
C. Murugesh¹, Dr. S. Murugan²

¹Assistant Professor, Programmer, Department of Computer and Information Science, Annamalai University, Annamalai Nagar
Email: 78murugesh[at]gmail.com

²Assistant Professor, Dr. M.G.R. Government Arts and Science College for Women, Villupuram.
Email: smuruganmpt79[at]gmail.com

Abstract: Wireless Sensor Networks (WSNs) serve as the backbone for numerous critical applications, including environmental monitoring, industrial automation, and healthcare. However, the inherent vulnerabilities of these networks to security threats necessitate robust Intrusion Detection Systems (IDS). Traditional rule-based IDSs often fall short of addressing the evolving nature of attacks in WSNs. This research presents an innovative approach that combines the power of Support Vector Machine (SVM) classifiers with the Grasshopper Optimization Algorithm (GOA) for effective intrusion detection (SVMGOA-ID) technique in WSNs. The SVMGOA-ID approach harnesses SVM, a well-established machine learning technique known for its proficiency in binary classification tasks. SVMs are trained to distinguish between normal network behavior and intrusion attempts, learning intricate patterns from a labeled dataset. However, the success of SVMs is highly dependent on appropriate parameter settings, and suboptimal choices can lead to reduced detection accuracy. To address this challenge, the Grasshopper Optimization Algorithm (GOA) simulated by the natural behavior of grasshoppers in search of optimal foraging spots, is introduced for parameter optimization. The GOA efficiently explores the parameter space of SVM models, seeking the ideal configuration that maximizes intrusion detection accuracy. Comprehensive experiments are conducted using benchmark datasets, evaluating the efficiency of the SVMGOA-ID methodology in detecting various intrusion types.

Keywords: Wireless Sensor Networks; Support Vector Machine; Intrusion Detection Systems; Grasshopper Optimization Algorithm

1. Introduction

Wireless sensor networks (WSNs) are not same as the standard computer network, however, it interconnects sensor networks through wireless connections without a centralized control network [1]. This can be similar to distributed transmission that adjacent nodes that are controlled by each node in a system. It has developed and is vulnerable to any attacks in a suitable atmosphere [2]. It is owing to extensive Internet use and numerous security difficulties that arise in different DoS forms. It is the major problem of every network security problem as it produces larger amount of data traffic to utilize entire allocated system sources and deactivates the transmission connection by stopping server from processing authorized requisitions for clients to perform transactions [3]. The DDOS practicable computational sources are the processing part and network bandwidth of memory and computers. Its capacity is compressed by the transmission channels [4]. If DDOS can be initiated, a massive quantity of unacceptable traffic data overflowed the allocated channel into a transmission link. Occasionally, attack targets to the network nodes protect a massive quantity of undesirable requests with all communication channels to the node [5]. Fig. 1 portrays the architecture of IDS.

![Intrusion Detection System](image)

Figure 1: Architecture of IDS

The Intrusion Detection System (IDS) is presented by Denning in 1987 then, with numerous enhanced methods and developments imposed on it [6]. IDS was demonstrated as an efficient technology to face cyber-attacks. IDS is categorized into two categories namely host- and network-based on position of IDS system in the networks [7]. A host...
based IDS monitors and detects application tasks for devices processing on system and network configurations. An advantage of a host based IDS has been analyzed in previous databases for detecting savvy intruders, which can be utilized in advanced methods and appears hard to detect in actualtime [8]. Nevertheless, it contains a few drawbacks which are processing time, storage on the hosts, other resources, and memory consumption. It can detect a large network with minimum use, simultaneously, it is an optimum-grained identification ability [9]. DL methods for ID are recently analyzed by several researchers. The huge volume of network data produced ID problems likable to DL approaches [10].

Fu et al. [11] developed a DDQN technique for arrangement position depending on Geography Position Information (GPI). GPI can prevent the complex computational procedure of channel state data. The DDQN technique could be presented for attaining the functional relation among GPI and the optimum UAV deployment position. Also, the incorporation of CNN-LSTM methods. In [12], the authors presented a hybrid method in that the nodes have been collectively clustered to create the CDS thus, the data transmission is increased. Packet distribution is executed depending on the CDS of nodes and unsuitable network failure to maintain similar node waking as frequently can be prevented.

