Intelligent Intrusion Detection System Based on Hierarchical BiGRU and Advanced Feature Selection Techniques

Faisal Alamri

Department of Computer Science and Information Technology, Jubail Industrial College, Jubail, Kingdom of Saudi Arabia. amrifa@rcjy.edu.sa

Abstract

An Intrusion Detection System (IDS) is a product that can identify the activity across more networks and alert the stakeholders to potentially harmful threats or malicious actions. Thus, the advancement in improving IDS is introduced in developing an Intelligent IDS (IIDS) centred on Deep Learning (DL) techniques for feature extraction. Moreover, the IDS methods generally underwent the issues such as false positive rates, adaptability to change and compute resources that can lead to ineffective detection of potential threats to a network during the real-time conditions. Hence, the proposed model uses a multi-strategy Sparrow Search Algorithms (SSA) and a Variational Mode Decomposition (VMD) that uses feature selection to select only the most important features, which reduces dimensionality when processing the data, by reducing the prior false positive rates accuracy. Moreover, a Hierarchical Bidirectional Gated Recurrent Unit (BiGRU) architecture with residual gates and a Time-Aware Attention Mechanism is used. It allows the model to more effectively capture the complex temporal dependencies and contextual information found in the network traffic data. The method is tested using the CIC IDS 2018 dataset which a benchmark dataset is containing many of the modern attack types so that it can be confident in the impact of the proposed method. The model enhances real-time network security from an advanced feature selection approach and a complex hierarchical BiGRU architecture that is adaptable, scalable and efficient. The advantages of the proposed model are demonstrated by its performance metrics with its high accuracy, precision, recall, and F1-score, signifying that the model is able to effectively identify the different intrusion types while also minimizing false alarms.

Keywords: Intrusion Detection System (IDS), Multi-Strategy Sparrow Search Algorithms (SSA), Variational Mode Decomposition (VMD), Hierarchical Bidirectional Gated Recurrent Unit (BiGRU), CIC IDS 2018 dataset

1. Introduction

In the rapidly evolving landscape of cybersecurity, safeguarding digital assets against malicious intrusions has become a paramount concern for organizations worldwide [1, 2]. Intrusion Detection Systems (IDS) serve as critical components in network security frameworks, tasked with monitoring network traffic, identifying suspicious activities, and informing supervisors towards possible threats [3]. Traditional IDS approaches, often based on signature matching or anomaly detection, have played a vital role; however, they face significant challenges in adapting to the dynamic and sophisticated nature of modern cyber threats [4-6].

Conventional intrusion detection techniques, such as signature-based detection, rely heavily on predefined patterns of known attacks [7]. While effective against known threats, they struggle to identify novel or evolving attack vectors, leading to high false-negative rates. Signature-based detection resembles a "most wanted" list for cybercriminals; it runs traffic against a database of known attack characteristics, or "signatures." If there's a match, it generates an alert. Moreover, it is quick, effective, and straightforward to apply against known threats. However, the major flaw in signature-based detection is that it can only detect attacks that have a signature [8]. The fundamental weakness of signature-based detection yields an overwhelming number of false negatives because legitimate attacks [9], which have never been seen before, can present nuisances as a signature database constantly lags behind rapidly innovative cybercriminals. Anomaly-based methods, on the other hand, attempt to model normal network behaviour and flag deviations as potential intrusions [10]. Anomaly detection provides an alternative method that seeks to characterize the "normal" behaviour of a network and identify significant changes as intrusions and therefore more flexible than signature-based methods for identifying new and emerging attack paths [11]. The challenge of course is accurately defining "normal" as networks can be dynamic and complex, and normal changes even though they are legitimate appear as anomalies, which are likely to have a high false positive rate, causing an overabundance of alerts that may confuse and overwhelm the efforts of security analysts, resulting in alert fatigue [12, 13].

Although more adaptable, these methods often suffer from high false-positive rates due to the difficulty in accurately modeling complex network behaviors. Additionally, traditional systems tend to process high-dimensional data inefficiently, leading to increased computational overhead and latency, which are detrimental in real-time detection scenarios [14]. Furthermore, the static nature of many traditional IDS models limits their ability to learn from new data continuously. As cyber threats become more sophisticated, there is an urgent need for intelligent, adaptive systems capable of capturing complex temporal patterns and subtle features indicative of malicious activities [15].

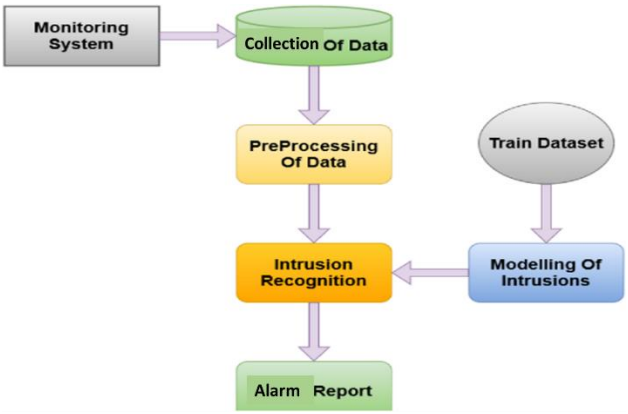

Figure.1 Flow of Intrusion Detection System [15]

To address these challenges, the proposed Intelligent IDS leverages advanced deep learning architectures combined with innovative feature selection techniques. The core of the model integrates a Hierarchical Bidirectional Gated Recurrent Unit (BiGRU) network with residual gates and a time-aware multi-head temporal attention mechanism. This architecture is designed to effectively capture temporal dependencies and contextual information within network traffic data, enabling more accurate and robust intrusion detection.

This dual approach to feature selection ensures that only the most informative features are fed into the classification model, significantly improving detection accuracy while reducing computational costs. The integration of VMD and SSA introduces a level of robustness and adaptability that surpasses traditional feature selection methods, enabling the system to generalize better across diverse network environments. The hierarchical structure allows the model to learn features at multiple temporal scales, capturing both short-term and long-term dependencies within network traffic sequences. Residual gates are incorporated to mitigate the vanishing gradient problem and facilitate deeper network training, ensuring that important features are preserved across layers. The time-aware attention mechanism further refines the model's focus by dynamically weighting different parts of the input sequence based on their relevance to the detection task. This attention mechanism allows the system to prioritize critical temporal segments, improving its sensitivity to subtle attack signatures.

Hence, the proposed IDS signifies a substantial progression in cyber security defence mechanisms. Through integrating a hierarchical BiGRU with residual gates and time-aware multi-head temporal attention, coupled with a robust multi-strategy feature selection framework based on VMD and SSA, the system aims to overcome the limitations of traditional IDS approaches. Its deployment on the CIC IDS 2018 dataset demonstrates its potential to deliver high accuracy, adaptability, and efficiency in real-world network security applications. As cyber threats continue to evolve, such intelligent and adaptive systems are essential for maintaining resilient and secure digital infrastructures.

1.1 Research Contribution

The main contribution of the proposed model is expressed below,

- To optimize the feature extraction and selection, for deriving the highly discriminative and relevant features from intricate network traffic, addressing the limitations on high-dimensional and noisy data in conventional IDS.
- To build a Hierarchical BiGRU architecture that can efficiently learn and model intricate temporal patterns in network data at various levels of abstraction, greatly enhancing the detection capabilities for complex, multi-stage, and subtle intrusions that are difficult to detect using traditional techniques.

- Incorporation of Time-Aware Multi-Head Temporal In addition to providing a critical degree of interpretability by emphasizing the most important features and time points that contribute to a detection decision, attention to explicitly consider the temporal ordering and inter-event timing allows the system to identify anomalies based on unusual sequences and timing.

- To attain better detection accuracy, lower false-positive rates with improved adaptability when it is compared with existing intrusion detection systems.

