Abstract

The COVID-19 pandemic, widely acknowledged as the most significant global health crisis of this century, has impacted virtually every sector, including software development. This unprecedented situation has led to significant implications for software projects and the professionals involved in software development, presenting substantial challenges. The primary objective of this study is to systematically analyze the evolving research trends within the Global Software Engineering (GSE) field. This study focuses on examining whether the global pandemic has led to an increased emphasis on software design research. Furthermore, we investigate the existing gap in social interaction during the software design phase of development. The research methodology consists of a two-phase systematic analysis of the existing literature. In the first phase involves the mapping of GSE research conducted over the two decades preceding the pandemic (2000-2020). In the second phase, this study employed a forward snowballing approach to review literature related to the software design phase published between 2020 and 2022. The analysis of 592 research studies in these two phases reveals various trends in GSE research. Evaluation research stands out as the most extensively explored research type across methods, processes, and human aspects of development. Despite the considerable impact of the COVID-19 pandemic, which led to an increased reliance on distributed teams, our findings suggest that, while software organizations have been extensively studied across all software engineering phases, the software design phase remains one of the least explored areas. Our contribution envisions a more collaborative and adaptable GSE field, providing guidance for future research endeavors aimed at supporting distributed teams.
Introduction

Wireless Sensor Networks (WSNs) [1, 2] is also termed as a heterogeneous system designed with the small controllers, sensors and generic processing components. Also, it is made up of thousands or hundreds of low-cost, self-organizing, wireless nodes that are used to monitor and regulate the environment. When creating a WSN, self-healing, dependability, adaptability, robustness, and security are the five primary factors that must be taken into account [3, 4]. It can also be used for a variety of military purposes, as well as for the monitoring of ocean, manufacturing equipment, earthquakes, and other natural disasters. In addition, it's likely that future applications may incorporate WSN concepts in their architectures, including those that monitor environment, transportation, site security, fires, and water quality. In this network [5], there may be one or more base stations, which are centralized control units. Then, a base station often serves as a gateway to another network, as well as a great data processing and storage facility and access point for human interaction. In order to retrieve data from the network and disseminate control information, it can also be utilized as a connector. It is imperative to guarantee a high level of security [6, 7] for the critical WSN applications in order to protect their data and infrastructure from breaches. In order to identify unusual activities and breaches, an Intrusion Detection System (IDS) [8] should be deployed. Moreover, it is a crucial component of security across any network type, since it provides the network with a high level of protection against potential dangers by stopping or identifying all intrusions and hosts. Its main objective is to make sure that every new attack may be detected. It is categorized into the types of misuse IDS and anomaly IDS [9], in which the anomaly IDS analyses statistical patterns and sophisticated ways to determine whether the behavior is healthy or not, whereas the misuse IDS uses signatures to find any new attacks. These techniques mostly involve intelligent classification methods and artificial intelligence algorithms. The pattern, detection rate, false alarm rate, and accuracy of each classifier can all be used to describe them. Moreover, various machine learning and deep learning based classification approaches are mainly used to accurately spot the intrusions in the network. For instance [10-13], the Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DR), Naïve Bayes (NB), and Artificial Neural Network (ANN) are the most popular and standard machine learning approaches used for developing an IDS framework. However, the traditional frameworks facing some complications during the detection of network intrusions, which includes the followings: increased false alarm rate, complexity in system modeling, reduced detection accuracy, and high dimensionality of features. Therefore, this paper motivates to construct a novel and intelligent IDS framework for ensuring the security and confidentiality of WSN. The original contributions of this paper are as follows:

- To generate the balanced dataset for improving the process of intrusion detection and categorization, the data preparation is performed at first, which includes the operations of data cleaning, normalization, splitting, and clustering.
- To choose the optimal set of features for tuning the parameters of classifier, an Intensive Binary Pigeon Optimization (IBiPO) mechanism is deployed, which highly increases the accuracy and detection rate of classifier.
- To exactly predict the normal and attacking data flows by training and testing the optimized features, the Bi-directional Long Short Term Memory (Bi-LSTM) classification model is utilized.
- To assess the results of the proposed IBiPO + Bi-LSTM model, an extensive simulation analysis is carried out during performance evaluation.