Devi et al. [13] recommended novel rule-based IDS, which contains a count of security methods. This technique uses a set of rules produced by the RF method. These created rules are dependent upon features extraction from WSNs data comprising hop count, energy consumption, and packet size. Integrated rule-based methods and ML techniques could be employed for enhancing the accuracy and effectiveness of intrusion detection approaches. Hemanand et al. [14] designed an intelligent IDS technique through the Cuckoo Search Greedy Optimization (CSGO) and Likelihood-SVM (LSVM) approaches. Firstly, the database pre-processing is carried out for normalizing the features. The optimum feature sets were chosen and provided to the input of CSGO method. Secondly, the LSVM-based ML classification method was employed to predict the classified label.

Umamaheshwari et al. [15] introduced efficient IDS for classification of attacks in a WSN employing ML approach. A baseline technique has been designed by features extraction from WSN-DS database applied a DT method. For minimizing the time for attack identification, an FS employing Statistical Analysis, Fisher Score, and Correlation Score by Kruskal-Wallis (KW) test, Relief, and MMRM techniques are analyzed. Srivastava and Bharti [16] recommended a Hybrid Model of One-class SVM and Isolation forest (HMOI) technique that is a ‘Classification + Classification’ framework. It can be 2 important stages. The first stage overcomes the problem of unlabelled data that has major popularity in real-time database of WSNs and transforms it into labelled data. The second stage has been determined for performing anomaly detection.

This research presents an innovative approach that combines the power of Support Vector Machine (SVM) classifiers with the Grasshopper Optimization Algorithm (GOA) for effective intrusion detection (SVMGOA-ID) technique in WSNs. The SVMGOA-ID approach harnesses SVM, a well-established ML technique known for its proficiency in binary classification tasks. SVMs are trained to distinguish between normal network behavior and intrusion attempts, learning intricate patterns from a labeled dataset. However, the success of SVMs is highly dependent on appropriate parameter settings, and suboptimal choices can lead to reduced detection accuracy. To address this challenge, the GOA simulated by the natural behavior of grasshoppers in search of optimal foraging spots, is introduced for parameter optimization. The GOA efficiently explores the parameter space of SVM models, seeking the ideal configuration that maximizes intrusion detection accuracy.

2. The Proposed Model

In this study, we have focused and development of the SVMGOA-ID technique in WSNs. The main purpose of SVMGOA-ID technique contains two phases namely SVM-based feature selection and hyperparameter tuning using GOA.

2.1 Feature selection-based SVM

In this stage, the SVMs are highly dependent on appropriate parameter settings, and suboptimal choices can lead to reduced detection accuracy. During the labeled database with N instances, the labels \( y \) are binary, taking a value of both 1 and −1 [17]. The feature vector \( x_i \) is a n-dimension vector that signifies the amount of accessible features and is determined by Eq. (1).

\[
D = \{(x_i, y_i) | x_i \in R^n, y_i \in \{-1, 1\}\}_{i=1}^{N}
\] (1)

The optimum hyperplane is defined by formula \( f(x) = w \cdot x + b \), with \( x \) as input, \( w \) the feature co-efficient, and \( b \) implies the bias. The purpose is to minimize \( ||w||^2 \) butadequatelimitations (Eq. (2)), purpose for maximizing the margin across the hyperplane as well as neighboring instances in 2-classes data. The SVM approach balances decreasing misclassifications and determining a hyperplane with important margin, dependent upon the elected kernel function (for instance, polynomial, linear, or RBF).

\[
\min \left( \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} \xi_i \right)
\]

Subject to \( y_i(x_i \cdot b + 1 - \xi_i) \geq 1 \quad \forall i = 1, ..., N \)

