CHAPTER 3
OVERALL SYSTEM ARCHITECTURE

3.1 INTRODUCTION

This section provides a description of the overall system architecture of the proposed framework, which projects the internal functional flow. The three variant supervised machine-learning techniques (Multi-Layer Perceptron, Extreme Learning Machine and Support Vector Machine) are presented as the background of existence, describing their need to improve the detection accuracy. Simultaneously, the Grey Wolf Optimization algorithm is discussed by describing their characteristics to search for the required fittest optimal solution.

3.1.1 Background of Existence-Machine Learning Based

The background of the existence from the machine learning-based approach is provided with the three variant supervised machine learning techniques as discussed above to exhibit their performance ability towards designing the diversified computational model against fault identification during existence.

3.1.1.1 Multi-Layer Perceptron

Multi-Layer Perceptron is predominantly determined as the intelligence neural network that can be built by on one’s or as own, created by the algorithm or both. The major characteristics of this technique use the back propagation to classify the instance. This uniqueness helps in achieving the highest reliability to overcome the futile rate of rule-based expert systems through the use of the technique deployed in the design of the Intrusion Detection System. The successive rate of Multi-Layer Perceptron-based Intrusion Detection System is obtained at a rate of 94.3% [91].

The structure of Multi-Layer Perceptron layered with the three variant modes is shown namely the input node, the hidden node, and the output node
is shown in Figure 3.1. Here, each element decides the feature vector and the nodes are interrelated by assigning the weights to the connection. Complex computational tasks can be performed by a probability distribution of the hidden node extracted from the input node and forwarding the network to output for finding accurate class instances. The findings of each computational task are gathered to obtain the conclusion.

![Figure 3.1 The Structure of Multi-Layer Perceptron](image)

Barapatre et al [92] have proposed the system that uses a Multi-Layer Perceptron neural network to show the development of the rate of detection by reducing the false alerts in designing the Intrusion Detection System. Experimentation was done by deploying a KDDCUP99 dataset. The system was trained using sigmoid activation function for accurate identification of every category of attack classes (Denial-Of-Service attacks, User-To-Root attacks, Root-To-Local attacks, and Probing attacks). The overall detection accuracy obtained by this technique is 81.96% and the false positive rate is 8.51%.

Abuadllah et al [93] have proposed the flow-based intrusion detection system using neural networks for detection and classification of attack network traffic patterns. For experimentation, a DARPA99 dataset was used with back-end support of MATLAB. Here, an approximation of twenty-one thousand eight hundred samples was taken into account. In stage one, the number of input nodes corresponds to the number of selected features of the
Net-Flow dataset. This stage is assigned with one input node, one hidden node, and two output nodes for finding both normal and attack classes. Moreover, the number of nodes in the hidden layer was attained by the computational tasks of backpropagation and the process of trial and error. The various results of detection rate obtained by this stage are 92.7%, 94.2%, and 91.1%, and the false positives are 3.6%, 3.4%, and 5.1% respectively.

In stage two, neural networks were used for the classification of attack categories. This stage is assigned with one input node, one hidden node, and five output nodes for finding the 'normal' class and four types of attack categories. The classified attack categories were Denial-Of-Service attacks, Port Scan attacks, Land attacks, unknown attacks and 'normal' traffic pattern. The performance of the hidden layer for stage two determined was the same as of stage one. The experimentation was performed from the DARPA dataset collection of twenty-one thousand samples approximately. The results of detection rate obtained by stage two are, 95.4% & 99.4% and false positives are 4.6% & 0.58% respectively.

The two-layer feedforward neural network system was designed for the accurate classification of attack categories [94]. The experimentation was performed using 10% subset of the KDDCUP99 dataset, out of which approximately twenty-six thousand instances were taken into account. The detection accuracy attained by the proposed system was 91.9%. Feature reduction techniques were proposed using the Multi-Layer feed forward neural network [95]. Here, two types of experiments were performed using the NSL-KDD dataset with reduced features from 41 to 29 neurons used. The experimentation one is executed using binary classification by deploying the BFGS quasi-Newton backpropagation algorithm.

