Advancing IoT Security: Integrative Machine Learning Models for Enhanced Intrusion Detection in Wireless Sensor Networks

Bhargavi Mopuru
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India
bhargaviphd83@gmail.com (corresponding author)

Yellamma Pachipala
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India
pachipala.yamuna@gmail.com

Received: 26 April 2024 | Revised: 5 May 2024 | Accepted: 11 May 2024

 DOI: https://doi.org/10.48084/etasr.7641

ABSTRACT

This paper introduces a breakthrough approach to enhancing intrusion detection capabilities within Wireless Sensor Networks (WSNs) by implementing the Enhanced Wireless Intrusion Detection System (EW-IDS). Leveraging a sophisticated blend of Machine Learning (ML) algorithms, including Principal Component Analysis (PCA) and Singular Value Decomposition (SVD), the proposed model effectively streamlines feature selection, resulting in a robust detection framework. Extensive evaluations demonstrate that EW-IDS not only achieves a high accuracy rate of 96%, but also consistently surpasses traditional models in precision, recall, and F1 Score metrics. These achievements underscore the model's superior ability to differentiate between benign and malicious network activities. The implementation of EW-IDS marks a significant advance in securing the Internet of Things (IoT) environments against a diverse range of cyber threats, enhancing both the security protocols and operational efficiency of WSNs. This study provides a novel intrusion detection solution and offers valuable insights into the application of ML in complex security settings.

Keywords-wireless sensor networks; intrusion detection systems; machine learning; PCA; enhanced security protocols; cyber rthreat detection; IoT

I. INTRODUCTION

In recent years, the Internet of Things (IoT) has transformed countless industries by connecting billions of devices and enabling them to communicate and collaborate in unprecedented ways. Among the most critical components of this technological revolution are the Wireless Sensor Networks (WSNs), which are foundational to applications ranging from environmental monitoring to smart grid management. However, the rapid expansion and increasing complexity of these networks have also heightened their susceptibility to cyber threats. These threats compromise the integrity and availability of data and most importantly they pose significant risks to physical infrastructure and personal safety. Traditional Intrusion Detection Systems (IDSs) for WSNs have primarily focused on basic threat detection strategies that are quickly becoming obsolete in the face of sophisticated cyber-attacks, such as network breaches, malware, and advanced persistent threats [1]. These conventional systems often struggle with high false positive rates, poor adaptability to new threats, and inefficient processing capabilities, which are exacerbated by the voluminous and high-dimensional data generated by modern IoT devices. Moreover, the static nature of traditional IDS algorithms makes them ill-suited to the dynamic environments in which contemporary WSNs operate. Recognizing these limitations, this research proposes EW-IDS, a novel framework that leverages cutting-edge Machine Learning (ML) techniques to revolutionize the detection and prevention of intrusions in WSNs. By incorporating PCA and SVD, EW-IDS ameliorates the accuracy of threat detection and additionally optimizes the process of feature selection. This optimization reduces the computational overhead and improves the scalability of intrusion detection systems, enabling them to efficiently handle the vast datasets typical of IoT environments [2].

The primary objective of this study is to evaluate the performance of EW-IDS under various operational scenarios and compare its effectiveness with traditional IDS solutions. Through rigorous testing using the comprehensive WSN-DS dataset, this paper aims to demonstrate that EW-IDS can

achieve superior performance metrics. By doing so, it seeks to establish a new standard for IDS in WSNs that is robust against evolving cyber threats, scalable for extensive IoT applications, and efficient in processing and response. In addition to its practical applications, this research also contributes to the theoretical foundations of cyber security in IoT networks. By exploring the interplay between ML algorithms and feature selection techniques, it provides valuable insights into the development of adaptive security mechanisms that are capable of withstanding the complexities of modern cyber-physical systems. This work not only advances the field of WSN security, but also supports the sustainable growth of IoT technologies in a secure and reliable manner.