\( \xi_i \geq 0 \) (2)

whereas \( \xi_i \) refers to the slack variable to evaluate the distance amount of the misclassified and hyperplane instances with penalty co-efficient (C). The KuhnTucker statehas been changedas a dual Lagrangian problem by establishing Lagrangian multipliers for limitations of the problems (Eq. (2)). The purpose is to evaluate the difference (\( \xi \)) among the hyperplane as well as incorrectly located instances, and resolve the problem utilizing the transformed Lagrangian dual formulation:

\[
\min \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i x_j
\]
By resolving the optimizer problem in Eq. (3), the Lagrange co-efficient \(\alpha_i\) (Lagrange co-efficient of \(i^{th}\) instance) is defined. This optimum value is utilized for calculating the hyperplane parameters \((b\ and \ w)\), leading to the subsequent classification function:

\[
f(x) = \text{sign} \left(\sum_{i=1}^{N} \alpha_i y_i (x_i \cdot x_j) + b\right)
\]  

(4)

If linear separation is difficult, nonlinear classifier approach such as SVM is employed. By deploying a mapping function represented in Eq. (5), the SVM transmissions data in low-to-high dimensional spaces, permitting simple separation among class borders. Eq. (6) presents the utilization of a non-linear function, \(\phi(x)\), for mapping input feature vectors \((x)\) in an \(n\)-dimension space to 1D feature space, increasing classification. This problem needs to determinize the kernel function \((x_i, x_j)\), as expressed in Eq. (7).

\[
\begin{align*}
\forall i, j \rightarrow \phi(x_j) \\
\phi(x) = (\phi_1(x), ..., \phi_N(x)) \\
K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)
\end{align*}
\]  

(5)

(6)

(7)

Next, the optimizer problem changes as the formula written as:

\[
\begin{align*}
\min & \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\
\text{Subject to} & : \sum_{i=1}^{N} \alpha_i y_i = 0 \\
& : 0 \leq \alpha_i \leq C
\end{align*}
\]  

(8)

If the non-linear kernel can be employed, Eq. (4) for the decision function canaller and change into Eq. (9).

\[
f(x) = \text{sign} \left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x_j) + b\right)
\]  

(9)

The RBF is a generally employed kernel in SVM applications. It is determined as Eq. (10):

\[
K(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)
\]  

(10)

SVMs are prominent for their computational efficacy and efficiency in controlling high-dimension data without being dependent on difficult methods.

### 2.2 Parameter tuning using GOA

To address this challenge, the GOA simulated by the natural behavior of grasshoppers in search of optimal foraging spots, is introduced for parameter optimization. GOA is a novel algorithm simulated by the swarming behaviors of grasshoppers [18]. All the grasshoppers in the swarm have their own location is the feasible solution to the optimization problems. GOA comprises three subcomponents, \(S_i\), the social interaction, \(G_i\), the gravitation force on \(i^{th}\) grasshoppers, and \(A_i\) the wind advection.

\[
X_i = S_i + G_i + A_i
\]  

(11)

In Eq. (11) \( Xi \) shows the location of \( i^{th} \) grasshoppers.

\[
S_i = \sum_{l=1}^{N} s (d_{ij}) d_{ij}
\]  

(12)

In Eq. (12), \( s \) shows the function to determine the strength of social force, \( d_{ij} = |X_j - X_i| \) the distance among \( i^{th} \) and \( j^{th} \) grasshoppers, and \((d_{ij} =) (X_j - X_i)/d_{ij}\) indicates a unit vector from \( i^{th} \) to \( j^{th} \) grasshoppers.

\[
s(r) = f e^{k((-r))}/1 - e^{-k(-r)}
\]  

(13)

In Eq. (13), \( f \) refers to the intensity of attraction, \( 1 \) shows the attractive length scale, within \([0,4]\) control attraction or repulsion among individual grasshoppers, and \( r \) denotes the force of repulsion. The distance should be standardized within \([1,4]\) since \( s \) function could not manage stronger forces with longer distance.