The overall detection accuracy is achieved at a rate of 79.9% for all 41 features. The experimentation two is performed to classify attack categories by deploying the Levenberg-Marquard back-propagation algorithm. The overall detection accuracy is achieved at a rate of 81.2% for reduced features from 41 to 29 neurons. Table 3.1 presents the details of the detection accuracy achieved by Multi-Layer Perceptron technique in existence.
<table>
<thead>
<tr>
<th>Author</th>
<th>Methodology</th>
<th>Dataset utilization</th>
<th>Detection Rate</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mukhopadhyay et al., 2011</td>
<td>Scaled conjugate gradient algorithm</td>
<td>KDDCUP99</td>
<td>Level 1: 95.6%</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Level 2: 73.9%</td>
<td></td>
</tr>
<tr>
<td>Baraputra et al., 2008</td>
<td>Training MLP neural network</td>
<td>KDDCUP99</td>
<td>81.96%</td>
<td>8.51%</td>
</tr>
<tr>
<td>Singh et al., 2015</td>
<td>Neuro solution software</td>
<td></td>
<td>94.3%</td>
<td>N/A</td>
</tr>
<tr>
<td>Abuadullah et al., 2014</td>
<td>Flow-based Anomaly Intrusion Detection System</td>
<td>DARPA dataset with 22,000 instances approximately</td>
<td>Experiment 1 92.7% 94.2% 91.1%</td>
<td>3.6% 3.4% 5.1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Experiment 2 95.4% 99.4%</td>
<td>4.6% 0.58%</td>
</tr>
<tr>
<td>Kumar et al., 2014</td>
<td>Two layer feed forward neural network</td>
<td>10% subset of KDDCUP99</td>
<td>91.9%</td>
<td>N/A</td>
</tr>
</tbody>
</table>

3.1.1.2 Extreme Learning Machine

Farias et al [96] and Xiang et al [97] addressed the significant characteristics of Extreme Learning Machine, a widely adopted and supervised machine learning technique used for the classification of normal and abnormal features. Extreme Learning Machine has been developed as a single layer feed forward neural network for parameter optimization in a dynamic state. Traditionally, the detection techniques hold the limits to the analysis of the presence of intrusion traces over network traffic. The significant benefits of Extreme Learning Machine overcome its limitations, which helps it turning out to be an emerging technique towards designing the Intrusion Detection System. The structure of Extreme Learning Machine is stated as the following quotes.
- Randomly assigns the input weights 'w;' and bias 'b;' leads to Constant Probability Distribution (CPD) function. This happens for mapping the input data into a feature space.
- Calculate the hidden layer of the output matrix.
- Calculate the outputs.

![Diagram of Extreme Learning Machine](image)

**Figure 3.2 The structure of Extreme Learning Machine**

Figure 3.2 illustrates the structure of Extreme Learning Machine (ELM). The hidden layer of ELM exists with combinational nodes. Each node is assigned a different set of compute nodes. Moreover, randomly assigned ELM parameter node leads to unique proficiency when compared to traditional backpropagation techniques. The characteristics of Extreme Learning Machine were discussed in [98] for designing an Intrusion Detection System. The fast learning rate and potential adaptation of the large collection of data patterns, have helped utilization of ELM to provide diversified standardization techniques. Here, it is proposed that the hybrid approach of Kernelized Extreme Learning Machine with Levenberg-Marquard learning
algorithm can obtain higher detection rate and lower false alarm rate than the regular ELM approaches. Experimentation was performed using KDDCUP99 dataset, which obtained a detection accuracy of 97.89%, and a false alarm rate of 1.06%.

Zhi-Xin et al [99] have addressed the two conditions required for multi-label classification. The first condition is that Extreme Learning Machine selects the most nearby value as the target label if ELM has only a single output node, among the multi-class labels. The second condition is that the index of the output node ranges with the highest output value and is then considered as the label of input data, in the case of ELM having multi-output nodes. Here, the framework using Extreme Learning Machine-Auto Encoder (ELM-AE) is proposed for feature extraction with the objective of achieving eminent performance in terms of accuracy and efficiency in multiple faults detection of wind turbines. The proposed work is layered to set the search region with the range from 0 to 1 at an interval of 0.01. Experimentation was composed of three components, namely, single acquisition module, feature extraction module, and fault identification module.

In the single acquisition module, the real-time dataset was used for recording the vibration of Wind Turbine Generator Systems (WTGS). Feature extraction module was induced with training and testing dataset from the collection of real-time accelerometers. ELM-AE forwarded the optimal parameter from the training to the testing bed. Fault identification module was induced with four variant classifiers (RVM, PNN, SVM, and ELM). ELM-AE holds all the information of the input data during the representation learning. The diversified computational methods are deployed using the classifier's optimal parameters as shown in Table 3.2.
Table 3.2 Zhin-Xin et al., (2016) represents a table for the evaluation of diversified computational models