The other sections of this paper are structured into the following units: Section 1 presents the complete related work of the existing intrusion detection methodologies used for protecting WSNs. Also, it investigates the advantages and disadvantages of each mechanism based on its working model and operating principles. Section 2 provides the clear explanation for the proposed AI based IDS framework with its overall workflow and illustrations. Section 3 validates the results of both existing and proposed anomaly detection methodologies by using various parameters and datasets. Finally, the overall paper is summarized with the findings and future scope in Section 4.
I. RELATED WORK

Paul, et al [17] implemented a neuro-fuzzy based IDS framework for improving the security of WSNs. The purpose of this work was to develop a lightweight security mechanism for protecting WSN against harmful networking attacks. Here, the centralized approach has been utilized to enable reliable and valid data exchange in networks. Moreover, the integration of WSN with IoT networks could be one of the most difficult tasks, due to many security challenges. Amouri, et al [18] introduced a cross-layered IDS approach for detecting anomalies in the WSN. Here, the Accumulated Measure of Fluctuation (AMoF) has been utilized to accurately classify the attacks in the network. Sarkunavathi, et al [19] presented a comprehensive analysis to examine the different types of machine learning and deep learning techniques used for developing an effective IDS framework. This paper objects to attaining an increased attack classification accuracy with reduced false positives. Typically, the IDS used in WSN is categorized into the following types:

- Anomaly detection IDS
- Misuse detection IDS
- Clustering based IDS
- Hybridized IDS
- Trust enabled IDS
- Zone-based IDS

Moreover, an efficient IDS framework should satisfy the following security parameters for ensuring better intrusion detection performance. It includes energy efficiency, accuracy, memory, and network topology. Salifadlin, et al [10] implemented an improved binary grey wolf optimization algorithm for constructing an effective IDS framework. This work mentioned that the feature selection was the most essential stage in the IDS, since it helps to obtain a high accuracy with reduced redundancy and maximized relevancy. Zhang, et al [11] implemented a Multi-Kernel Extreme Learning Machine (MK-ELM) model for strengthening the security of WSN against horrible intrusions. The original contribution of this work was to obtain an increased detection accuracy with ensured Quality of Service (QoS). This algorithm incorporates the operations of both ELM and multi-kernel SVM for increasing the robustness and detection accuracy. Alwan, et al [8] utilized a Slime Mould Algorithm (SMA) for developing a new IDS framework for WSN. Elsaid, et al [9] developed an optimized collaboration based IDS framework for increasing the security of WSNs. This paper intends to improve the robustness, detection rate, and reduce the false alarm rate of classification. Typically, the WSN is highly vulnerable to network intrusions and attacks, hence it must be protected for ensuring the security and reliability of the network. Hence, the IDS is one of the most suitable option for WSN security, which supports to spot the intrusions or unauthenticated activities in the network by analyzing the features of network and data. Pan, et al [14] deployed a lightweight and intelligent intrusion detection model for guaranteeing the privacy, security, and confidentiality of WSN. Here, the K-Nearest Neighbor (KNN) algorithm incorporated with the Sine Cosine Algorithm (SCA) was utilized to minimize the false alarm rate and increase the classification accuracy of this detection framework. Moreover, the alarm response generated by the IDS could be used to block the intrusions or attacks in the network. In addition to that, the Polymorphic Mutation Strategy (PM) has been utilized to choose the features for analyzing the characteristics of attacks. The advantages of this work are minimal computational complexity, ensured system robustness and reliability. However, the time required to train and test the data samples are increased, which degrades the efficacy of the suggested system.

Gowdhaman, et al [15] used a Deep Neural Network (DNN) approach to deal with the unbalanced attacks in the WSNs. In this case, the cross correlation has been used to effectively choose the pertinent features from the datasets for correctly identifying the intrusions. Sood, et al [16] utilized a conditional Generative Adversarial Network (GAN) model for protecting the WSN against the harmful network intrusions. This research work focused on an unsupervised learning method and how it may be used to create secure IDS. Also, it generated some fictitious data to mislead the attacker. In contrast to other deep learning based IDS models, it can secure the network and transport data between the sender and the recipient. However, it failed to prove the detection accuracy and QoS of the suggested model, which could be major limitation of this work. Masengo, et al [2] suggested an AI based anomaly detection model for an integrated Software Defined WSN (SDWSN) platform. The purpose of this paper was to analyze the efficiency and performance of various classification techniques such as DT, NB, and deep ANN for developing a computationally intelligent IDS framework for securing SDWSN. In order to analyze the efficacy of these models, the prediction time, run time, and memory size have been estimated in this work. Also, this study stated that the deep ANN model outperforms other techniques with improved performance values. Karthic, et al [18] introduced a hybrid optimized DNN for detecting intrusions in the WSNs. Here, the standard CNN model is incorporated with the LSTM framework for identifying and categorizing the class of intrusions. Moreover, an enhanced conditional random field based feature selection mechanism was also used to simplify the process of feature learning. Due to an efficient learning of features, the overall detection accuracy of the suggested framework was highly improved. However, it could be difficult to understand the system model, due to the complexity in computational operations.