Authors in [3] conducted a comprehensive literature review on intrusion detection using ML techniques, focusing on single, hybrid, and ensemble classifiers. Their study compared related works based on classifier design, datasets used, and experimental setups. They highlighted the increasing popularity of hybrid classifiers, particularly integrated-based hybrid methods, and the limited consideration of ensemble classifiers in recent years. They also discussed the need for future research in areas, like baseline classifier selection, architecture of multiple classifiers, and feature selection to enhance IDSs. The review provided a valuable understanding of the current state of IDSs and identified potential research directions for further advancement in the field. Authors in [4] demonstrated the superior performance of the FA-ML technique in accurately identifying intrusions in WSN-IoT systems, achieving a maximum accuracy of 99.34% and outperforming existing models like KNN-PSO and XGBoost. The authors also explored potential future research directions, namely integrating semi-supervised or unsupervised learning methods and addressing specific types of intrusions like sybil attacks and routing attacks to further enhance the FA-ML technique. Authors in [5] proposed Apollon, a robust defense system against Adversarial Machine Learning (AML) attacks. In [6], the suggested system utilizes the Multi-Armed Bandits (MAB) algorithm with Thompson sampling to dynamically select the optimal classifier for each network traffic request, effectively preventing attackers from learning the IDS behavior and generating adversarial traffic. The MAB algorithm balances the trade-off between exploiting classifiers with the highest expected accuracy and exploring new classifiers that may yield higher accuracy in the future. This approach enhances the system's responsiveness to evolving security threats. Authors in [7] conducted a comprehensive literature review on the employment of ensemble learning strategies for improving the performance of autonomous learning models, particularly in the context of IDSs. The study stresses the significance of learning algorithms and data quality in determining the effectiveness of intrusion detection approaches. The DT-EnSVM intrusion detection framework, which combines ensemble learning and data transformation techniques, was introduced and its superiority in terms of precision, detection rate, false alarm rate, and learning speediness was demonstrated. The methodology followed involves transforming original characteristics to marginal density logarithmic thresholds to create new and enhanced training data, and then training base learners using Support Vector Machine (SVM) models. The study also highlights the advantages of ensemble learning and feature augmentation in creating a robust intrusion detection framework, while acknowledging the limitations of traditional methods of automated learning. Additionally, the importance of effective representations and dimensionality reduction in maximizing the classification results for single- and multiple- classification algorithms is examined. Authors in [8] conducted a comprehensive literature review on the use of machine learning and deep learning approaches for cyber event forecasting. The study focused on time series-based techniques, involving autoregressive time series and neural network models, as well as other forecasting methods, such as ARIMA, linear regression, SMOreg, Gaussian process, and multilayer perceptron. In [9], the authors highlighted the advantages of time series-based techniques for forecasting future events, particularly in the context of cyber attacks. These methods offer the ability to detect vulnerabilities in internet browsers and address adversarial attacks. However, the review also noted that no single technique outperforms others in time series-based anomaly detection, and the accuracy of machine learning and deep learning models is contingent on the quality of the datasets used. The study emphasized the importance of forecasting cyber events as a complement to intrusion detection, aiming to provide proactive strategies for mitigating potential cyber threats. Authors in [10] proposed a novel IDS for unmanned aerial vehicles (UAVs) deploying an Artificial Neural Network (ANN) and a Genetic Algorithm (GA) with a new dimensionality reduction technique. The method involves utilizing the Pearson's correlation coefficient and information gain for dimensional reduction, and the GA to optimize the weights of the ANN. The advantages of this approach include increased prediction accuracy and time efficiency. However, the study does not focus on image or video recognition, limiting its applicability in those areas. The network is systematically organized into distinct clusters, each of which symbolizes a specialized domain and is equipped with a distinct collection of sensor nodes and communication devices within this sophisticated IoT security architecture. These clusters function as complexly interwoven strands within a broader interconnected network, thereby enhancing the overall resilience of the proposed IoT ecosystem. Central to this architectural design is a wireless switch, which serves as a critical component enabling uninterrupted communication among the discrete clusters. By coordinating the movement of data, this intelligent switch guarantees a seamless interchange of information across the various clusters. By serving as a pivotal node, it optimizes the operation and unifies the entirety of the network.

II. THE PROPOSED METHOD

The Enhanced Wireless Intrusion Detection Algorithm (EWIDA) is a comprehensive and precise methodology that has been specifically developed to substantially improve the capabilities of intrusion detection in WSNs. The algorithm operates by executing a sequence of precisely delineated stages, with each stage being considered significant in determining its overall efficacy. Figure 1 illustrates the suggested framework.