<table>
<thead>
<tr>
<th>Feature Extraction</th>
<th>Classifier</th>
<th>Overall Accuracies attained</th>
</tr>
</thead>
<tbody>
<tr>
<td>WPT+TDSF+KPCA</td>
<td>PNN</td>
<td>83.76%</td>
</tr>
<tr>
<td></td>
<td>RVM</td>
<td>81.21%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>90.78%</td>
</tr>
<tr>
<td></td>
<td>ELM</td>
<td>90.89%</td>
</tr>
<tr>
<td>EMD+TDSF</td>
<td>PNN</td>
<td>84.52%</td>
</tr>
<tr>
<td></td>
<td>RVM</td>
<td>83.21%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>94.35%</td>
</tr>
<tr>
<td></td>
<td>ELM</td>
<td>94.32%</td>
</tr>
<tr>
<td>LMD+TDSF</td>
<td>PNN</td>
<td>84.52%</td>
</tr>
<tr>
<td></td>
<td>RVM</td>
<td>83.21%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>93.27%</td>
</tr>
<tr>
<td></td>
<td>ELM</td>
<td>94.44%</td>
</tr>
<tr>
<td>ELM-AE</td>
<td>PNN</td>
<td>84.52%</td>
</tr>
<tr>
<td></td>
<td>RVM</td>
<td>83.21%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>93.27%</td>
</tr>
<tr>
<td></td>
<td>ELM</td>
<td>94.42%</td>
</tr>
</tbody>
</table>

The overall accuracies of Support Vector Machine (SVM) and Extreme Learning Machine (ELM) are found to be attained with marginal differences. The ELM-AE is proposed for learning representations and feature extraction [100]. Experimentation was performed with the images of handwritten digits. The overall 60,000 samples taken from MNIST dataset were used for training and 10,000 samples were employed for testing purposes. The overall testing accuracy obtained by ELM-AE is 99.03% with the training time as 444.65 seconds.

Huang et al [101] have addressed the interest gain of ELM applied over feature extraction, learning representations and other tasks. The effectiveness in producing the best generalization procedure has led to the discussion of further enhancements. The analysis of high dimensional data was seen as a challenging task for the traditional techniques. A need was seen for designing the improved ELM that could handle the high dimensional data with intent to
strengthen the random feature mapping by reducing over-fitting. The parameters of ELM should generate the probability distribution function without affecting the approximation capability in real-time applications.

The Online Sequential ELM technique [102] is proposed for Intrusion Detection. The proposed technique uses the alpha profiling and beta profiling methodologies. The characteristic of alpha profiling is used for reducing chronological complexity and removing irrelevant features whereas beta profiling is used for reducing the size of the training dataset. Experimentation has been conducted using two varied datasets and they are Kyoto University benchmark dataset and NSL-KDD dataset. Here, NSL-KDD has experimented with binary-class and multi-class types. Experimentation results collected from Kyoto University benchmark dataset showed achievement of the detection accuracy at a rate of 96.73% and the false alarm of 5.76%. On the other hand, the experimentation results of binary-class NSL-KDD dataset showed the detection accuracy of 98.66% and the false alarm of 1.74% along with computational learning time taken around 2.43 seconds. The experimentation results of the multi-class NSL-KDD dataset obtained the detection accuracy of 97.67% and false alarm at a rate of 1.74% along with learning time around 2.65 seconds.

3.1.1.3 Support Vector Machine

The state-of-art for support vector machine (SVM) designed for Intrusion Detection Systems [103] is a well-known supervised machine learning classifier derived from the principle of structured risk minimization. This intervention paved the way for the creation of hyperplane with intent to extract classification and regression by maximizing the vector space. Likewise, utilizing the SVM in designing the Intrusion Detection Systems made the Researchers widely impressive in the designing of diversified set-ups. The significant characteristics of SVM are given below:

- Ability to seek an opinion from which one can find the lowest prone to errors.
• Ability to update the training pattern dynamically when there is a new pattern for classification.

• It also guarantees the most separation margin between the positive and negative classes. This characteristic made the SVM an idealistic approach, while the performance of the other classifiers was found to be difficult to achieve.

• SVM reduces the over-fitting problem existing in the model, with too many parameters related to the number of observations found.

![Diagram of Support Vector Machine](image)

Figure 3.3 The structure of Support Vector Machine

Figure 3.3 delineates the structure of SVM technique and its idealistic approach towards creating the hyperplane between positive and negative class. The working structure of SVM technique is classified into three phases. In the first phase, preprocessing module is evolved for feature extraction. Here, a set of training data is randomly picked up as input data and passed on to the training phase. In the second phase, the parameters of SVM are trained to adopt the input data and boost up the performance by utilizing the kernel evaluation process. Finally, in the testing phase, the outcome of trained SVM classifier's detection accuracy and its false alarm rate are delivered.