Rezvi, et al [19] implemented a new data mining technique for developing an effective IDS framework. Here, the dataset preprocessing was performed to characterize the attack types into the discrete values. Then, the different types of classification models such as ANN, KNN, SVM, LR, and NB have been validated to choose the most efficient technique for an accurate intrusion detection and classification. In addition, the SMOTE analysis was performed to estimate the prediction rate of the suggested framework. Di Mauro, et al [29] suggested a Weightless Neural Network (WNN) for an effective IDS framework. Here, the attack types of were classified according to their features such as coarse grained features, flow based features, time based features, byte based features, packet based features, and flag based features. Yet, it required to minimize the classification time, which affects the performance of classifier. Halbouni, et al [30] utilized a CNN-LSTM classification technique for designing a competent IDS with increased accuracy and highest detection.
rate. This framework includes the operating stages of data encoding, normalization, optimization, and classification. The survey found that existing IDS frameworks are primarily concerned with increasing detection rates, lowering false positives, and boosting learning effectiveness.

However, it has the following issues:

- The features’ testing and training take a lot of time.
- Complicated feature extraction and selection processes.
- Oversampling.
- Lack of reliability.
- Difficult to implement.

Therefore, the goal of the proposed work is to create a new security paradigm that will shield WSN against damaging intrusions or abnormalities.

II. PROPOSED METHODOLOGY

This section provides the clear explanation for the proposed IDS framework with its overall workflow and illustrations. The original contribution of this paper is to develop a computationally intelligent IDS framework for securing WSN from network intrusions. For this purpose, a novel and efficient Intensive Binary Pigeon Optimization (IBiPO) technique incorporated with a Bi-directional Long Short-Term Memory (Bi-LSTM) models are implemented. The overall workflow of the proposed system is depicted in Figure 1, which includes the following operations:

- Data preparation
- Intensive Binary Pigeon Optimization (IBiPO)
- Bi-Directional LSTM (Bi-LSTM) classification
- Performance evaluation

![Figure 1 Work flow of the proposed system](Image)

Here, the popular and public IDS datasets are used for system implementation, which are preprocessed at the initial stage with the data cleaning, normalization, splitting, and clustering operations. Then, the balanced dataset is used for further optimization and classification processes. The IBiPO technique is mainly used to optimally select the most relevant features for accurately predicting the intrusions with reduced false alarms. Moreover, this technique helps to simplify the process of intrusion identification and classification with high accuracy. Then, the obtained features are passed to the Bi-LSTM classifier for training and testing. This classifier predicts the normal and anomalous data based on the training features. The advantages of the proposed IBiPO + Bi-LSTM model are reduced overfitting, false alarms, time consumption, and high detection accuracy.

\[ DS_N = \frac{(DS_{\max} - DS_{\min})}{DS_{\max} - DS_{\min}} \]  

(1)

The data is normalized by the formula listed in equation (1) where, \( DS \) indicates the IDS dataset, \( DS_N \) is the normalized dataset, \( DS_{\min} \) and \( DS_{\max} \) are the minimum and maximum values of dataset. Subsequently, the data splitting and clustering operations are also performed to generate the balanced dataset, where training dataset is split into two such as training and validation. In which, the training set is to train the entire model, and the validation set is used to test the model at the time of parameter tuning. Then, the distance-based clustering mechanism is applied to reduce the size of dataset. During this process, the random centroid is selected at first, and the data points are assigned to the nearest cluster according to its distance or similarity. Each cluster's centroid is calculated as the average of all the data points that belong to that cluster after all the data points have been assigned to the closest group. The procedure of assigning the data points to the new cluster's centroid is then repeated until the centroid's values remain consistent. Finally, the preprocessed dataset is generated and used for further operations.