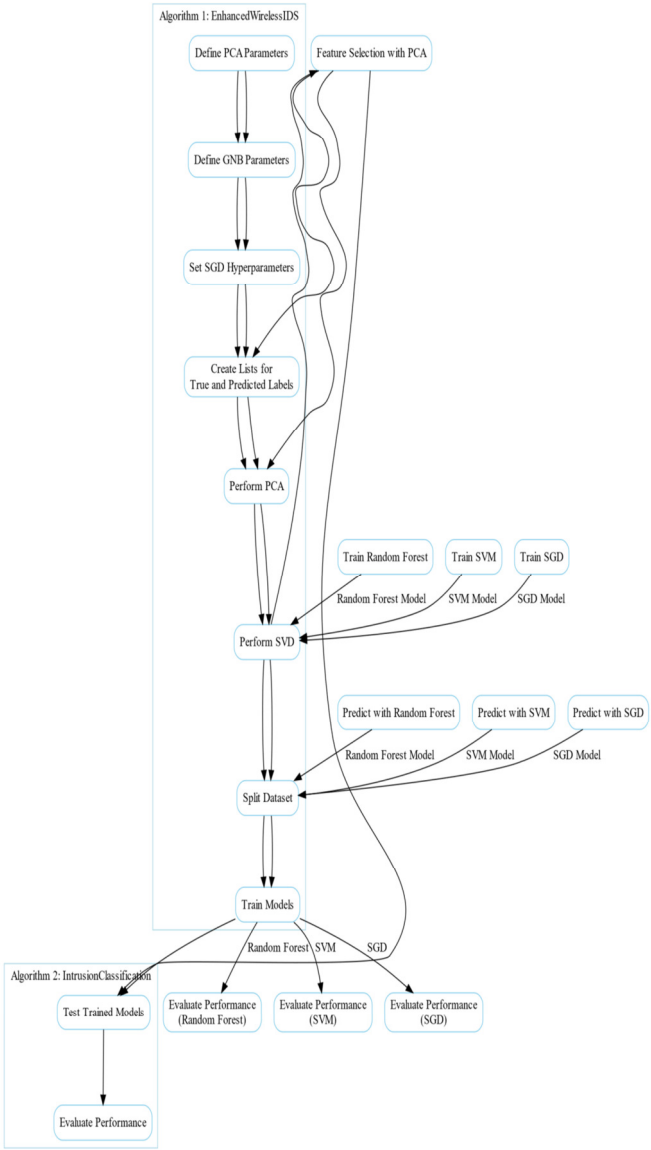

Fig. 1. Architectural framework.

When initiating the algorithm there are parameters that are pertinent to the process of feature selection. An important aspect to consider is the feature selection method, which may include the specification of the number of features (num_features) to be evaluated. After the selection of features, the algorithm proceeds with the initialization of parameters for three separate machine learning models, namely Stochastic Gradient Descent (SGD), SVM, and Random Forest (RF). SGD utilizes stochastic gradient descent optimization, while RF employs an ensemble of decision trees, and SVM implements support vector-based classification. An aspect of the algorithm that sets it apart is its implementation of sophisticated feature selection methods, such as PCA. The latter is implemented in order to effectively decrease the dimensionality of the dataset without compromising critical information. By prioritizing the most informative features, this stage optimizes the algorithm's performance and consequently enhances its capability to identify nuanced patterns that may indicate intrusions. The dataset is subsequently divided into training and testing sets, with training receiving 70% of the resources and testing 30%. This partition facilitates the autonomous training of ML models on a specific subset while permitting the assessment of their efficacy on the other. Once the models have been trained, they are applied to the testing set by the algorithm. Feature selection with PCA is performed for every data point in the testing set, and forecasts are generated utilizing the three models. Labels that are anticipated are retained for subsequent assessment. The algorithm completes a rigorous performance assessment phase. An assortment of assessment metrics, involving Accuracy, Precision, Recall, and F1 Score, are calculated for each of the three models in comparison to the actual predicted labels. This exhaustive assessment offers a nuanced comprehension of the algorithm's capability to differentiate between benign and malicious conduct.

The Intrusion Classification Algorithm (ECA) is a methodical framework that combines PCA, Gaussian Naive Bayes (GNB), and SGD to identify and categorize network intrusions. The algorithm commences its operation by establishing parameters for PCA, which involve specifying the quantity of components required to efficiently reduce the dimensionality of the dataset. Following this, the parameters for GNB are initialized, which consist of probabilities, mean vectors, and covariance matrices. Furthermore, hyperparameters are established for SGD, including the learning rate and the number of iterations. For evaluation purposes, genuine labels and predicted labels are stored in lists. Following this, the algorithm advances through the data preprocessing stages. The dataset is subjected to PCA, which decreases its dimensionality to the predetermined number of components. The SVD is subsequently executed to facilitate additional feature selection and data preprocessing. The resulting set is divided into two subsets: assessment (30%) and training (70%) sets. The training phase comprises two essential stages. To begin, the GNB model undergoes training for every class in the dataset. This involves determining the probability for each class and calculating the mean and covariance matrix. Subsequently, the SGD model is trained iteratively, wherein the training data is randomized, and model parameters are modified in accordance with the computed gradient. During the testing phase, the trained models are assessed using the testing set. PCA and SVD transformations are implemented on every data point in the testing set, and forecasts are generated utilizing the GNB and SGD models. Using its method, the GNB model calculates class probabilities and designates as the predicted label the class with the highest probability. Classification by the SGD model is determined by the decision boundary that is constructed using SGD-trained parameters.

Intrusion Detection Systems for Wireless Sensor Networks (WSN-IDS) are crucial for protecting the security and integrity of WSNs. These systems have been purposefully engineered to identify and thwart potential security breaches and threats that may occur in wireless sensor environments. In light of the limited resources available to sensor nodes, WSN-IDS implement specialized methodologies including anomaly detection, signature-based detection, and ML algorithms that

are customized to suit the unique attributes of WSNs. Through the ongoing surveillance of network operations, examination of communication patterns, and identification of irregular conduct, WSN-IDS plays a pivotal role in the timely identification and averment of a multitude of threats, such as physical interference, denial of service, and sybil attacks, among others. Ensuring the dependability and effectiveness of WSNs in a wide range of applications, including environmental monitoring, healthcare, and industrial automation, these systems are indispensable. The increasing prevalence of the IoT devices underscores the growing importance of resilient WSN-IDS in safeguarding connected sensor ecosystems. Table I depicts the different attack types in the WSN-DS dataset.

