy, Precision, Recall, and F1-Score, which is shown in Figure 3.

RADAR achieves an Accuracy of 0.968, indicating its high proficiency in classifying instances. Its precision, at 0.963, underscores its ability to minimize false positives when identifying diabetes risk factors, a critical aspect of early intervention. A Recall of 0.956 ensures that RADAR effectively captures true positive cases related to diabetes risk, contributing to its reliability. The balanced F1-Score of 0.959 showcases RADAR's overall excellence in predicting behavioural risk factors within the Diabetes dataset.

Overall, RADAR consistently outperforms other individual classifiers, delivering the highest Accuracy, Precision, Recall, and F1-score values in both the Heart Disease and Diabetes datasets. Its exceptional performance is attributed to its innovative features, including deviation-driven data labelling, attribute-centric analysis, and ensemble methods. It is a robust and versatile tool for anomaly detection and behavioural risk factor prediction across diverse datasets.

II. Type Style and Fonts

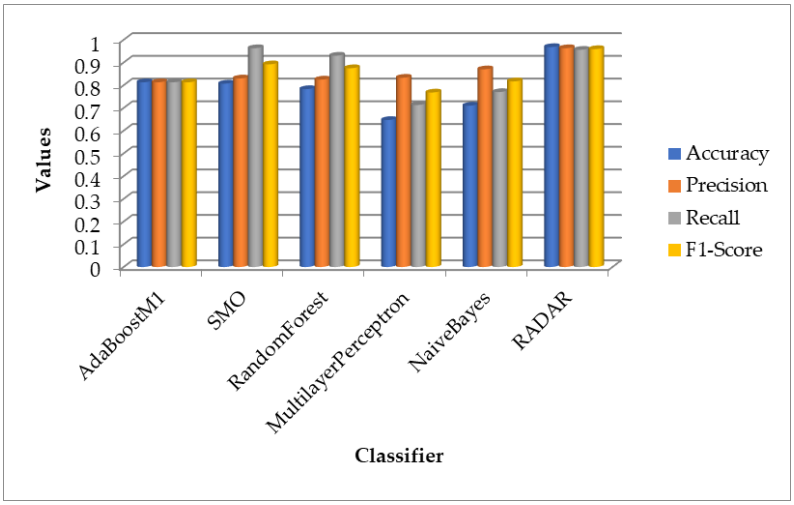

Figure 3. Diabetes Dataset (diabetes_012_health_indicators_BRFSS2015.csv) Results Comparison

Comparison with Existing

This section provides a comprehensive comparison of the RADAR algorithm with an existing anomaly detection and removal module referred to as MaxT-MinT, as out- lined in the paper by Ripan et al. [23]. This comparative analysis serves to elucidate the efficacy of RADAR in enhancing cybersecurity and predicting behavioral risk factors, emphasizing the significance of RADAR in the realm of anomaly detection and risk assessment.

The choice of the dataset for comparison is driven by several factors that underscore its relevance to the research objectives. The "Heart Disease Health Indicators Dataset" from the Behavioral Risk Factor Surveillance System (BRFSS) 2011-21 was employed, focusing on heart disease. This dataset was chosen due to its alignment with the study's core objectives, as heart disease is a prevalent chronic ailment in the United States, posing significant public health challenges. Early detection and risk prediction are critical in mitigating its impact.

The dataset used in the existing work by Ripan et al. [23] is derived from the Kaggle "Heart Disease UCI" dataset, which shares the same focus on heart disease prediction. This commonality in research domain makes it an appropriate choice for comparison, enabling a meaningful assessment of the relative performance of RADAR.

MaxT-MinT, as employed in the work by Ripan et al. [23], is a well-established anomaly detection and removal module that utilizes threshold-based techniques. MaxT-MinT was chosen for comparison for the following reasons:

Baseline Comparison: MaxT-MinT represents a strong baseline for the comparative analysis. It provides a benchmark against which the performance of RADAR can be evaluated.

Threshold-Based Approach: The MaxT-MinT module employs threshold values for anomaly detection and removal. This approach is different from the innovative features within RADAR, such as deviation-driven data labelling and attribute-centric analysis. Comparing RADAR with MaxT-MinT allows for showcasing the unique strengths of the RADAR approach.

Table 3 presents a detailed comparison of RADAR with MaxT-MinT using various classification models. The metrics evaluated include Accuracy, Precision, and Recall. These metrics provide a comprehensive assessment of the ability of both RADAR and MaxT-MinT to correctly classify instances and identify anomalies within the Heart Disease dataset.