The \( G \) element consists of two different parts, \( g \) denotes the gravitational constant and \((g \cdot g)^\wedge\) indicates a unity vector nearby the center of earth.

\[
G_i = -g e^g
\]  

(14)

The wind advection \( A \) is evaluated by Eq. (14):

\[
A_i = u e^w
\]  

(15)

In Eq. (15), \( e^w \) shows the unity vector in the wind direction and \( u \) denotes the constant drift.

\[
X_i = \sum_{j=1}^{N} s_{j} |x_j - j_j| \frac{|x_j - j_j|}{d_{ij}} - g e^g + u e^w
\]  

(16)

The balance among exploitation as well as exploration in a stochastic algorithm assist in searching for global optima. Any special parameters are added to show exploitation and exploration in dissimilar phases of optimizer:

\[
X_i^d = \left(\sum_{j=1}^{N} \left(\frac{u b_d - l b_d}{2} s |x_j - j_j| \frac{|x_j - j_j|}{d_{ij}}\right)\right) + T_d
\]  

(17)

In Eq. (17), the component of \( G \) is ignored assuming no wind direction and gravitational force is often towards a target. \([lb]\) and \([ub]\) shows the upper and lower bounds at the \( d \) dimensional space and \( T_d \) refers to the values of \( d^{th} \) dimension in the target. The inner ‘\( c \)’ decreases attraction or repulsion forces among grasshoppers relative to the amount of iterations, whereas outer ‘\( c \)’ maintain the balance between exploitation and exploration. The reducing coefficient ‘\( c \)’ is applied twice in Eq. (17) to control forces between grasshoppers and is updated using Eq. (18).

\[
c = c_{max} - L \frac{c_{max} - c_{min}}{L}
\]  

(18)

Where \( L \) represents the maximal iteration counter, \( c_{max} = 1 \) denotes the maximal value, \( c_{min} = 0.00001 \) indicates the minimal value, and \( L \) shows the existing iteration.

### 3. Experimental Validation

The ID results of the SVMGOA-ID approach are studied here. Fig. 2 exhibits the confusion matrices produced by the SVMGOA-ID methodology at 80:20 and 70:30 of TRP/TSP. The resultspoint out the effective recognition and classification of 5 classes accurately.
The ID results of the SVMGOA-ID approach with 80:20-TRP/TSP are studied in Table 1 and Fig. 3. The results imply the effectual recognition of five classes. On 80%-TRP, the SVMGOA-ID approach accomplishes average $\text{accu}_y$, $\text{sens}_y$, $\text{spec}_y$, $F_{\text{score}}$, and MCC of 99.61%, 95.73%, 99.49%, 93.59%, and 92.62% correspondingly. Then, on 20%-TSP, the SVMGOA-ID methodology realizes average $\text{accu}_y$, $\text{sens}_y$, $\text{spec}_y$, $F_{\text{score}}$, and MCC of 99.62%, 96.60%, 99.56%, 94.04%, and 93.11% correspondingly.