2. Literature Review

In the prevailing study [16] a hybrid sampling method (ADASYN + RENN) is applied. Feature selection is then performed using a combination of Random Forest and Pearson correlation. Subsequently, a Convolutional Neural Network (CNN) extracts spatial features, enhanced by average and max pooling. Bidirectional Gated Recurrent Units (BiGRU) extract long-range dependencies for comprehensive feature learning. Finally, Softmax classifies the data. Moreover, for data imbalance bias [17] the existing method uses a mixed sampling strategy for data resampling. It also includes denoising methods to manage any noise presented during sampling. Furthermore, feature redundancy is eliminated using a combination of Pearson correlation coefficient RF for feature selection, which improves the model's ability to capture crucial information from high-dimensional attack traffic. Likewise, in GIGA[18], a two-stage feature selection method, optimizes ID systems by reducing dimensionality and enhancing detection accuracy. It uses Gini impurity to pre-filter features, then a Genetic Algorithm to select the most relevant subset. This approach improves accuracy and reduces false positives/negatives for ML models. Further, an intrusion detection method [19] is suggested using ResNet-BiGRU with PSO-GA hyperparameter optimization. ResNet extracts parallel local features, which BiGRU then processes for long-distance dependencies. An attention mechanism is added to weight features, improving the comprehensive capture of important network intrusion characteristics and detection performance. The study has suggested [20] implied and explicit attention mechanisms. Implicit attention employs a self-supervised task that asks a model to perform rotation classification on an input image in order to focus the deep learning model's algorithms on certain helpful features that are present in the image. Explicit attention is implemented using a multi-headed self-attention mechanism in a Vision Transformer that learns to attend to important locations in the input image and map global features to semantic space while training. AN extensive experiment produced state-of-the-art results, producing the highest harmonic mean on three datasets.

Subsequently, the study [21] has presented a bidirectional GRU-based NIDS model with a hierarchical attention mechanism, treating intrusion as a time-series event. The study investigates the impact of varying lengths of prior traffic on performance. Experiments on the UNSW-NB15 dataset show the proposed model achieves a satisfactory detection accuracy of over 98.76%.Moreover,the MAML algorithm [22] is integrated to enable the BiGRU model's rapid adaptation and improved generalization to new tasks. Parameters are refined via meta-learning, summing losses across instances and using quadratic gradient descent. Empirical results show significant advancements, with goodness-of-fit decision coefficients of 0.926983 and 0.934452.Moreover,the TRBMA model [23]enhances ResNet18 with Temporal Convolutional Networks (TCNs), BiGRUs, and Multi-Head Self-Attention for comprehensive temporal feature learning. A 1D ResNet processes time-series data, and the AdamW optimizer improves convergence and generalization. On the CIC-IDS-2017 dataset, TRBMA achieves 98.66% accuracy in predicting malicious traffic. In the study [24] a hybrid sampling algorithm combining ADASYN and RENN is employed. Feature redundancy is tackled by combining RF and Pearson correlation analysis. A CNN extracts spatial features, which are further refined by fusing Averagepooling and Maxpooling. An attention mechanism assigns feature weights, reducing overhead and improving model performance.The prevailing study [25] improves object instance labeling and boosts robustness against adversarial attacks. It achieves significantly higher mAP, F1, and AUC scores, and an 8-point higher mAP against FFF or UAP attacks. This improvement is attributed to integrating both contextual and visual features from scenes, demonstrating that a simple context module can greatly enhance detector reliability.

ResNeSt-biGRU framework [26] updates IDS by using a dual-layered mechanism to exploit temporal and spatial features in network data, enhancing detection accuracy. Then, the PreIoT dataset is introduced (IoT network behaviors) for training and evaluating IDS models to identify intrusion traffic. The suggested scheme achieves average accuracies of 99.90% on N-BaIoT and PreIoT datasets, and 94.45% on UNSW-NB15.In the study[27] an ML-driven network intrusion detection model using Random Oversampling (RO) to handle the imbalance, stacking and embeddings and fitting based on the clustering and PCA, specifically for large with high imbalance datasets. The evaluation was based on evaluating the model using 3 benchmark datasets. The results indicated accuracy of 99.595% and 99.95% for the RF and ET models, respectively, on the UNSW-NB15.The prevailing study [28] has suggested IDS using the NSL-KDD dataset. It employs Recursive Feature Elimination (RFE) with

a Decision Tree for pre-processing and optimal feature selection. Various deep learning models (ANN, LSTM, BiLSTM, CNN-LSTM, GRU, BiGRU) were evaluated, with CNN-LSTM achieving the best performance: 95% accuracy, 0.89 recall, and 0.94 f1-score.The existing approach employs [29] a multistage feature extraction process for IoT security data. Autoencoders (AEs) first extract robust features from unstructured data. LSTM networks then analyze these features for temporal patterns of abnormal behavior. Finally, these refined features are fed into CNNs for classification, effectively leveraging each model's strengths. Similarly, the approach [30] uses a multistage feature extraction for IoT security data. Autoencoders (AEs) extract robust features, then LSTMs analyze these for temporal patterns of abnormal behavior. Finally, CNNs classify these refined features, effectively combining each model's strengths.

Correspondingly, the process [31] begins with Embedding Bilateral Filter (EBF) for image decomposition into base and detail layers, improving detection. An IR with attention-based Conv-ViT fusion model (efficiently captures shape and texture. Further, the Improved Shark Smell Optimization Algorithm (ISSOA) enhances feature selection and reduces redundancy. Finally, a Multi-scale Contextual Semantic Guidance Network (MCS-GNet) confirms strong image classification through incorporating multi-layer features, preventing data loss.The study has suggested [32] implicit and explicit attention mechanisms. Implicit attention utilizes a self-supervised image angle rotation task, focusing on task-relevant image features. Explicit attention, implemented via a multi-headed self-attention Vision Transformer, maps image features to semantic space during training. Extensive experiments on AWA2, CUB, and SUN datasets establish the projected mechanisms' effectiveness, achieving state-of-the-art harmonic mean performance. Moreover, an IDS architecture [33] comprised of CNN for extraction of spatial features and four variations RNN. The combination of these networks can more efficaciously detect and classify malicious cyberattacks in IoT network traffic than traditional methods. In the study[34] , an LSTM-SVAE extracts relevant features. Then, a BiRNN-HAID identifies intrusions with enhanced focus and memory. Finally, CE-CIA refines predictions using cognitive principles to balance sensitivity and specificity, reducing false positive.

This study [35] recommends an attention-based ZSL model for unseen class recognition. It uses a Vision Transformer-adapted attention mechanism to learn discriminative attributes from image patches. Experiments on AWA2, CUB, and SUN show state-of-the-art harmonic mean results, proving its effectiveness. Likewise MFEI-IDS [36] is an intrusion detection system that uses an FCN–Transformer for multi-level feature extraction, capturing local and global patterns to reduce manual engineering. Its inductive learning component maps features to robust class representations, improving generalization for unseen attacks. MFEI-IDS is designed for data imbalance and small-sample scenarios. Similar to that ,Packet-level adversarial traffic generation (PATG) [37] is projected to attack IIoT NIDSs by evading DL-based systems while maintaining domain constraints. A reversible abstract traffic representation allows effective, functional modification of original traffic. A packet-level generative adversarial network crafts adversarial traffic by learning benign data distribution and simulating evasion. Additionally, two defence schemes are designed to enhance system resilience against these attacks.The suggested Electricity Theft Detection (ETD) model [38] combines RFE, KNNOR, BiLSTM, and LogitBoost into a BiLSTM-LogitBoost stacking ensemble. This four-stage model achieved 96.32% precision, 94.33% F1 score, and 89.45% accuracy on SGCC data, outperforming benchmarks The study [39] has recommended a hybrid network classifier, combining improved residual network blocks and BiGRU. An improved autoencoder first reduces network data dimensionality. The processed data is then fed to the hybrid classifier for detection. Experiments on NSL-KDD and UNSW-NB15 datasets show the method achieves high accuracies of 93.40% and 93.26%, respectively.

2.1. Problem Identification

The research gaps obtained by the existing studies are expressed are depicted.

- There is no deeply embedded and optimized feature selection that specifically enables a BiGRU architecture to learn hierarchical temporal dependencies [18] [36].
- The prevailing studies have affected on adaptability and issues related to data, however, the explicit focus on robustness to concept drift and the enduring adaptation of hierarchical BiGRU models to the evolving patterns of an attacker is not provided extensively [22].
- While there is presence of attention mechanisms, there is limited attention focused on providing thorough, interpretable and actionable explainability for the complex, multi-level, temporal reasoning of a hierarchical BiGRU in the context of an IDS. [17] [21].