The Kernel Independent Component Analysis (KICA) was proposed [104] to extract the features from the trained dataset as input data and deployed to the SVM technique with intent to create hyperplane between
classification and regression. The experimentation was carried out using KDDCUP99 dataset. The overall detection accuracy was achieved at a rate of 97.4% and false alarm rate obtained was 1.1%. Yuan et al [105] have proposed the hypothesis test theory to SVM with the objective of getting increase in accuracy in the classification of the trained dataset. Here, the RBF kernel function was used and the results were compared with those of the sigmoid and polynomial kernels. The experimentation was carried out using KDDCUP99 dataset, the overall detection accuracy rate obtained is 93.7% and false alarm rate 0.11%, respectively.

The novel approach using a combined SVM with Ant Colony Networks (CSCOACN) was proposed for handling high dimensional data for designing the real-time Intrusion Detection Systems [106]. The overall detection accuracy and false alarm rate are 86.10% and 0.36%, respectively. Table 3.3 illustrates the comparison of methodologies using SVM in existence.

Table 3.3 Comparison of methodologies using SVM in existing techniques

<table>
<thead>
<tr>
<th>Author</th>
<th>Methodology</th>
<th>Dataset</th>
<th>Detection Rate</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kausar et al., 2012</td>
<td>Feature Transformation using PCA-SVM</td>
<td>KDDCUP99</td>
<td>99.46%</td>
<td>0.52%</td>
</tr>
<tr>
<td>Zaman et al., 2009</td>
<td>Fuzzy based Enhanced Support Vector Decision Function</td>
<td>KDDCUP99</td>
<td>99.66%</td>
<td>0.24%</td>
</tr>
<tr>
<td>Atefi et al., 2013</td>
<td>Support Vector Machine-Genetic Algorithm</td>
<td>KDDCUP99</td>
<td>99.49%</td>
<td>1.78%</td>
</tr>
<tr>
<td>Sagale et al., 2014</td>
<td>SVM-Naive Bayes</td>
<td>KDDCUP99</td>
<td>99%</td>
<td>6.8%</td>
</tr>
<tr>
<td>Wei et al., 2017</td>
<td>Grid based IGWO-SVM</td>
<td>Enrollment of students database in real-time applications</td>
<td>87.86%</td>
<td>5.29%</td>
</tr>
</tbody>
</table>
3.1.2 Background of Existence-Evolutionary Based

In this section, the background of the existence of evolutionary based approach is reviewed with the descriptive structure of Grey Wolf Optimization algorithm, which is effectively utilized in other applications. In addition, the specified characteristics of Grey Wolf Optimizer and their difference from the other optimization techniques were also discussed.

3.1.2.1 Grey Wolf Optimizer

Mirjalili et al [107] have developed the novel evolutionary-based approach known to be Grey Wolf Optimizer algorithm based on the real behavior of Western Wolf which pretends candid wildness in nature. In another term, Grey Wolf Optimizer algorithm has been notified as a new meta-heuristics that made a privileged point course of action for an optimization difficult in a proficient way. Two major components of any meta-heuristics required were 'Exploitation' and 'Exploration'. The Exploration focuses the search in a local region by knowing that current good solution is found in this region. While Exploitation generates diverse solutions in order to explore the search space on a global scale. There should be the healthy balance between these two approaches in improving the convergence of algorithm by later providing the best solution required during a selection process.

Generally, Grey Wolves are subspecies of Canis lupus family familiar for hunting, and encircling and with the attacking attitude towards the prey. Based on these characteristics, algorithm was developed to provide an optimal solution required for the specified problem. Grey Wolf Optimizer algorithm designed with hierarchical behavior had categories ‘alpha (Xα)’, ‘beta(Xβ)’, ‘delta(Xδ)’ and ‘omega(Xω)’. ‘Alpha(Xα)’-based search agent was a decision maker in organizing the other categories to find the location of the fittest optima. ‘Beta(Xβ)’-based search agent subordinated with the decision maker to organize the plan for providing the fittest optimal solution. The remaining search agents seem to be better options to subordinate the alpha and beta search agents but ‘omega(Xω)’ search agent were dominated by
‘delta ($X_{d}$)’ search agents to offer the third fittest optimal for the required solution.