### Table 1: ID outcome of SVMGOA-ID approach with 80:20-TRP/TSP

<table>
<thead>
<tr>
<th>Class</th>
<th>$\text{Accu}_y$</th>
<th>$\text{Sens}_y$</th>
<th>$\text{Spec}_y$</th>
<th>$F_{\text{score}}$</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRP (80%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>99.14</td>
<td>99.23</td>
<td>98.28</td>
<td>99.53</td>
<td>95.06</td>
</tr>
<tr>
<td>C2</td>
<td>99.72</td>
<td>98.03</td>
<td>99.76</td>
<td>94.92</td>
<td>94.82</td>
</tr>
<tr>
<td>C3</td>
<td>99.74</td>
<td>98.83</td>
<td>99.78</td>
<td>96.72</td>
<td>96.61</td>
</tr>
<tr>
<td>C4</td>
<td>99.70</td>
<td>84.76</td>
<td>99.83</td>
<td>83.32</td>
<td>83.18</td>
</tr>
<tr>
<td>C5</td>
<td>99.76</td>
<td>97.79</td>
<td>99.79</td>
<td>93.46</td>
<td>93.43</td>
</tr>
<tr>
<td>Average</td>
<td>99.61</td>
<td>95.73</td>
<td>99.49</td>
<td>93.59</td>
<td>92.62</td>
</tr>
<tr>
<td>TSP (20%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>99.16</td>
<td>99.21</td>
<td>98.63</td>
<td>99.53</td>
<td>95.18</td>
</tr>
<tr>
<td>C2</td>
<td>99.74</td>
<td>98.31</td>
<td>99.78</td>
<td>95.27</td>
<td>95.19</td>
</tr>
<tr>
<td>C3</td>
<td>99.73</td>
<td>98.77</td>
<td>99.77</td>
<td>96.67</td>
<td>96.55</td>
</tr>
<tr>
<td>C4</td>
<td>99.73</td>
<td>88.84</td>
<td>99.83</td>
<td>85.63</td>
<td>85.55</td>
</tr>
<tr>
<td>C5</td>
<td>99.75</td>
<td>97.86</td>
<td>99.78</td>
<td>93.10</td>
<td>93.09</td>
</tr>
<tr>
<td>Average</td>
<td>99.62</td>
<td>96.60</td>
<td>99.56</td>
<td>94.04</td>
<td>93.11</td>
</tr>
</tbody>
</table>
The ID result of the SVMGOA-ID methodology with 70:30-TRP/TSP is studied in Table 2 and Fig. 4. The outcome value referred the effective recognition of five classes. On 70%-TRP, the SVMGOA-ID method realizes average $acc_y$, $sens_y$, $spec_y$, $F_{score}$, and MCC of 99.61%, 97.25%, 99.48%, 94.22%, and 93.25% correspondingly. Next, on 30%-TSP, the SVMGOA-ID system achieves average $acc_y$, $sens_y$, $spec_y$, $F_{score}$, and MCC of 99.62%, 97.36%, 99.46%, 94.63%, and 93.68% correspondingly.

**Table 2: ID outcome of SVMGOA-ID methodology with 70:30-TRP/TSP**

<table>
<thead>
<tr>
<th>Class</th>
<th>$Acc_y$</th>
<th>$Sens_y$</th>
<th>$Spec_y$</th>
<th>$F_{score}$</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>99.10</td>
<td>99.19</td>
<td>98.26</td>
<td>99.51</td>
<td>94.84</td>
</tr>
<tr>
<td>C2</td>
<td>99.76</td>
<td>98.46</td>
<td>99.79</td>
<td>95.63</td>
<td>95.55</td>
</tr>
<tr>
<td>C3</td>
<td>99.69</td>
<td>97.34</td>
<td>99.78</td>
<td>96.02</td>
<td>95.86</td>
</tr>
<tr>
<td>C4</td>
<td>99.77</td>
<td>94.07</td>
<td>99.81</td>
<td>87.43</td>
<td>87.53</td>
</tr>
<tr>
<td>C5</td>
<td>99.72</td>
<td>97.17</td>
<td>99.77</td>
<td>92.50</td>
<td>92.47</td>
</tr>
<tr>
<td>Average</td>
<td>99.61</td>
<td>97.25</td>
<td>99.48</td>
<td>94.22</td>
<td>93.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class</th>
<th>$Acc_y$</th>
<th>$Sens_y$</th>
<th>$Spec_y$</th>
<th>$F_{score}$</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>99.12</td>
<td>99.23</td>
<td>98.08</td>
<td>99.51</td>
<td>94.95</td>
</tr>
<tr>
<td>C2</td>
<td>99.77</td>
<td>98.75</td>
<td>99.80</td>
<td>95.84</td>
<td>95.77</td>
</tr>
<tr>
<td>C3</td>
<td>99.69</td>
<td>97.40</td>
<td>99.78</td>
<td>96.10</td>
<td>95.95</td>
</tr>
<tr>
<td>C4</td>
<td>99.76</td>
<td>94.69</td>
<td>99.81</td>
<td>88.01</td>
<td>88.12</td>
</tr>
<tr>
<td>C5</td>
<td>99.76</td>
<td>96.73</td>
<td>99.82</td>
<td>93.71</td>
<td>93.64</td>
</tr>
<tr>
<td>Average</td>
<td>99.62</td>
<td>97.36</td>
<td>99.46</td>
<td>94.63</td>
<td>93.68</td>
</tr>
</tbody>
</table>
The SVMGOA-ID approach is compared with other ML classifiers in Table 3. In Fig. 5, the ID outcomes of the SVMGOA-ID approach are examined in terms of $accu_y$ and $F_{score}$. The results exhibited the effectual performance of the SVMGOA-ID approach. Based on $accu_y$, the SVMGOA-ID approach reaches improving $accu_y$ of 99.62% while the Adaboost, GB, XGBoost, KNN, and KNN-PSO methodologies attain decreasing $accu_y$ of 96.30%, 94.23%, 95.91%, 96.40%, and 96.47% respectively. In addition, based on $F_{score}$, the SVMGOA-ID systemachieves enhance $F_{score}$ of 94.63% while the Adaboost, GB, XGBoost, KNN, and KNN-PSO methods reach lesser $F_{score}$ of 91.09%, 92.43%, 90.70%, 90.79%, and 92.59% correspondingly.