3. Proposed Methodology

Traditional IDS have significant weaknesses: signature-based approaches use known patterns to identify attacks (false-negative issues) and anomaly-based approaches may present numerous false positives when characterizing an unpredictable and dynamic network. Both approaches are time-inefficient when working with high-dimensional datasets, creating latency and computational overhead, and are static; they do not allow for continuous adaptation to evolving and possible future threats. Hence, the proposed IDS mitigates these shortcomings by combining a Multi-Strategy SSA-VMD for Feature Selection method with the Hierarchical BiGRU with Residual Gates and Time-Aware Attention Mechanism. Where, the Multi-Strategy SSA-VMD derived features allow for optimal feature extraction, reduced dimensionality, and dimmed noise ensuring a time-efficient dataset downstream. The Hierarchical BiGRU incorporates Residual Gates that learn complex, multi-scale restrict temporal patterns from the derived features, with the goal of improving the detection of new attacks and eliminating false positives. The Time-Aware Attention makes it possible to attend to the precise timings of events, which makes it possible to achieve greater anomaly detection with interpretability and the overall flow is given in the figure.2 which is depicted.

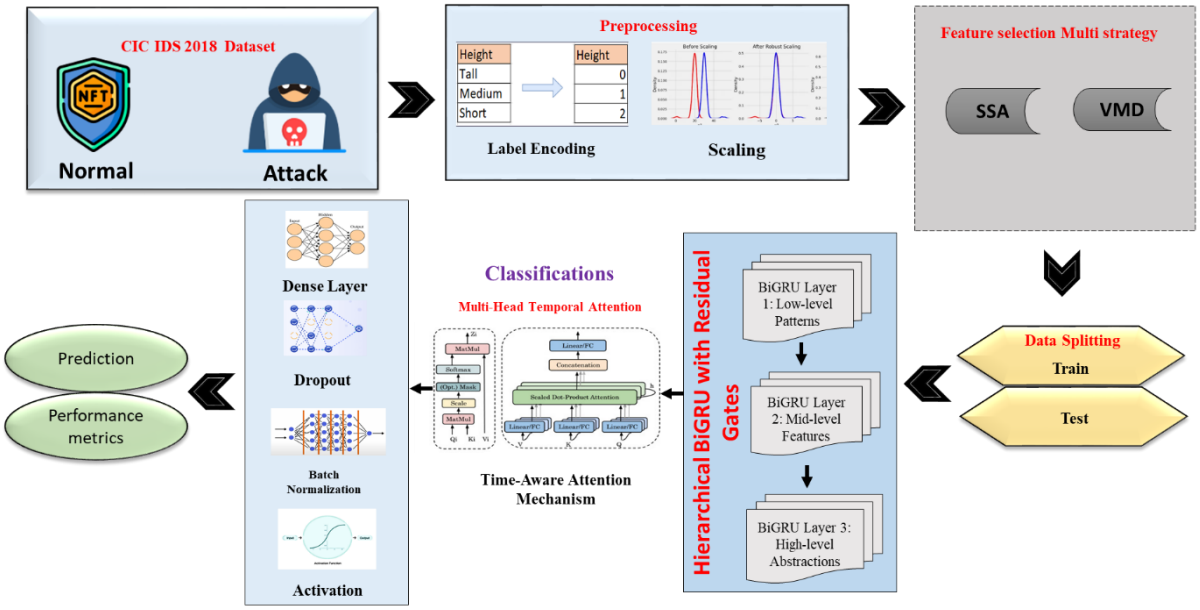

Figure.2 Overall Flow of Proposed Model

Figure 2 illustrates the complex process for intrusion detection starts with the CIC IDS 2018 Dataset, which is a strong example of a dataset with labelled instances representing "Normal" (or "Non-attack") or "Attack" types of network traffic. Then in pre-processing steps, including Label Encoding to make categorical features into their numerical equivalencies, and Scaling to lessen the impact of different contributions from each feature of the data, thereby constraining the dataset to be uniform, enabling the model to operate optimally.

After the pre-processing steps, a Complex Multi-strategy Feature Selection method is applied, which Combine Variational Mode Decomposition (VMD) with the Multi-strategy Sparrow Search Algorithms (SSA). VMD works by decomposing network traffic features into intrinsic mode functions, and in this method, despite the presence of noise or non-stationary data, then useful features are extracted using VMD. The Multi-Strategy SSA can intelligently search combinations of feature subsets and provides a different feature combinations that can help in reducing the dimensionality of the data, and makes more effective and efficient models. Prior to training, the refined dataset is appropriately split into Train and Test dataset to assess the model's performance on distinct data. The workflow is driven using the Hierarchical Bidirectional Gated Recurrent Unit (BiGRU) model which is comprised of three separate layers, sequentially where, Layer 1 focuses on low-level trends then return fundamental features; Layer 2 focuses on mid-level features as it relates to more complex relationships; and Layer 3 for high-level abstractions and deeper meanings.

Moreover, the architecture has a Time-Aware Attention Mechanism with Multi-Head Temporal Attention - allowing the model to focus on different time frames and their significance. The processed data then channels through the classification module where predictions are made using the learned representations from the BiGRU layers. Here, Residual Gates is used to overcome the vanishing gradient problem to train deeper networks while preserving important features between layers. The proposed time-aware attention mechanism help the model focus

even further by dynamically weighting and focusing on different parts of the input does sequence based on which time frames are relevant to the detection task (both normal and attack). Hence, model performance is evaluated using performance metrics.

3.1. Data Pre-processing

There are some specific steps that are important to take in order to change the raw data to a state that is ready to model. One of the major processes used is Label Encoding; that is the process of structured coding of categorical features to a numerical representation; allowing machine learning algorithms to use those categorical features. For instance, different regions or categories in the categorical variable could be assigned different integer values, allowing the rest of your model to process that in a numerical form. Moreover, Label Encoding can artificially place an ordinal hierarchy on nominal data that could possibly be misinterpreted by some models.

A further critical method is Scaling. Scaling generally means that data is normalized or standardized so that the features are weighted equally when calculated by the model. Scaling is especially important for distance-based algorithms, where one feature could dominate the others if they are a larger number. Normally scaling simply means moving the values of the data to some range (like 0-1 in the case of normalization) or to some fixed distribution centered at zero with unit variance (like standardization). Since large value features can dominate the learning process, scaling (normalizing or standardizing) the values of each feature prevents this from happening. Regardless, scaling data is a necessary and productive pre-processing activity to ensure that the proposed model can perform and generalize better.