A. Data Preparation

Before processing and evaluating the data, it is highly essential to prepare the balanced dataset. The data preparation holds the major operations of data cleaning, normalization, transformation, clustering, and compression. In which, the process of eliminating redundant or irrelevant entries and addressing missing data is known as data cleaning. It is a crucial step in making sure the data is reliable, accurate, and useable. Moreover, it is more important to eliminate the duplicate entries in the given dataset to keep the classifiers from learning rare records and from being biased toward the most common records. Moreover, the transformation of symbolic data into numerical values and label transfer are considered as the additional data cleaning operations. In this work, three different datasets (NSL-KDD, CICIDS 2018 and UNSW-NB15) have the class labels with symbolic values like “normal” or “intrusion type.” Then, the created IDS tries to
distinguish the legitimate and malicious communications without disclosing the nature of the assault. Consequently, the process of scaling or changing each feature’s data values into a proportional range is known as data normalization. Our data processing can be seen in Figure 1.

B. Intensive Binary Pigeon Optimization (IBiPO)

In this stage, the features of the preprocessed dataset are extracted by using the IBiPO technique and can be seen in Figure 1. It is mainly used to obtain an optimal set of features for reducing the dimensionality of dataset, which also supports to increased detection accuracy and performance of classifier. The IBiPO is a meta-heuristic optimization technique, which comprises three operators such as landmark, map, and compass. Among other optimization techniques, the key merits of using this approach are as follows: high convergence speed, reduced overfitting, and easy to deploy. In this technique, the pigeons perceive the geomagnetic fields in the map and compass operators to create a map for homing. Let consider that, the searching space having N dimensions, and i pigeons of swarms as represented in below:

Let consider that, the searching space having N dimensions, and i pigeons of swarms as represented in below:

\[ S_i = (S_{i1}, S_{i2}, \ldots, S_{iN}) \]  

Then, the velocity of pigeon is represented based on its changing location in the N dimensional vector as shown in below:

\[ Q_i = (Q_{i1}, Q_{i2}, \ldots, Q_{iN}) \]

Similarly, the visited locations of the ith pigeons are represented as follows:

\[ Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iN}) \]  

Consequently, the global optimal location of the pigeons are considered as \((K_{i1}, K_{i2}, \ldots, K_{iN})\), and all pigeons can fly in the searching space by using the following model:

\[ Q_i(h + 1) = Q_i(h) \times e^{-\alpha h} + \alpha \times (S_g - Q_i(h)) \]  

\[ S_i(h + 1) = S_i(h) + Q_i(h + 1) \]  

Where, \(S_i(h)\) denotes the present location of pigeon at time h, G denotes the map and compass factors, a is the arbitrary value ranging between 0 to 1, \(S_g\) is the global optimal solution, \(Q_i(h)\) represents the velocity at time h.

C. Bi-Directional LSTM (Bi-LSTM) Classification

After feature optimization, the selected subset of features are used for classification, where the Bi-LSTM model is used to accurately spot the intrusions in WSN as can be seen in Figure 1. Typically, the Bi-LSTM is a kind of sequence processing mechanism, which comprises two LSTM models. In which, one LSTM will receive input going forward, and the other will receive input going backward. The effectiveness of the model is increased when the LSTM is applied twice because it changes how long-term interdependencies are learned. These dependencies can be monitored as the sequence progresses. The LSTM is made to avoid the long-term dependence issue by recollecting the data for a lengthy period of time and incorporating a memory cell. Moreover, it comprises three gates such as input gate, forget gate, and output gate, in which the input gate determines how much additional data will indeed be transferred to the memory, the output gate determines if the current value in the cell subsidizes to the output, and the forget gate determines whether to retain or discard available data. In neural networks, activation functions are used to estimate the weighted sum of inputs and biases, then it determines whether a neuron can activate or not.