TABLE I. ATTACK TYPES IN THE WSN-DS DATASET

Attack Class	Description
Denial of Service (DoS)	Overwhelming the network with excessive traffic or depleting the energy resources of sensor nodes.
Physical Attacks	Tampering, destruction, or compromise of physical components of sensor nodes.
Sybil Attacks	A compromised node claiming multiple identities, leading to various malicious activities.
Routing Attacks	Disrupting or manipulating communication paths in the network, leading to data loss or unauthorized access.
Eavesdropping Attacks	Unauthorized interception of communication between sensor nodes.
Traffic Analysis Attacks	Analyzing network traffic patterns to gain insights into the behavior of the WSN.
Hello Flood Attacks	Flooding the network with a large number of "hello" messages, disrupting normal operations.
Selective Forwarding Attacks	Compromised nodes selectively forwarding or dropping packets, affecting data delivery reliability.
Gray Hole Attacks	Nodes selectively forwarding packets but dropping certain portions of the data.
Wormhole Attacks	Creating a tunnel between two distant points, potentially allowing attackers to inject false data.

III. DATA PRE-PROCESSING

During the preliminary stages of data preprocessing for the WSN dataset, managing the label attribute is a significant undertaking. This attribute, which was initially composed of alphabetic characters, signifies the diverse attacks that the network has encountered. In order to improve the compatibility of the dataset with analytical methodologies, an essential conversion is executed: the alphabetic identifiers are transformed into numeric values. When examining the WSN-IDS dataset, one encounters a complex assortment of attack types that present numerous obstacles to the network's security and functionality. The dataset comprises a variety of attack types, such as DoS maneuvers, which deplete the energy resources of sensor nodes or inundate the network with excessive traffic. Physical assaults pose a concrete danger as they involve the manipulation, destruction, or compromise of physical elements comprising sensor nodes. Sybil attacks introduce intricacy through the participation of compromised nodes that assume multiple identities, thereby instigating an array of malevolent activities.

The disruption or manipulation of network communication paths by routing attacks may result in unauthorized access or data loss. Eavesdropping assaults pertain to the unauthorized interception of sensor node communication, thereby compromising the confidentiality of the transmitted data. Additional intricacies emerge in the form of traffic analysis attacks, wherein adversaries scrutinize patterns of network traffic to obtain knowledge regarding the actions of the WSN. Network operations are disrupted by Hello Flood attacks, which inundate the system with an excessive volume of "hello" messages. Selective Forwarding Assaults transpire when compromised nodes forward or discard packets in a particular manner, thereby compromising the dependability of data delivery. Gray Hole Attacks, which involve nodes judiciously forwarding packets while discarding specific segments of the data, present nuanced yet consequential risks. By constructing tunnels between remote locations, Wormhole Attacks may enable malicious actors to inject erroneous data into a network. Effectively handling these varied threats necessitates a rigorous preprocessing strategy that transforms qualitative data into quantitative values and establishes the foundation for resilient intrusion detection methodologies. The WSN-IDS dataset accurately represents the intricacies of the real world. By thoroughly addressing these various assaults, one can establish a wireless sensor network environment that is both resilient and secure.

PCA is a highly effective technique utilized in the fields of ML, data science, and signal processing for dimensionality reduction. The former technique transforms a dataset comprising numerous interrelated variables into a distinct set of uncorrelated elements referred to as principal components. Capturing the most data variance allows a concise depiction without losing much information. The PCA process has several critical steps. First, the dataset's mean is calculated and subtracted from each data point to center the data. Calculating the covariance matrix, which shows variable relationships, follows. Then, the covariance matrix eigenvectors and eigenvalues are calculated. Eigenvalues measure variance along maximal variance axes, while eigenvectors demonstrate them. After that, the eigenvectors are ranked by eigenvalue and the highest ones are kept getting the principal components. These pieces provide a new dataset foundation for a lower-dimensional representation.

Regarded as highly acclaimed and versatile ML algorithms, SVMs are widely used in classification and regression tasks. The SVM has shown great success on complex datasets and adaptability across various fields. SVM finds the best hyperplane to divide the feature space into classes. It uses kernels to handle high-dimensional data and nonlinear relationships. The algorithm finds the hyperplane with the biggest margin, which is the distance between it and the data points from each class nearest to it. This strategy improves model overfitting resistance and resilient generalization to unseen data. The SVM algorithm excels at linear and nonlinear decision boundaries. When data cannot be separated linearly, SVMs use kernel functions to transform the basic feature space into a higher-dimensional space where a hyperplane may efficiently differentiate classes. Sigmoid, RBF, and polynomial kernel functions are common.