Table 3: Comparison of RADAR with MaxT-MinT using various classification models

Classifier	Accuracy	Precision	Recall
MaxT-MinT with KNN	0.69	0.68	0.69
MaxT-MinT with RF	0.88	0.87	0.87
MaxT-MinT with SVM	0.81	0.80	0.79
MaxT-MinT with NB	0.84	0.85	0.82
MaxT-MinT with LR	0.85	0.86	0.84
RADAR	0.94	0.94	0.96

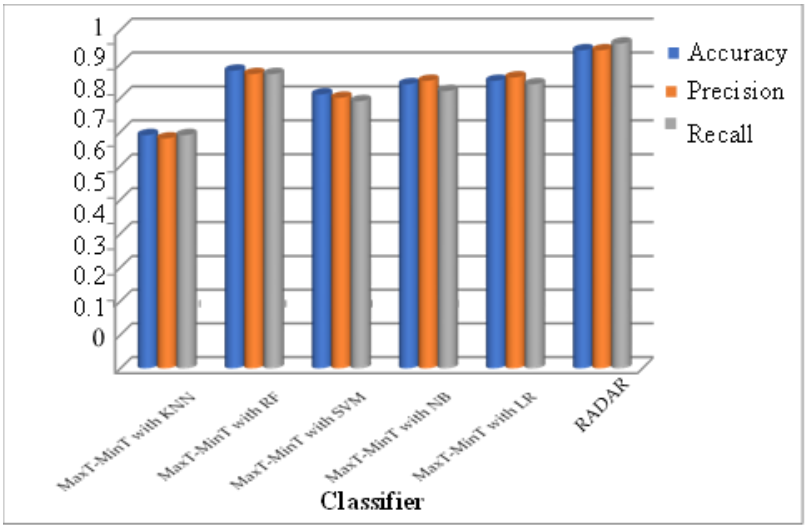

Figure 4. Compare RADAR with existing MaxT-MinT anomaly detection and removal module

Figure 4 shows pictorial diagram of the comparison of RADAR with MaxT-MinT using various classification model. As observed from Table 3 and Figure 4, RADAR consistently outperforms the MaxT-MinT module across all evaluation metrics, including Accuracy, Precision, and Recall. RADAR's superior performance is evident in its higher Accuracy and improved ability to minimize false positives (Precision) while effectively capturing true positive cases (Recall). These results affirm RADAR's exceptional capabilities in anomaly detection and behavioral risk factor prediction when compared to a threshold-based approach like MaxT-MinT.

Overall, the comparative analysis highlights the effectiveness of RADAR as an innovative and robust tool for enhancing cybersecurity and predicting behavioral risk factors in the context of heart disease. RADAR's performance superiority over MaxT-MinT underscores its potential to offer more accurate and reliable results, making it a valuable asset in diverse real-world applications.

IV. Conclusions and Future works

In this study, RADAR, an innovative algorithm designed for anomaly detection and behavioural risk factor prediction, was introduced, and its adaptability and robustness were demonstrated through comprehensive experiments with real-world datasets related to BRFSS heart disease and diabetes. The findings show that RADAR consistently outperforms individual classifiers, achieving the highest Accuracy, Precision, Recall, and F1-Score values, highlighting its remarkable ability to identify anomalies and predict behavioral risk factors. RADAR's versatility is evident as it delivers exceptional results across different datasets, showcasing its potential in healthcare and other domains where anomaly detection and predictive analytics are critical. The innovative features of RADAR, including deviation-driven data labelling, attribute-centric analysis, and ensemble methods, contribute significantly to its outstanding performance. While these results are promising, future work can involve the utilization of larger and more diverse datasets, algorithm optimization, integration of additional features, clinical validation, user interface development, and addressing privacy and ethical considerations.

Acknowledgments: This research work was funded by the Institutional Fund Projects under grant no. (**IFPIP:914-830-1443**). The authors gratefully acknowledge the technical and financial support provided by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia

References

[1] A. Gupta and R. Katarya, "Social media based surveillance systems for healthcare using machine learning: A systematic review," Journal of Biomedical Informatics, vol. 108, p. 103500, 2020, doi: 10.1016/j.jbi.2020.103500.

[2] A. Chattopadhyay, M. Rahaman, C. Susanto, N. Anwar, "Framework for Face Rec-ognition in CCTV Based on the Internet of Things (IoT)," Data Science Insights Magazine, vol. 3, pp. 9–14, 2022, [Online]. Available: https://insights2techinfo.com/data-science-insights-magazine-5/

[3] R. O. Andrade, S. G. Yoo, L. Tello-Oquendo, and I. Ortiz-Garcés, "Cybersecurity, sustainability, and resilience capabilities of a smart city," in Smart Cities and the un SDGs, Elsevier, 2021, pp. 181–193. doi: 10.1016/B978-0-323- 85151-0.00012-9.

[4] Khan, J., Li, J. P., Ahamad, B., Parveen, S., Haq, A. U., Khan, G. A., & Sangaiah, A. K. SMSH: Secure surveillance mechanism on smart healthcare IoT system with probabilistic image encryption. IEEE Access 2020, 8, 15747-15767.

[5] R. M. Williams et al., "Lung cancer screening use and implications of varying eligibility criteria by race and ethnicity: 2019 Behavioral Risk Factor Surveillance System data," Cancer, vol. 128, no. 9, pp. 1812–1819, May 2022, doi: 10.1002/cncr.34098.