3.2. Feature Selection- Multi Strategy SSA – VMD

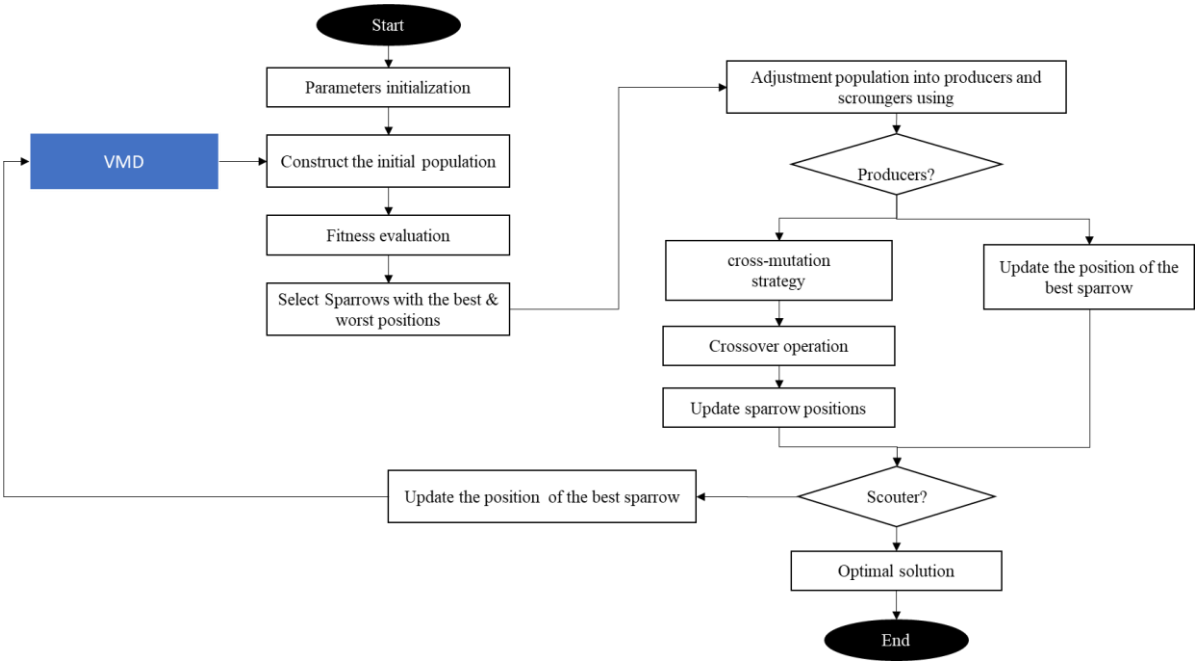

Figure.3 Illustration of Feature Selection Using SSA-VMD Approach

Figure.3 depicts the operation of a refined feature selection method in an Intelligent Intrusion Detection System (IDS) based on a Variational Mode Decomposition (VMD) strategy motivated by the sparrow's behavior. It starts with initializing parameters and building an initial population of potential solutions, followed by fitness evaluation to determine their performance in feature selection. Sparrows are classified according to their positions, specifying the top and low performers. The algorithm then separates the producers from the scroungers in the population, using a cross-mutation strategy and crossover operations to improve diversity among possible solutions. Best-performing sparrows' positions are updated in each iteration, with the algorithm monitoring scout roles to venture into new regions in the solution space. This proceeds until an optimal solution is determined, thus enhancing the efficiency and accuracy of the intrusion detection system via efficient feature selection.

Sparrow Search Algorithm (SSA) is plagued with slow convergence, local optima trapping, and loss of diversity, which results in suboptimal feature selection in Intrusion Detection Systems (IDS). These problems are addressed by Variational Mode Decomposition (VMD), which improves exploration by using multi-mode decomposition, enabling parallel search across various regions of the solution, hence improved convergence rate. Population

diversity is also ensured by cross-mutation approaches and adaptive parameter adjustment, which avoid premature convergence. Variational mode decomposition (VMD) is a totally non-recursive feature processing method based upon Wiener filtering and Hilbert transform. VMD is fundamentally an iterational variational technique which can automatically search the optimal solution of the variational problem: finding the minimum total bandwidth of each component. The original input feature can be decomposed via VMD into a series of band-limited discrete sub-features, that is, with most of the energy of each mode concentrated about the centred frequency.

Correspondingly, VMD is employed to expand the original feature into IMF components and determine the optimal variational mode solution, which is a fully non-recursive model.

For the best solution of the constrained variational model, the principle is as follows

The mode equation is described as an amplitude frequency feature F_g, where l in the input data and it is defined as follow in equation (1).

$$F_g(l) = o_g(l)\cos[\phi(l)] \tag{1}$$

Where $o_g(l)$ is the IMF vibration amplitude, the IMF frequency is the derivative of ϕ (l) and the VMD's principle is given as below:

1. The systematic feature of each $F_g(l)$) is obtained by Hilbert transform to acquire the frequency of single-sided defined as follows Equation (2)

$$\left(F(l) + \frac{y}{\pi l}\right) * F_g(l) \tag{2}$$

Here, the impact function is $F(l)$, and it is shown in Equation (3):

$$F(l) = \begin{cases} \infty & l = 0 \\ 0 & l \neq 0 \end{cases} \tag{3}$$

2. Add the e −jwkt and send the range of every modal component to the pertinent base band is illustrated as follows Equation (4)

$$\left[\left(F(l) + \frac{y}{\pi l}\right) * F_g(l)\right](e^{-yu_k l}) \tag{4}$$

3. The VMD principle is reformulated as an optimization problem with restrictions and it is expressed in Equation (5)

$$\underset{F_g u_g}{sml}\left\{\sum_{g=1}^{g} \left\|\partial_l \left[\left(F(l) + \frac{y}{\pi l}\right) * F_g(l)\right](e^{-yu_g l})\right\|_2^2\right\}, F.l \sum_g F_g(l) = f(l) \tag{5}$$

Here, cod $F_g(l)$ is represented in Equation (1), u_g is the center of frequency of each modal component, g is the iteration times, $f(l)$ is the original feature, $F.l$ is the bound term, * is the symbol of convolution calculation.

4. Include a penalty factor α and a refined Lagrange formula for solving Equation (5), transform the ordinary variation problems into unconstrained variation problems. A comprehensive Lagrange expression is derived as follows Equation (6)

$$\begin{aligned} Q(F_g, u_g, \lambda = \alpha \sum_{g=1}^{g} \left\|\partial_l \left[\left(F(l) + \frac{y}{\pi l}\right) * F_g(l)\right](e^{-yu_g l})\right\|_2^2 \\ + \left\|f(l) - \sum_{g=1}^{g} F_g(l)\right\|_2^2 + \left\langle \lambda(l), f(l) - \sum_{g=1}^{g} F_g(l)\right\rangle \end{aligned} \tag{6}$$

Here the 2nd penalty factor is α which is used to adjust in order to get the completeness of the VMD.

5. After transformation, the constrained variational model's optimal solution expression is derived as follows Equation (8):

$$\widehat{F_g^{n+1}}(u) = \frac{\hat{f}(u) - \sum_{i \neq g} \hat{F}_l(u) + \frac{\lambda(u)}{2}}{1 + 2\alpha(u - u_g)^2} \tag{7}$$

$$u_g^{n+1} = \frac{\int_0^{\infty} |\hat{F}_g(u)|^2 du}{\int_0^{\infty} |F_g(u)|^2 du} \quad (8)$$

The VMD approach, a non-recursive feature decomposition method to decompress a feature into intrinsic mode function (IMF) components. The fundamental concept is to determine an optimal variational mode solution by solving a constrained optimization problem. This is done through multiple steps: first, computing the analytic feature of every IMF component via a Hilbert transform (Equation 2 and 3); second, shifting each modal component's spectrum to its corresponding baseband (Equation 4); third, defining the VMD principle as a constrained optimization problem (Equation 5); and lastly, solving for this problem by adding in a penalty factor α and a more sophisticated Lagrange formula (Equation 6) to obtain optimal solutions for both the IMFs (Fg) and their centre frequencies (ug) (Equations 7 and 8). Mostly, VMD tries to decompose a feature into a collection of band-limited IMFs, with each IMF containing a compact frequency representation at certain central frequency.

Therefore, the SSA algorithm based VMD parameter optimization framework, which can reach the optimized parameters more rapidly and precisely. The SSA algorithm optimizes by simulating a target foraging behaviour of the sparrow population and has high optimization capability and fast convergence speed. The VMD decomposes the unique feature to K modal components. Uncertainty the modal element is clean and fewer noise-influenced, the information corresponding to the feature of the original time series data will be more evident and have smaller sample entropy. The proportion of sample entropy is used as a total evaluation index to develop the objective function. Hence, the process of optimizing the VMD parameters is changed to pursue the ratio of minimum sample entropy defined by SSA.