In Fig. 6, the ID analysis of the SVMGOA-ID approach is examined in terms of $sens_y$ and $spec_y$. The outcome values depicted the effectual performance of the SVMGOA-ID approach. Based on $sens_y$, the SVMGOA-ID methodology reaches improving $sens_y$ of 97.36% while the Adaboost, GB, XGBoost, KNN, and KNN-PSO methods attain lowers $sens_y$ of 94.96%, 96.95%, 94.75%, 96.99%, and 94.10% correspondingly. Moreover, based on $spec_y$, the SVMGOA-ID approach reaches greater $spec_y$ of 99.46% while the Adaboost, GB, XGBoost, KNN, and KNN-PSO methodologies gainminimals $spec_y$ of 94.47%, 94.55%, 94.14%, 96.20%, and 94.21% correspondingly.

Table 3: Comparative outcome of SVMGOA-ID methodology with ML systems

<table>
<thead>
<tr>
<th>Methods</th>
<th>$Accu_y$</th>
<th>$Sens_y$</th>
<th>$Spec_y$</th>
<th>$F_{Score}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Proposed Model</td>
<td>99.62</td>
<td>97.36</td>
<td>99.46</td>
<td>94.63</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>96.30</td>
<td>94.96</td>
<td>94.47</td>
<td>91.09</td>
</tr>
<tr>
<td>GB</td>
<td>94.23</td>
<td>96.95</td>
<td>94.55</td>
<td>92.43</td>
</tr>
<tr>
<td>XGBoost</td>
<td>95.91</td>
<td>94.75</td>
<td>94.14</td>
<td>90.70</td>
</tr>
<tr>
<td>KNN</td>
<td>96.40</td>
<td>96.99</td>
<td>96.20</td>
<td>90.79</td>
</tr>
<tr>
<td>KNN-PSO</td>
<td>96.47</td>
<td>94.10</td>
<td>94.21</td>
<td>92.59</td>
</tr>
</tbody>
</table>
4. Conclusion

In this study, we have focused and development of the SVMGOA-ID technique in WSNs. The SVMGOA-ID approach harnesses SVM, a well-established ML technique known for its proficiency in binary classification tasks. SVMs are trained to distinguish between normal network behavior and intrusion attempts, learning intricate patterns from a labeled dataset. However, the success of SVMs is highly dependent on appropriate parameter settings, and suboptimal choices can lead to reduced detection accuracy. To address this challenge, the GOA simulated by the natural
behavior of grasshoppers in search of optimal foraging spots, is introduced for parameter optimization. The GOA efficiently explores the parameter space of SVM models, seeking the ideal configuration that maximizes intrusion detection accuracy.

References