SGD is a popular and effective optimization technique in ML and deep learning. Being an iterative and stochastic

gradient descent variation, SGD reduces the cost or loss function of model training efficiently. It handles large datasets and complex models well due to its scalability and versatility. SGD optimizes model parameters by iteratively adjusting them based on cost function gradients. The algorithm's "stochastic" quality comes from using a randomly picked mini batch of training data at each iteration. This randomization accelerates convergence and may help prevent local minima by causing disturbance. SGD is particularly competent in large datasets where processing the complete dataset in each iteration would be computationally expensive. SGD balances computing efficiency and well- founded generalization to the entire dataset by altering parameters for a subset of data. Randomness through mini-batch selection may make convergence paths harder to anticipate than the conventional gradient descent. The learning rate and mini-batch size must be precisely adjusted to counteract this.

SVD is widely deployed in ML, signal processing, and linear algebra. SVD decomposes a matrix into three extra matrices to reveal its structure and attributes. The SVD function converts a matrix A into a product of three matrices: U, V^T, and Σ, where H represents the singular values of A and U and V are orthogonal. Singular values indicate the relevance of each dataset mode of variation. SVD reduces dimensionality and extracts features well in data analysis. By retaining only, the most important singular values and associated vectors, data can be represented in a smaller space without losing information. The data's intrinsic structure is disclosed and their compute performance is improved by this approach. SVD is essential for collaborative filtering in recommendation systems and signal processing noise reduction. SVD also supports PCA, a popular tool for discovering dominating patterns and lowering dataset complexity.

GNB is a probabilistic Naive Bayes classifier. It employs Bayesian probability. It excels at classification tasks using continuous Gaussian (normal) features. Bayes' theorem underpins naive Bayes classifiers like GNB. In light of fresh evidence, Bayes' theorem lets this study adjust its hypothesis probability estimates. It calculates the likelihood of a class given observable features in GNB. GNB is "naive" since it assumes feature independence, which states that one feature does not affect another. While this assumption decreases computational complexity, its applicability to real-world settings may be limited. Despite repeated violations of this assumption, GNB often performs effectively. GNB represents the Gaussian probability distribution of every feature for every class. For continuous features, GNB calculates feature mean and standard deviation within classes. When a new data point is added, the algorithm determines its probability deploying each class's Gaussian distribution. GNB adopts Bayes' theorem to combine likelihoods with prior probabilities (i.e. class probabilities without feature considerations) to classify. After considering the observed features, the data point is assigned to the class with the highest posterior probability.GNB is useful for small datasets and resilient to extraneous features. It has worked in medical diagnosis, sentiment analysis, and email spam detection where the feature distribution across classes is Gaussian.

IV. EXPERIMENTAL SETUP

A thorough experimental setup was constructed with the intention of assessing the effectiveness of the proposed intrusion detection model in a WSN environment. The fundamental basis of the performed investigation is the application of the WSN-DS dataset, an extensive and varied compilation of data that encompasses a multitude of facets pertaining to sensor node behavior as well as potential security risks. Its diversity enables a comprehensive assessment of the capabilities of the suggested model in the face of various threat scenarios.

In order to guarantee a comprehensive assessment, this work partitioned the dataset into separate training and testing sets, reserving 30% of the records for testing purposes and 70% for training. By implementing this partitioning strategy, the proposed model is able to engage in an extensive learning process on a significant portion of the dataset. Simultaneously, this permits an impartial evaluation of the model's ability to generalize to previously unseen data.

Accuracy has been selected as the evaluation metric to quantify the performance of the model. This metric quantifies the ratio of instances that were accurately classified to the overall number of instances in the testing set. Accuracy is a dependable metric, particularly when applied to intrusion detection, where the precise differentiation between benign and malevolent activities is critical. Through the computation of accuracy, valuable insights are obtained regarding the general efficacy of the recommended model in differentiating benign activities from potentially hazardous ones within the WSN. The performance measures are [9]:

$$\text{Recall} = \frac{\text{TP}}{(\text{TP}+\text{FN})} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP}+\text{FP})} \quad (2)$$

$$\text{F1 Score} = \frac{2 * \text{Precision} + \text{Recall}}{(\text{Precision} - \text{Recall})} \quad (3)$$

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{(\text{TP}+\text{FN} +\text{FP} + \text{FN})} \quad (4)$$

where TP represents True Positives, TN represents True Negatives, FP represents False Positives, and FN represents False Negatives.

V. RESULTS AND DISCUSSIONS

Conducting an exhaustive research evaluation, this study initiated a sequence of carefully planned experiments to evaluate the effectiveness of the proposed EW-IDS. The principal aim of this research was to assess the influence of feature extraction on the performance of EW-IDS, with a specific focus on the utilization of PCA and SVD. The investigations were carried out using feature dimensions of 10, and 15, to determine the most effective configuration for intrusion detection. Expanding its focus beyond feature extraction, this work aimed to conduct a comparative evaluation of EW-IDS in relation to alternative classification strategies based on ML in order to assess the former's efficacy in comparison to established methods, including SGD and GNB. This comparative analysis' objective was to highlight the

unique benefits and resilience of EW-IDS when compared to traditional methods, underscoring its potential as an innovative intrusion detection system. Moreover, the conducted investigation broadened to include a thorough assessment of the robustness of EW-IDS against a variety of network threats. EW-IDS was subjected to realistic threat scenarios through the meticulous simulation of four distinct categories of attacks: DoS Attacks, Sybil Attacks, Routing Attacks, and Physical Attacks. Through this comprehensive assessment, the efficacy of EW-IDS in detecting and preventing diverse forms of network intrusions was examined.