3.3. Proposed Classification Model- Hierarchical BiGRU with Residual Gates and Time-Aware Multi-Head Temporal Attention

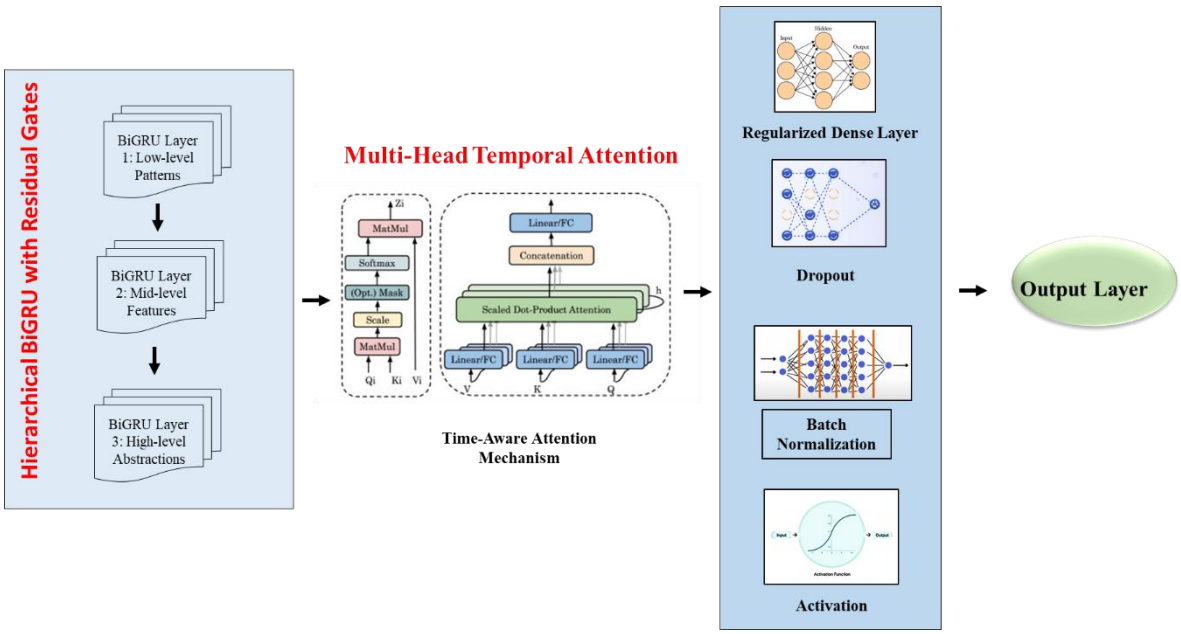

Figure.4 Proposed Hierarchical BiGRU with Residual Gates and Time-Aware Multi-Head Temporal Attention Model

Figure.4 shows the proposed model features a unique form of intrusion detection based on the Hierarchical BiGRU architecture. The core structure of this architecture include three BiGRU layers that operate in sequence, Layer 1 models low-level patterns, Layer 2 models mid-level features interactions, and Layer 3 extracts high-level abstractions to get a complete understanding of the sequential data, while processing information in both forward and backward directions. This is significant because the BiGRUs also have Residual Gates in all layers which helps eliminating the vanishing gradient problem and allows effective training of deeper networks composed of BiGRUs. Also aware of the importance of time, the architecture also implements a Multi-Head Temporal Attention method that computes the weighted sum of its input features through three steps. They are matrix multiplication, Softmax normalization and scaled values and concatenate the different heads together to diversify the effects of attention across the model, for example the head can look at the input data with more time

consideration - this is done prior to a final linear projection. The processed output is then fed into a Regularized Dense Layer, which incorporating techniques like dropout and batch normalization to prevent over fitting that is followed by an Activation Function to introduce non-linearity. The final Output Layer produces predictions, classifying network traffic as "Normal" or "Attack." Hence, the hierarchical structure is proposed for nuanced data understanding, the bidirectional nature of BiGRU for comprehensive temporal context, the adaptive focus provided by the temporal attention mechanism, the training stability offered by residual connections, and the overall robustness achieved through comprehensive regularization techniques. This integrated design significantly enhances the model's capability for accurate and effective intrusion detection in complex network environments.

Correspondingly, the streamlined version of LSTM is GRU, which is similar to LSTM and makes the model to be simple. Here, GRU combines forget and input gates to be an "update gate". Where, the unit and hidden states are merged together. Moreover, the hidden state is evaluated for the current status and hence the h˜ t is estimated and the reset gate is Q_t is provided in the equation (9).

$$Q_t = \sigma\left(W_Q[h_{t-1}, x_t]\right) \tag{9}$$

Moreover, the sigmoid function is employed to constrain the gate value between [0, 1]. When the reset gate approaches 0, the current candidate value h˜ t disregards the past hidden state ht−1 and is calculated with the present input xt. This essentially allows the hidden state to discard some unnecessary information contained in the future:

$$\tilde{p_t} = \tanh\left(W_p[r_t * p_{t-1}, x_t]\right) \tag{10}$$

After computing the candidate value h˜ t, the update gate determines how much information of the previous hidden state can be passed onto the current hidden state. The update gate zt is defined as:

$$o_t = \sigma(W_o[p_{t-1}, x_t]) \tag{11}$$

The recent hidden state ht at the present time is computed as:

$$p_t = (1 - z_t) * p_{t-1} + z_t * \tilde{p_t} \tag{12}$$

Moreover, the GRU neural network employs a recurrent structure for encoding and decoding information about its input, and the NN uses the information from the previous moment at the point and does not use a future moment state, so it just can't improve accuracy of prediction. The BiGRU network employs a future layer and can therefore have the sequence of data predicted in the opposite direction to fix this issue.

The network has 2 hidden layers to gather information in the past and the future, and both are connected to the same output layer. In BiGRU:

$$\vec{p_t} = GRU(x_t, \vec{p}_{t-1}) \tag{13}$$

$$\bar{p_t} = GRU(x_t, \mathrm{p} \ \leftarrow_{t-1}) \tag{14}$$

$$p_t = u_t p_t + v_t p_t + b_t \tag{15}$$

In the hidden layers between, the nonlinear mapping function for the input data. u_t and v_t are weights for the forward hidden layer, and state hE t and the reverse hidden layer state ← ht respectively. Moreover, b_t is the hidden layer state at time t of the corresponding compensation.

Pass the hidden state H to the attention mechanism, and fuse the BiGRU output y ' i−1 at time i to compute the attention weight of each h in the current time. Y' 0 is taken from the final hidden state hm of the encoder. The calculation of the weights is given in the equation (16):

$$k_{ij} = V^{\mathrm{T}} \tanh\left(Wp_j + Uy'_{i-1} + bias\right), i \in \mathbb{N}, 1 \le i \le n, j \in \mathbb{N}, 1 \le j \le m \tag{16}$$

Where V ∈ R T, W ∈ R T×k and U ∈ R T×k are parametrical weight matrices.

Normalize eij by using the Soft max function as described in Equation 10.

$$U_{ij} = \frac{\exp(e_{ij})}{\sum_{j=1}^{m} \exp(e_{ij})} \quad (17)$$

The e_{ij} is the feature correlation coefficient and it is scaled by its consistent hidden state value hj to produce a weighted cell state C ∈ R n×k ,which reflects the varying contributions of the features.

The hidden state Y 0 is transformed according to Equation (9):

$$J'_t = f_2(c_t, J'_{t-1}) \quad (18)$$

f2 represents the BiGRU network unit in the decoder, where the input is no longer the original cell state, then the weighted data differential according to the influence size. The encoder part can calculated and it is extracted by the contribution rate of each input data with the attention mechanism, resulting in improved prediction performance. Moreover, the algorithm for time aware attention mechanism is depicted.

Algorithm :Time aware Attention Mechanism

1: Coverd = encoder.bigru.output
2: for v = 1 to TargetLength do
3: for y = 1 to InputLength do
4: $e_{vy} = MIP(y'_{v-1}, T_y)$
5: end for
6: $B_v = \frac{\exp(e_{vy})}{\sum_{y=1}^{Target Length} \exp(e_{vy})}$
7: $k_v = B_v \odot$ Covered
8: $y'_v = f_2(k_v, y'_{v-1})$
9: end for
10: return L'

Moreover, the model can concurrently attend to data from multiple representation subspaces at different positions thanks to multi-head attention.

$$MultiHead(q,k,v) = Concat(h1,\ldots,hh)wo \quad (19)$$

$$\text{Where } hi = Attention(qw\ q_i, kwk_i, vwv_i) \quad (20)$$

The parameter matrixes are projected as wq $i \in R\ dmodel \times dk$, wk $i \in R\ dmodel \times dk$, $wvi \in R\ dmodel \times dv$ and $wo \in R\ hdv \times dmodel$.

Averaging stops it with a single attention head. Each layer of the encoder and decoder has a fully connected feed-forward network, which is implemented individually and identically to each position, with the exception of attention sub-layers. It consists of a ReLU activation sandwiched between two linear transformations.

$$FFN(x) = \max(0, xW1 + b1)\ W2 + b2 \quad (21)$$

On the other hand, layer-to-layer linear transformations are comparable across different states with multiple parameters.