The major characteristics of Grey Wolf algorithm relate to its heterogeneity i.e., enclosure of diversified set of elements made the experimental to high-level instructional practices while compared to homogeneous families of Ant Colony Optimization, Particle Swarm Optimization etc. Further, Grey Wolf Optimizer algorithm is designed to differ from the other Swarm Intelligence techniques by making them move in fairly huge together which appear to be live in the pack. Another characteristic is provide resilience to execute a stealthy operation towards its prey. The pheromone behavior of Ants, and Particle Swarm exhibiting the contact with their subordinates makes the duration exceed to reach over the search plane whereas, Grey Wolves break this kind of pheromone attitude by shortening the run time of the search. Therefore, this is the specified reason to project the Grey Wolf Optimizer(GWO) algorithm for the enhancement of the Support Vector classifier detection accuracy as the Proposed Hybrid approach.

A novel approach was proposed using a combined technique of GWO and SVM as the Improved Grey Wolf Optimizer-Support Vector Machine for providing an optimal feature solution from the real-time data events [108]. The subject of experimentation was perceived from Wenzhou Vocational College of Science and Technology. A set of 402 students majored in Digital Media Technology, out of which 195 students selected a graphic design as the major subject and the remaining 207 students selected the video production as the major subject. The objective of this experimentation was to update the current position of students from discrete search space. The IGWO-SVM obtained a classification accuracy of 87.36% and lifted the irrelevant and redundant features i.e. false leads as of 5.29%.

Another novel framework was proposed using a combined approach of GWO and ELM as the Improved Grey Wolf Optimizer-Kernelized Extreme Learning Machine for medical diagnosis [109]. The objective of this experimentation was to update the optimal feature extraction from the discrete search space. The experimentation was performed using two variant
datasets. The dataset one, called WDBC dataset was used for breast cancer prediction. The overall classification accuracy obtained by IGWO-KELM was 95.61%. The second dataset, Parkinson dataset, had an impact on the range of biomedical voice measurements. The overall classification accuracy obtained by IGWO-KELM is of 97.45%.

3.2 OVERALL SYSTEM ARCHITECTURE

The overall hybrid classification techniques are represented in Figure 3.4. The architecture diagram illustrates the three variant proposed hybrid classification techniques, namely, Multi-Layer Perceptron - Grey Wolf Optimizer (MLPGWO), Extreme Learning Machine-Grey Wolf Optimizer (ELMGWO), and Support Vector Machine - Grey Wolf Optimizer (SVMGWO) which resulted in the improvement of detection accuracy in classification of network traffic instances by reducing their false leads in a minimal learning time.

![Architecture Diagram](image)

Figure 3.4 Architecture of Proposed Hybrid Classification

The entire architecture of the proposed framework is classified into three phases. The first known as preprocessing module urged the collection of the instances from randomly assigned KDDCUP99 dataset. This dataset made use of the four main simulated attack categories and they are Type 1:
Denial-Of-Service attacks, Type 2: User-To-Root attacks, Type 3: Root-To-
Local attacks and Type 4: Probing attacks. Additionally, 'normal' traffic pattern
is imposed in this dataset. Each record in this dataset was persuaded with 41
features i.e. inspired by (34 continuous and 7 discrete features found) and
additionally one feature is to label the class as neither 'attack pattern' nor
'normal behavior'. The text form of dataset (.csv) Comma Separated Value
format can be converted into Attribute Relational File Format (.ARFF) as
extension file.

Phase two known as training the classifier module made use of feature
extraction process, which made automatic search of the best subset of
attributes from the randomly assigned KDDCUPP dataset by means of the
state space search method. The priority of this method was to control the
position of the discrete search space and locate the best combination for
improving the performance of selecting the attributes and by reducing the
over-fitting. This subjective was involved with the proposed hybrid
classification techniques by executing the experimentation from WEKA
simulator tool in MS Windows operating system background.

Phase three was involved the performance metrics module endeavored
to maximize the proposed hybrid classifier's detection accuracy by reducing
the misclassification of True Negatives i.e., False Positives. Some of the
parameters were taken into account for process evaluation were True
Positives (TP), False Positives (FP), ROC (Receiving Operating
Characteristics) curve, Precision, Kappa Statistics and Matthews Correlation
Coefficient) to build the model. The results expository of assigned parameters
for the proposed hybrid classification techniques were compared among
themselves in existence to show the best performance obtained by proposed
strategies in achieving the detection accuracy to a maximum extent.

3.3 CONCLUSION

This chapter has designated the proposed system architecture which
projects the internal functional flow in detail. The background of the existence
from machine learning techniques such as Multi-Layer Perceptron (MLP),
Extreme Learning Machine (ELM) and Support Vector Machine (SVM) is presented, describing their need to improve the detection accuracy from minimizing its false leads elaborately. Simultaneously, the background of existence from evolutionary based technique i.e., Grey Wolf Optimization (GWO) has been discussed, describing its characteristics in its search for the required fittest optimal solution.