Great emphasis was placed on the criticality of precision and recall metrics when evaluating the integration potential and usability of EW-IDS's IDS. These metrics function as crucial indicators of the system's capability to detect and categorize normal and malicious network activity with precision. In addition to comparing EW-IDS to conventional classification techniques, such as GNB and SGD, this study also examined the former in the context of advanced models including Deep Neural Networks (DNNs) and Deep Convolutional Neural Networks (Deep CNNs). By adopting this comprehensive methodology, a nuanced comprehension of the performance of EW-IDS in various scenarios was achieved, solidifying its status as an advanced and versatile solution for intrusion detection in WSNs.

TABLE II. PERFORMANCE METRICS OF INTRUSION DETECTION ALGORITHMS

Algorithm	Accuracy	Precision	Recall	F1 Score
DNN	0.80	0.78	o.82	0.80
Deep CNN	0.82	0.8	0.84	0.82
EW-IDS GNB	0.95	0.94	0.96	0.95
EW-IDS SGD	0.96	0.95	0.97	0.96

The competence of each approach is effectively evaluated through a comparative analysis of the intrusion detection algorithms implemented on the WSN-DS dataset without a feature selection phase, as indicated by the performance metrics. The DNN exhibits a noteworthy accuracy of 87%, accompanied by consistently high values of precision, recall, and F1 score. Table II conducts a comparative analysis of the ML models against intrusion detection algorithms in the context of WSNs. The presented results are derived from various algorithms, encompassing conventional methods, such as CNN and DNN, in addition to the suggested EW-IDS, which utilizes GNB and SGD phases.

The assessment of ML algorithms holds significant importance in the field of intrusion detection. The outcomes derived from three classifiers, i.e. RF, SGD, and GNB, provide thought-provoking observations. The RF classifier is distinguished by its remarkable accuracy value of 0.98. The algorithm exhibits exceptional precision, recall, and F1-score values across various classes, thereby showcasing its resilience in discerning and categorizing occurrences of intrusions. Significantly, it attains an impeccable precision of 1.00 for the majority class (class 3), thereby demonstrating its adeptness in managing the commonplace normal instances. On the other hand, the accuracy of the SGD classifier is 0.91. Although it achieves a moderate level of overall accuracy, it encounters difficulties in several courses regarding precision and recall. The difficulties become especially apparent in class 0, as precision and recall are essentially diminished, signifying the potential for challenges in accurately classifying instances. However, it manifests exceptional performance in class 3, attaining a remarkable precision of 0.91 and a flawless recall.

The performance of the GNB classifier is 85%. Although the model achieves praiseworthy precision and recall rates for class 3, it faces difficulties in accurately distinguishing instances belonging to other classes. The trade-off between precision and recall is clearly observed in class 0, where the classifier attains a substantial recall but sacrifices precision. This observation implies a propensity to incorrectly designate a greater number of instances as class 0, which may result in the generation of FP. The findings highlight the diverse capabilities and drawbacks of each classifier with respect to intrusion detection. RF's flawless accuracy establishes it as a formidable option, whereas SGD and GNB exhibit nuanced performances characterized by unique compromises between recall and precision across distinct classes. The exhaustive classification reports furnish an elaborate analysis, thereby providing valuable insights for the purpose of selecting the most appropriate algorithm in an IDS, taking into consideration particular requirements and priorities. Table III portrays the results of the proposed method versus those of GNB.

TABLE III. RESULT COMPARISON

Proposed Model Result Accuracy: 0.98 Classification Report				
Episode	**Precision**	**Recall**	**F1-score**	**Support**
1	0.99	0.99	0.99	2043
2	0.94	0.98	0.96	631
3	0.99	0.99	0.99	2985
4	1.00	1.00	1.00	67965
5	1.00	0.93	0.96	1309
Average			0.98	74933
Macro Avg	0.98	0.98	0.98	74933
Weighted avg	1.00	1.00	1.00	74933
GNB Result Accuracy: 0.85 Classification Report				
Episode	**Precision**	**Recall**	**F1-score**	**Support**
1	0.56	0.93	0.70	2043
2	0.23	0.87	0.36	631
3	0.18	0.52	0.26	2985
4	0.99	0.87	0.93	67965
5	0.84	0.37	0.52	1309
Average			0.85	74933
Macro Avg	0.56	0.71	0.55	74933
Weighted avg	0.94	0.85	0.88	74933

By selecting features from a sample set, the feature selection module enables to reduce the number of dimensions in a dataset or improve the quality of the estimators on high-dimensional datasets. Both PCA and SVD, the two techniques utilized in this investigation, employed 10 to 15 characteristics. The graph in Figure 2 provides an extensive comparison of classification metrics for different IDAs in the context of a WSN. The algorithms are represented along the x-axis, whereas the performance metrics are displayed along the y-axis.