Therefore, several existing IDS mainly focus on feature extraction or classification and uses static approaches. These approaches are failed in adapting the intricate occurred on the dynamic network traffic that results in loss of key temporal information. However, in the proposed model a feature selection module and a classification model are the two complementary stages of the proposed Intelligent IDS. Besides, raw network traffic is intelligently processed during the feature selection phase, which is based on Multi-Strategy SSA and VMD. The Multi-Strategy SSA then extracts robust features from these IMFs, guaranteeing thorough capture of temporal dependencies, after VMD first breaks down the complex network data into both noise and appropriate patterns. The Hierarchical BiGRU with Residual Gates and Time-Aware Multi-Head Temporal Attention classification model is then fed these refined features. With the help of residual connections for stable training, this DL component learns high-level and fine-grained temporal patterns from both past and future contexts using a multi-layered Bidirectional GRU. Significantly, the model's Time-Aware Attention mechanism explicitly incorporates timing information for highly accurate and context-aware intrusion detection, enabling it to selectively focus on the most important time steps and aspects of the network data. Subsequently, the model resolves the limitations

occurred in the existing methods and bridges the gap between advanced feature extraction and dynamic classification in finding network anomalies and the algorithm for the Multi-Strategy SSA-VMD + Hierarchical BiGRU for Intrusion is depict.

Algorithm: Multi-Strategy SSA-VMD + Hierarchical BiGRU with Time-aware Attention Mechanism for Intrusion

- **Input:** CIC IDS 2018 & BoTNeTIoT datasets
- **Output:** "Normal" or "Attack" classification

1. **Pre-processing:**
   Load, label encode, scale, and split data.
2. **Feature Selection (SSA-VMD):**
   Initialize sparrows; iteratively evaluate and update positions.
   Decompose features via VMD; optimize parameters using entropy.
   Select optimal feature subset.
3. **Classification (Hierarchical BiGRU):**
   Three BiGRU layers extract temporal features (low to high level).
   Use residual gates and skip connections.
   Apply multi-head time-aware attention for focused temporal info.
   Pass through dense layer with dropout & batch norm.
   Output prediction with activation.
4. **Evaluation:**
   Measure accuracy, precision, recall, F1-score.
   Return classification result.

Table.1 Hyperparameter Specifications

Parameters	Values
Batch_size	1024
Timesteps	3
Activation	Relu
Monitor	'Val_loss'
Factor	0.5
Patience	5
Verbose	1
Min_lr	1e-6
Validation_split	0.2
Epochs	50
batch_size	32
class_weight	class_weights
verbose	1

4. Results and Discussion

This section discuss in detail about the dataset usage, performance metrics, performance analysis and its comparison with the existing studies.

4.1 Dataset Description

The CIC-IDS2018 dataset is recognized as a popular dataset for research in intrusion detection as it presents better attributes than the previous benchmark datasets including KDD Cup 99, NSL-KDD, and the data from CIC-IDS2017. Datasets from years and decades ago, older benchmarks are often too dated as they do not contain the relevant attack patterns of today, they are often made of repeated records and over-anonymized for real world applicability in evaluating an advanced IDS. Unlike older datasets, CIC-IDS2018 was carefully produced to represent current network traffic and current/relevant attack scenarios that included brute-force, Heart bleed,

Botnet, DoS, DDoS, web attacks including infiltration. The CIC-IDS2018 dataset is composed of a full set of flows labelled as benign and malicious with 79 features, extracted using CICFlowMeter-V3, and totaled more than 160,000 flows that are ideal for representing real-world scenarios. As a well-documented, and relatively recently produced dataset, it allows for researchers to produce and thoroughly test an intrusion detection system with a greater sense of fidelity and relevance to present online attack threats [40].

Subsequently, the BoTNeTIoT dataset is aimed to evaluate the anomaly detection in IoT that include a varied range of network traffic conditions, which contains benign and malicious activities. Moreover the dataset has a significantly unbalanced class distribution; some attack types, like port scanning and DDoS, are overrepresented relative to others, which makes it difficult to train models efficiently without favouring the more common classes. The dataset also poses difficulties because different types of attacks have similar attack patterns, making it difficult to distinguish between them and possibly resulting in misclassifications. Moreover, functioning with the BoTNeTIoT dataset necessitates the use of advanced methodologies for effective intrusion detection in IoT systems, as this overlap calls for complex feature extraction and classification techniques to improve accuracy in identifying and mitigating threats. The dataset has 27 features, which were derived from packet captures within 10-second time windows. These features encompass several statistical attributes like mean, variance, count, magnitude, radius, covariance, and correlation coefficient [41].

4.2 Performance Metrics

When estimating the DL models for classification, typically confusion matrix based metrics will be used. Besides, confusion matrix shows the relationship between the predicted class labels of the given input image and the ground truth class label for the image. These conditions can be captured by the metrics: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Similarly, performance metrics for the assessment of an image classifier can be measured by accuracy, recall, precision and F1 score. In fact, the accuracy is the rate of accurate predictions made by the model and estimated as the rate of accurate predictions taken by the total overall predictions. Then, recall is a rate utilized to measure the model capability of detecting all the positive samples, and it is evaluated as the rate of true positive detections made by the actual total positive samples. Another metrics is precision, where the capability of the model is measured of finding true positive instances and is considered by the rate of positive predictions by the total overall positive predictions. The F1 score, is defined while the recall and precision rates are calculated using the harmonic mean, then this is used to weight recall and precision during conflict. The formula description of these metrics are given as follows:

$$Accuracy\ (A) = \frac{(TP+\ TN)}{(TP+\ FP+ FN+\ TN)} \tag{22}$$

$$Recall\ (R) = \frac{TP}{(TP+\ FN)} \tag{23}$$

$$Precision\ (P) = \frac{TP}{(TP+\ FN)} \tag{24}$$

$$F-1\ Score\ (F1) = \frac{[2\times (P\times R)]}{(P+R)} \tag{25}$$

4.3 Performance Analysis

This section explains in detail about the performance of the proposed model and with its description.

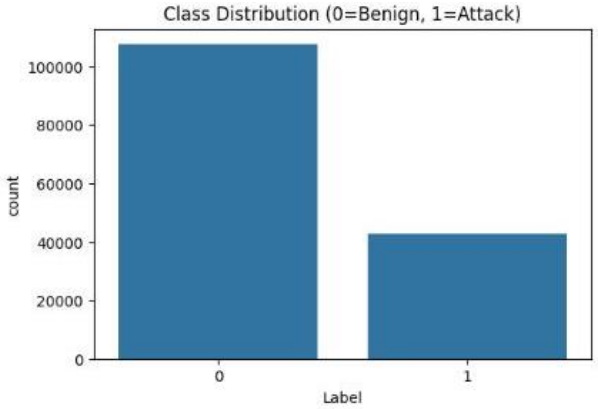

Figure.5 Class Distribution of Proposed Model

The bar chart shown in the figure.5 depicts the distribution of the assignments in the data set, exhibiting clear class imbalances. The x-axis indicates "label" (0 is "Benign" while 1 is "Attack") and the y-axis indicates the total instances. The bar chart displays class imbalances in that a large number of the instances are labelled "Benign" (about 100,000) and a substantially lower label count as "Attack" (approximately 40,000). Recognizable class imbalances in the data, like what was illustrated here, are important for two reasons; first, they impact in what way ML models are trained and subsequently evaluated. This is important because if they are trained on a data set that has class imbalances, it is likely the model can perform well on the majority label but poorly on the minority label.