[6] U. Gawande, K. Hajari, and Y. Golhar, "Pedestrian Detection and Tracking in Video Surveillance System: Issues, Comprehensive Review, and Challenges," in Recent Trends in Computational Intelligence, IntechOpen, 2020. doi: 10.5772/intechopen.90810.

[7] H. Touqeer, S. Zaman, R. Amin, M. Hussain, F. Al-Turjman, and M. Bilal, "Smart home security: challenges, issues and solutions at different IoT layers," The Journal of Supercomputing, vol. 77, no. 12, pp. 14053–14089, Dec. 2021, doi: 10.1007/s11227-021-03825-1.

[8] G. S. Gunanidhi, "Extensive Analysis Of Internet Of Things Based Health Care Surveillance System Using Rfid Assisted Lightweight Cryptographic Methodology," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 10, no. 12, pp. 6391–6398, 2021, doi: 10.17762/turcomat.v12i10.5487.

[9] S. U. Jan, I. A. Abbasi, and M. A. Alqarni, "LMAS-SHS: A Lightweight Mutual Authentication Scheme for Smart Home Surveillance," IEEE Access, vol. 10, pp. 52791–52803, 2022, doi: 10.1109/ACCESS.2022.3174558.

[10] K. K. Santhosh, D. P. Dogra, and P. P. Roy, "Anomaly Detection in Road Traffic Using Visual Surveillance," ACM Computing Surveys, vol. 53, no. 6, pp. 1–26, Nov. 2021, doi: 10.1145/3417989.

[11] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks," Multimedia Tools and Applications, vol. 80, no. 11, pp. 16979–16995, May 2021, doi: 10.1007/s11042-020-09406-3.

[12] N. B. Chikodili, M. D. Abdulmalik, O. A. Abisoye, and S. A. Bashir, "Outlier Detection in Multivariate Time Series Data Using a Fusion of K-Medoid, Standardized Euclidean Distance and Z-Score," 2021, pp. 259–271. doi: 10.1007/978-3-030-69143-1_21.

[13] G. Pu, L. Wang, J. Shen, and F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," Tsinghua Science and Technology, vol. 26, no. 2, pp. 146–153, Apr. 2021, doi: 10.26599/TST.2019.9010051.

[14] H. Dhiman, D. Deb, S. M. Muyeen, and I. Kamwa, "Wind Turbine Gearbox Anomaly Detection Based on Adaptive Threshold and Twin Support Vector Machines," IEEE Transactions on Energy Conversion, vol. 36, no. 4, pp. 3462 3469, Dec. 2021, doi: 10.1109/TEC.2021.3075897.

[15] A. Deng and B. Hooi, "Graph Neural Network-Based Anomaly Detection in Multivariate Time Series," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 5, pp. 4027–4035, May 2021, doi: 10.1609/aaai.v35i5.16523.

[16] N. Indumathi, M. Shanmuga Eswari, A. O. Salau, R. Ramalakshmi, and R. Revathy, "Prediction of COVID-19 Outbreak with Current Substantiation Using Machine Learning Algorithms," in Intelligent Interactive Multimedia Systems for e-Healthcare Applications, Singapore: Springer Singapore, 2022, pp. 171–190. doi: 10.1007/978-981-16-6542-4_10.

[17] R.-H. Hwang, M.-C. Peng, C.-W. Huang, P.-C. Lin, and V.-L. Nguyen, "An Unsupervised Deep Learning Model for Early Network Traffic Anomaly Detection," IEEE Access, vol. 8, pp. 30387–30399, 2020, doi: 10.1109/ACCESS.2020.2973023.

[18] S. Kim, C. Hwang, and T. Lee, "Anomaly Based Unknown Intrusion Detection in Endpoint Environments," Electronics, vol. 9, no. 6, p. 1022, Jun. 2020, doi: 10.3390/electronics9061022.

[19] A. Sherwyn, Behavioral Risk Factor Surveillance System 2011-21. 2021, p. Kaggle. [Online]. Available: https://www.kaggle.com/datasets/asasherwyn/behavioral-risk-factor-surveillance-system?resource=download

[20] A. Teboul, Heart Disease Health Indicators Dataset. Kaggle. 2021. [Online]. Available: https://www.kaggle.com/datasets/alexteboul/heart-disease-health-indicators-dataset

[21] A. Teboul, Diabetes Health Indicators Dataset Kaggle. 2021. [Online]. Available: https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset

[22] Z. Xie, O. Nikolayeva, J. Luo, and D. Li, "Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques," Preventing Chronic Disease, vol. 16, p. 190109, Sep. 2019, doi: 10.5888/pcd16.190109.

[23] R. C. Ripan et al., "A Data-Driven Heart Disease Prediction Model Through K-Means Clustering-Based Anomaly Detection," SN Computer Science, vol. 2, no. 2, p. 112, Apr. 2021, doi: 10.1007/s42979-021-00518-7.