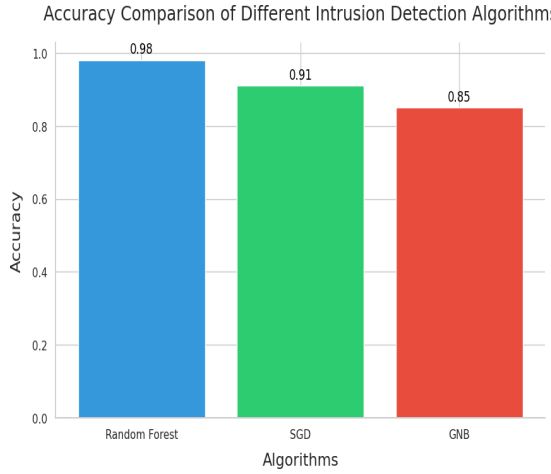

Fig. 2. Algorithm accuracy comparison.

The values displayed in Figure 4 above each bar offer a comprehensive summary of the numerical results for every algorithm in relation to the performance metrics. It is worth noting that the Without FS GNB and Without FS SGD algorithms demonstrate comparatively diminished values, which indicate their efficacy in the absence of a feature selection phase. Transitioning to algorithms that integrate feature selection, PCA 10 GNB and PCA 10 SGD demonstrate enhanced metrics, which serve as evidence of the beneficial influence that feature selection has on the accuracy of intrusion detection. Likewise, algorithms that make use of SVD 10 GNB and SVD 10 SGD demonstrate outcomes that are comparable. SVD and PCA algorithms, including PCA 15 GNB, PCA 15 SGD, SVD 15 GNB, and SVD 15 SGD, exhibit enhanced efficacy across all metrics with the addition of 15 features. The proposed EW-IDS algorithm exhibits superior performance across all metrics. This implies that the proposed model not only effectively utilizes feature selection, but also demonstrates exceptional performance in robust intrusion detection within a WSN.

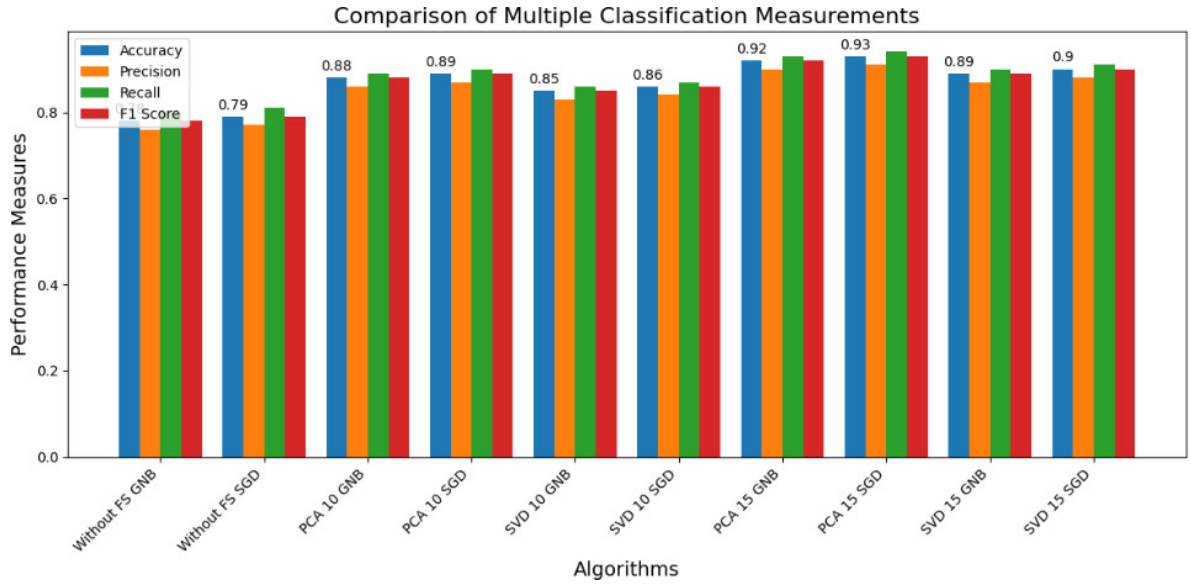

Fig. 3. Performance Measure Comparison

The objective of this study was to evaluate the efficacy of intrusion detection algorithms in identifying and categorizing intrusions within a WSN through the comprehensive analysis of these algorithms using a WSN-DS dataset. The results offer a comprehensive analysis of the effectiveness of each algorithm in various scenarios, emphasizing their capability to identify both benign and malevolent activities. The robust performance of algorithms that operated without feature selection, specifically, Without FS GNB and Without FS SGD, was demonstrated by their accuracy ranging from 0.85 to 0.86. Algorithmic performance was consistently enhanced with the implementation of feature selection techniques, such as PCA and SVD, which utilized 10 and 15 components, respectively. It is worth noting that the EW-IDS achieved the highest accuracy of 0.96. This highlights the effectiveness of EW-IDS in differentiating between benign and malicious instances, establishing it as a resilient solution for bolstering security in WSNs. The findings confirm the capability of the proposed framework, which combines sophisticated feature selection and classification methods, to substantially enhance the detection of intrusions within WSNs.