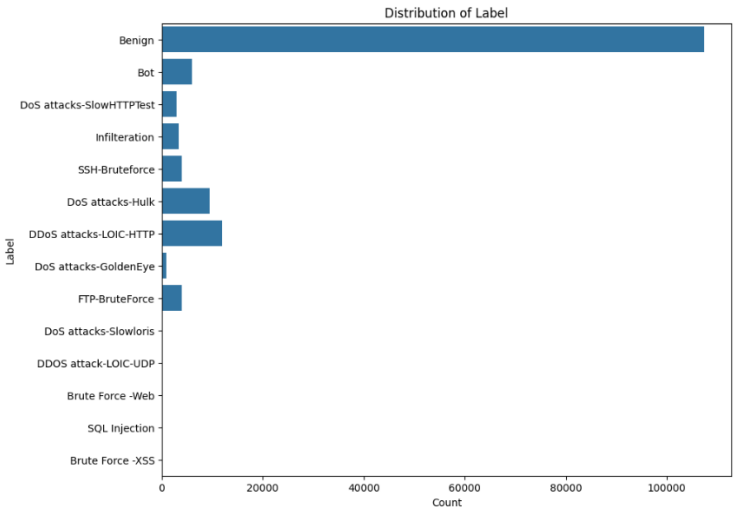

Figure.6 Multi-Class Distribution of Proposed Model

Figure.6 illustrates the frequency of categories within what is likely a network traffic or network intrusion detection dataset, with each category representing a network activity or attack type. The x-axis lists those categories: "Benign" (representing standard network traffic), "Bot" (traffic coming from a botnet), several different Denial-of-Service (DoS) attacks, "SlowHTTPTest," "LOIC-HTTP," and "GoldenEye," "Infiltration" (access attempts), a variety of "Brute Force" attack contexts ("Brute Force - Web," "Brute Force - XSS"), and "SQL Injection" (attempts to exploit a database). The y-axis details the counts for the labels where the height of each bar indicates the frequency of that label. A compelling detail from the bar chart is the preponderance of the "Benign" category which contains far greater counts compared to all of the other labels. This creates a large class imbalance for benign traffic in the dataset. The much more limited counts of the other attack types should be considered, especially during model training to ensure no strong bias toward the majority classes.

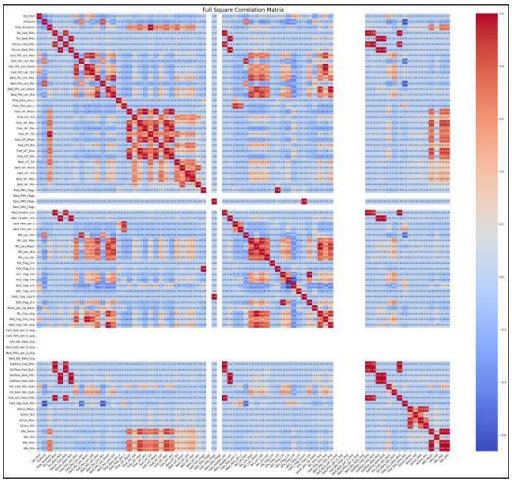

Figure.7 Correlation Heat map of Multi-class Classification Model

Figure.7 demonstrates an extensive correlation matrix to visualize the relationships between many of the features in the dataset. The x-axis and the y-axis show the same features, which make the comparison for any two features easy. The colors on the scale range from blue, which indicates negative correlation, to red, which indicates positive correlation, and the darker the blue or red, the stronger the correlation is. The diagonal elements represent that each feature is correlated to itself at 1.0. The off-diagonal elements reveal the correlation coefficients between different features: +1 are values of strong positive correlation, -1 are values of strong negative correlation. The correlation matrix provides important insights about features that may be redundant because they are highly correlated, and it often means we can delete one of these highly correlated features to simplify the model. In addition, the correlation matrix also provides important value in identifying features which contribute in a similar manner to predicting outcomes, improving the informed decision regarding the feature selection and feature engineering process. Furthermore, many features being strongly correlated, demonstrate complex relationships with the data, including interdependencies, potentially transforming the performance of the model.

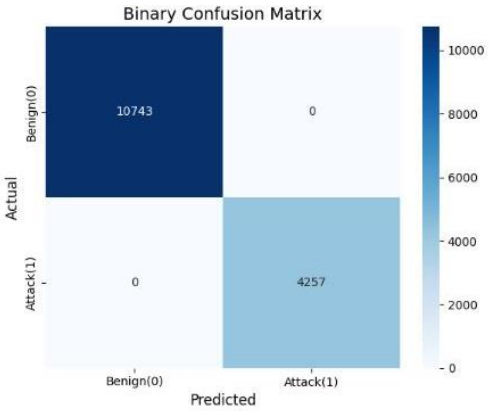

Figure.8 Confusion Matrix of Binary Classification Model for CSE-CIC-IDS 2018 Dataset

Figure.8 shows a binary confusion matrix employed to measure the performance of a model that classifies between two classes: "Benign" (0) and "Attack" (1). There are 10,743 true positives for well-classified benign instances in this matrix, no benign cases misclassified as attacks (0 false negatives), and no attacks classified as benign (0 false positives). There are also 4,257 true positives for attack instances. This shows that the model performs with perfect classification without any errors in separating benign and attack instances.

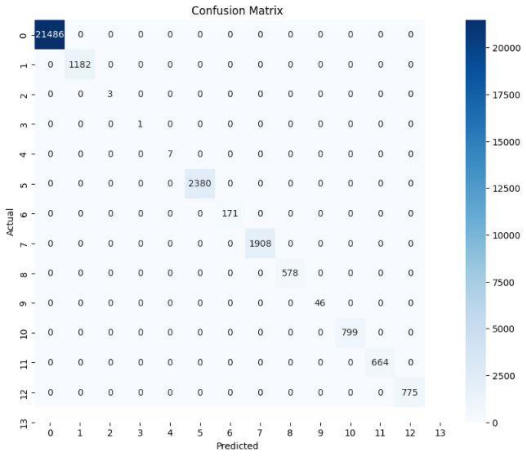

Figure.9 Confusion Matrix of Multi-Class Classification Model for CSE-CIC-IDS 2018 Dataset

The confusion matrix in figure.9 represents a deeper view of the classification performance of the model for 14 classes (0 to 13), with rows as actual classes and columns as predicted classes. The diagonal values indicate the number in which each class was correctly classified (21,486 for class 0, 2,380 for class 5 etc.) with larger being more accurately classified. The actual misclassification are in the off-diagonal values. For class 1, it had 1,182 correct classified, but many more were classified as class 0 where they could be class 1. For class 7, the model misclassified an instance as class 5 171 times, but had positive 1,908 correct classified instances in class 7 results. The significant number of diagonal values demonstrates overall good accuracy of the model. In contrast dive off-diagonal values likely shows instances of misclassification going forward that could be worth looking at again as consideration for future model construction or investigation. The color ranges from light blue to dark blue to show count sizing, with darker showing larger counts. This color matching is a good visual aide for rapidly seeing the correctly classified and misclassified instances. The error analysis is obtained by observing the data presented in the columns. This shows that the multi-class has less values and certain columns have more values.

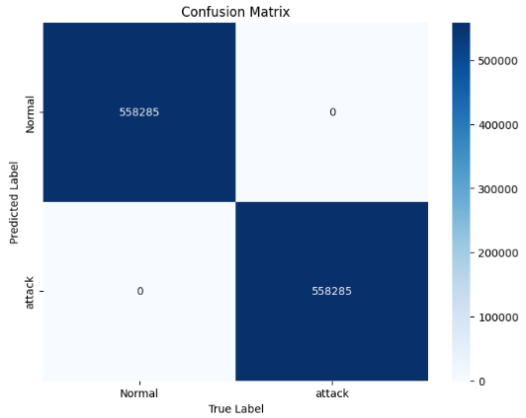

Figure.8 Confusion Matrix of Binary Classification Model for BoTNeTIoT-L01-v2 Dataset

The binary confusion matrix presented here reflects that the classification model works extremely well with perfect accuracy. It reflects 10,743 true negatives (correctly identified benign instances) and 4,257 true positives (correctly identified attack instances), without any false positives or false negatives. This brings about an accuracy rate of 100%, reflecting the model's capability to perfectly classify between benign and attack instances. Such outcomes indicate a successful model; nonetheless, further examination on measurement such as precision, recall, and F1-score would show us more about its general performance and stability.

Therefore, the implementation of SSA-VMD method make the datasets as a clean, linearly separable with minimal overlap during the feature distributions. It identifies a highly discriminative features by removing the irrelevant or redundant values. This can make the model to select the features precisely capture the variations among classes and hence the perfect scores (100%) are achieved. Moreover, the conventional security models requires more time for calculating a large data in detecting the risks. Whereas, a bad user may requires an unauthorized access to data for obtaining sensible information and changes the data could be affects the user negatively.

4.4 Comparative Analysis

This section explain in detail about the comparison of existing models with proposed model and it is discussed in detail.