III. RESULTS AND DISCUSSION

This section validates the results and performance of the proposed IDS model by using various evaluation indicators and datasets. For this assessment, the different types of network intrusion datasets have been used, which hold NSL-KDD, UNSW-NB 15, and CICIDS-2018. In which, four different types of assaults, including DoS, Probe, R2L, and U2R, are included in the NSL-KDD dataset. The remaining data is classified as typical data and only comes within the four categories. Under the four categories, 39 various attacks are grouped together, and rather than detailing each attack individually, each attack is mapped into the corresponding group. The performance metrics (7-12) used in this study are computed as follows:

\[ \text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{p} + \text{n}} \]  

\[ \text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \]  

\[ \text{Recall or TRP} = \frac{\text{tp}}{\text{tp} + \text{fn}} \]  

\[ \text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \]  

\[ \text{FPR} = \frac{\text{fp}}{\text{tp} + \text{fn}} \]  

\[ \text{FNR} = \frac{\text{fn}}{\text{tp} + \text{fn}} \]

Figure 2 shows the comparison of precision for traditional algorithms and proposed model, and it shows that the proposed IDS model is more precise than traditional machine learning and deep learning-based IDS methods. A comparison analysis for the recall parameter is shown in Figure 3, where the investigation shows that the proposed model has maximum recall values.
The estimated results indicate that the proposed model can detect the greatest number of intrusions while traditional SVM, DT, RF, and DNN models perform less well. The proposed model has a recall percentage that is 30% higher than that of SVM, DT, RF, and DNN. Moreover, Figure 4 shows a thorough comparison of all the techniques used for the various types of attacks in the dataset. As can be seen, the suggested model's detection rate is higher than that of other approaches for all attacks.

Table 1 provides a summary of the overall performance of the proposed model versus traditional machine learning & deep learning models. The investigation shows that the proposed IDS model has a maximum accuracy of 98.8%, which is higher than the other techniques. Due to the best feature processing and selection, the proposed model has a bettercomputationability and systemefficacy. The use of proposed optimization + classification model highly improves the accuracy of intrusion detection performance, but the performance of standard machine learning and deep learning algorithms declines as a result of inadequate feature selection and processing. Finally, the findings show that the suggested model may successfully identify intrusions in sensor networks as can be seen in Figure 5.

Additionally, Table II compares the proposed IDS model with other models by using UNSW-NB 15 dataset. For this evaluation, the existing SVM, ELM, and MK-ELM [19] techniques are considered, which have comparable accuracy. However, the accuracy of the proposed IDS model is the highest, and the accuracy of all three algorithms has decreased when compared to the proposed technique.

Table III presents the overall comparative analysis of the existing and proposed anomaly detection methodologies.
based on the parameters of false alarm rate, detection rate, accuracy, and execution time. Based on the study, it is determined that the proposed model outperforms other approaches with better prediction results.

<table>
<thead>
<tr>
<th>IDS Methods</th>
<th>False alarm rate</th>
<th>Detection rate</th>
<th>Accuracy</th>
<th>Execution time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust based IDS</td>
<td>Low</td>
<td>Very High</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td>ARIMA – traffic anomaly detection</td>
<td>Low</td>
<td>Very High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Lightweigh t IDS</td>
<td>Very Low</td>
<td>Very High</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td>Sensor anomaly Detection</td>
<td>Low</td>
<td>Very High</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td>PSO-IDS</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td>Evolutionary NN-IDS</td>
<td>Very High</td>
<td>Very High</td>
<td>Very</td>
<td>High</td>
</tr>
<tr>
<td>Proposed IBiPO + Bi-LSTM</td>
<td>Very Low</td>
<td>Very High</td>
<td>Very</td>
<td>Very Low</td>
</tr>
</tbody>
</table>

### IV. CONCLUSION

In this paper, a novel and computational efficient IDS framework, named as, IBiPO + Bi-LSTM model is proposed for securing WSN. The original contribution of this paper is to highly protect the network from the harmful intrusions or anomalies. In this context, the system is implemented using the well-known and open IDS datasets, which were initially preprocessed using data cleaning, normalization, splitting, and clustering processes. The balanced dataset is then applied to additional procedures of optimization and classification. The IBiPO method is mainly used to choose the most pertinent information best in order to anticipate intrusions effectively with less false alarms. Additionally, this method aids in the high accuracy detection and simplification of intrusion classification. Following that, the Bi-LSTM classifier receives the collected features for training and testing. Based on the learning features, this classifier predicts the normal and anomalous data. The suggested IBiPO + Bi-LSTM model has the advantages of low overfitting, quick processing, and excellent detection accuracy. During performance analysis, the system is assessed against recent relevant works in terms of DR, FPR, accuracy, precision, recall, and time consumption. Moreover, three well-known IDS datasets (NSL-KDD, CICIDS 2018 and UNSW-NB15) were utilized in all experiments for evaluation. Using the three aforementioned datasets, the suggested system performs better than the existing models.

### REFERENCES