VI. CONCLUSION

In this study, a WSN-DS dataset was utilized to methodically analyze the effectiveness of various intrusion detection algorithms within Wireless Sensor Networks (WSNs). Multiple attack types, including Denial of Service (DoS), Physical, Sybil, Routing, Eavesdropping, Traffic Analysis, Hello Flood, Selective Forwarding, Gray Hole, and Wormhole attacks, were meticulously investigated. The research carried out rigorously evaluated a suite of neural network models, specifically Gaussian Naive Bayes (GNB),

Stochastic Gradient Descent (SGD), Deep Neural Networks (DNNs), and Deep Convolutional Neural Network (Deep CNNs). The experimental results clearly indicated that the proposed Enhanced Wireless Intrusion Detection System (EW-IDS) consistently outperformed other evaluated models, achieving an accuracy of 0.96, which was significantly higher than DNN (0.87), Deep CNN (0.89), GNB (0.85), and SGD (0.91). Precision, recall, and F1 Score of EW-IDS all exceeded 0.94, confirming the robustness of the proposed model. Crucially, the implementation of advanced feature selection techniques, including Singular Value Decomposition (SVD) and Principal Component Analysis (PCA), was found to substantially enhance the discriminative capability of the detection algorithms deployed. The integration of these sophisticated feature selection mechanisms, along with the employment of robust machine learning classifiers, proved indispensable in elevating the accuracy and reliability of the detection system. This enhancement is particularly critical in the current landscape of evolving and complex cyber threats. EW-IDS's superior performance and resilience highlight its potential as a foundational component in strengthening the security architecture of WSNs. This research not only advances the state-of-the-art in WSN intrusion detection, but also provides a valuable framework for future explorations and developments in securing IoT and WSN infrastructures against increasingly sophisticated cyber threats. The insights derived from this study serve as a pivotal reference point for the enhancement of security protocols within the broader ecosystem of connected devices.

REFERENCES

[1] R. Doriguzzi-Corin, L. A. D. Knob, L. Mendozzi, D. Siracusa, and M. Savi, "Introducing Packet-Level Analysis in Programmable Data Planes to Advance Network Intrusion Detection." arXiv, Jan. 04, 2024, https://doi.org/10.48550/arXiv.2307.05936.

[2] M. A. Talukder *et al.*, "Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction." arXiv, Jan. 22, 2024, https://doi.org/10.48550/arXiv.2401.12262.

[3] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: A review," *Expert Systems with Applications*, vol. 36, no. 10, pp. 11994–12000, Dec. 2009, https://doi.org/10.1016/j.eswa.2009.05.029.

[4] M. Karthikeyan, D. Manimegalai, and K. RajaGopal, "Firefly algorithm based WSN-IoT security enhancement with machine learning for intrusion detection," *Scientific Reports*, vol. 14, no. 1, Jan. 2024, Art. no. 231, https://doi.org/10.1038/s41598-023-50554-x.

[5] A. Paya, S. Arroni, V. García-Díaz, and A. Gómez, "Apollon: A robust defense system against Adversarial Machine Learning attacks in Intrusion Detection Systems," *Computers & Security*, vol. 136, Jan. 2024, Art. no. 103546, https://doi.org/10.1016/j.cose.2023.103546.

[6] A. M. S. Saleh, "A Power-Aware Method for IoT Networks with Mobile Stations and Dynamic Power Management Strategy," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12108–12114, Dec. 2023, https://doi.org/10.48084/etasr.6352.

[7] C. SaiTeja and J. B. Seventline, "A hybrid learning framework for multi-modal facial prediction and recognition using improvised non-linear SVM classifier," *AIP Advances*, vol. 13, no. 2, Feb. 2023, Art. no. 025316, https://doi.org/10.1063/5.0136623.

[8] Y. Ahmed, M. A. Azad, and T. Asyhari, "Rapid Forecasting of Cyber Events Using Machine Learning-Enabled Features," *Information*, vol. 15, no. 1, Jan. 2024, Art. no. 36, https://doi.org/10.3390/info15010036.

[9] S. Zafar, G. Miraj, R. Baloch, D. Murtaza, and K. Arshad, "An IoT Based Real-Time Environmental Monitoring System Using Arduino and Cloud Service," *Engineering, Technology & Applied Science Research*, vol. 8, no. 4, pp. 3238–3242, Aug. 2018, https://doi.org/10.48084/etasr.2144.

[10] S. Chopparapu and J. B. Seventline, "An Efficient Multi-modal Facial Gesture-based Ensemble Classification and Reaction to Sound Framework for Large Video Sequences," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11263–11270, Aug. 2023, https://doi.org/10.48084/etasr.6087.