Table -2 Comparative Analysis of Existing Models with Proposed Model for CSE-CIC-IDS 2018 Dataset[42]

Algorithm	Accuracy	Precision	Recall	F1 score
AdaBoostM1	0.862	0.961	0.883	0.92
Bagging	0.988	0.996	0.989	0.992
Bayes Net	0.965	0.968	0.989	0.972
CDT	0.987	0.997	0.988	0.993
DDTNB	0.986	0.996	0.988	0.992
IB1	0.978	0.986	0.988	0.987
J48	0.988	0.977	0.988	0.993
KStar	0.979	0.989	0.987	0.988
Proposed Model	**1**	**1**	**1**	**1**

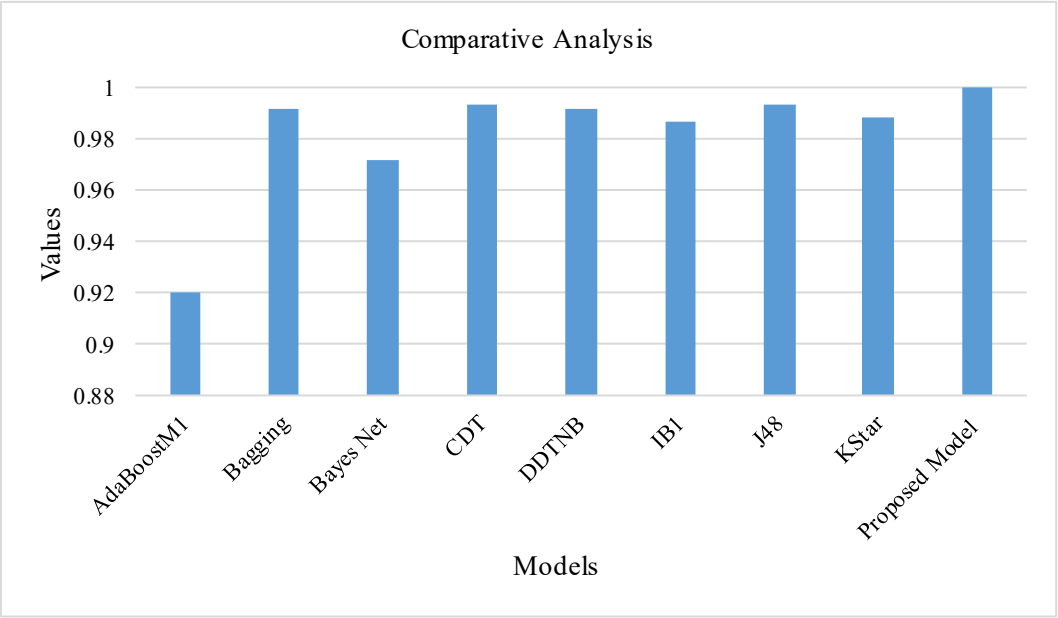

Figure.10 Illustration of Comparison of Performance Metrics of Proposed Model with Existing Models for CSE-CIC-IDS 2018 Dataset [42]

Table.2 and Figure.10 shows that the "Proposed Model" is evaluated and achieved a perfect score of 1.0 for Accuracy, Precision, Recall, and F1 score. Bagging, CDT, and J48 performed well and acquired an accuracy scores of 0.988 and F1 scores of 0.992 and 0.993. The Bayes Net, DDTNB, IB1, and KStar algorithms all performed well with an accuracy of 0.965 to 0.986, F1 score of 0.972 to 0.988. AdaBoostM1 performed the lowest scores in all categories, and lowest Accuracy of 0.862 and F1 score of 0.92. It would be beneficial to build a more robust analysis and demonstrate that the proposed model is high by comparing against a wider variety of ML algorithms.

Table.2 Comparison of Accuracy among Existing and Proposed Model for CSE-CIC-IDS 2018 Dataset [43]

Reference	Accuracy	Precision	Recall	F1 score
Autoencoder [44]	97.70%	98.00%	98.00%	98.00%
DSSTE [45]	96.99%	97.346%	96.97%	97.04%
]LSTM [46]	98.38%	97.79%	97.74%	97.76%
DNN [47]	95.79%	95.38%	95.79%	95.11%

TCN-LSTM [48]	97.77%	97.94%	97.53%	97.73%
PCA-DNN [49]	97.73%	99.97%	97.42%	98.68%
MLP-BP (Binary Classification) [50]	96.25%	98.75%	96.80%	97.76%
MLP-PSO(Multi-class Classification) [50]	98.97%	99.98%	98.80%	97.76%
Proposed (Binary Classification)	**100%**	**100%**	**100%**	**100%**
Proposed (Multi-class Classification)	**100%**	**100%**	**100%**	**100%**

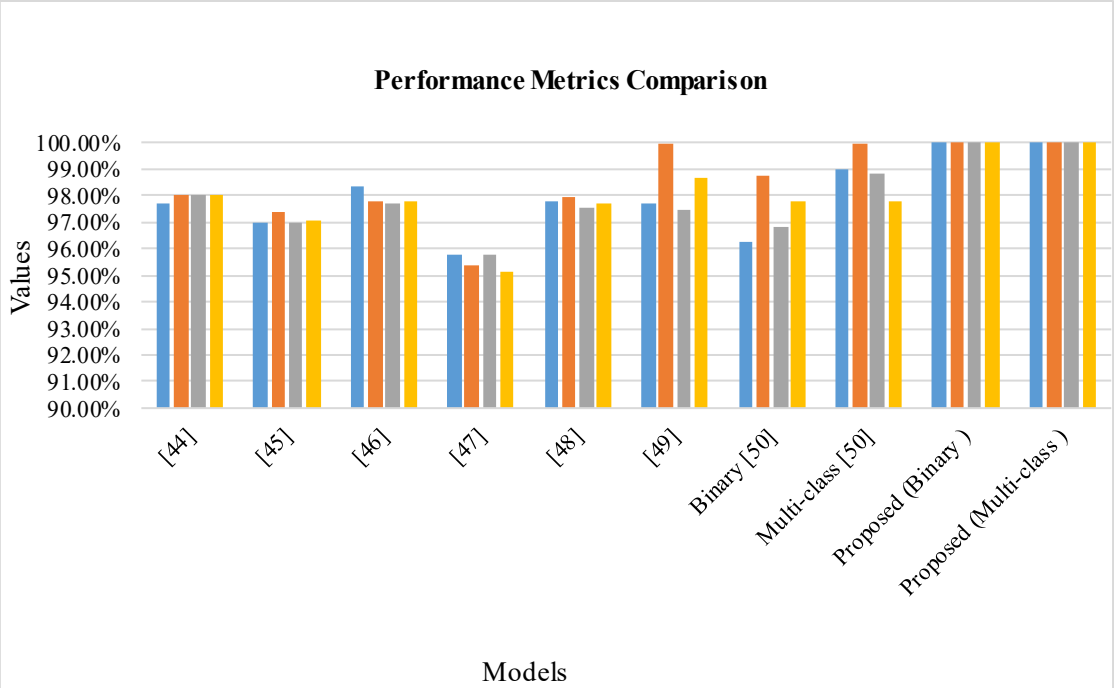

Figure.11 Comparison of Accuracy among Existing and Proposed Model for CSE-CIC-IDS 2018 Dataset[43]

Based on the table.2 and figure.11 the "Proposed (Binary Classification)" and "Proposed (Multi-class Classification)" models exhibit perfect scores of 100 percent for every metric as follows: Accuracy, Precision, Recall and F1 score. Of the other models shown, the "MLP-PSO (Multi-class Classification)" model has the next best scores of 98.97% for Accuracy and 99.98% for Precision. Whereas LSTM and PCA-DNN had high scores, with 98.38% and 97.73% for Accuracy and their F1 scored were also very high. The other reference models, including Autoencoder, DSSTE, DNN and TCN-LSTM and, MLP-BP (Binary Classification) demonstrated high performance scores where Accuracies ranged from 95.79% to 97.77%. The lowest achieved score of the reference models was from DNN, which had a low Accuracy of 95.79% and an F1 score of 95.11%. Overall the proposed scores lead all others referenced and benchmarks models in all measured categories across all models.

Table-4Comparative Analysis between LSTM and Proposed Model for CSE-CIC-IDS 2018 Dataset [51]

DL and ML Model	Accuracy
LSTM+AM [52]	96.2%
Light GBM+HBGB [53]	97.5%
CFBLS and BLS [54]	97.46%
LSTM [55]	99%
Proposed	**100%**

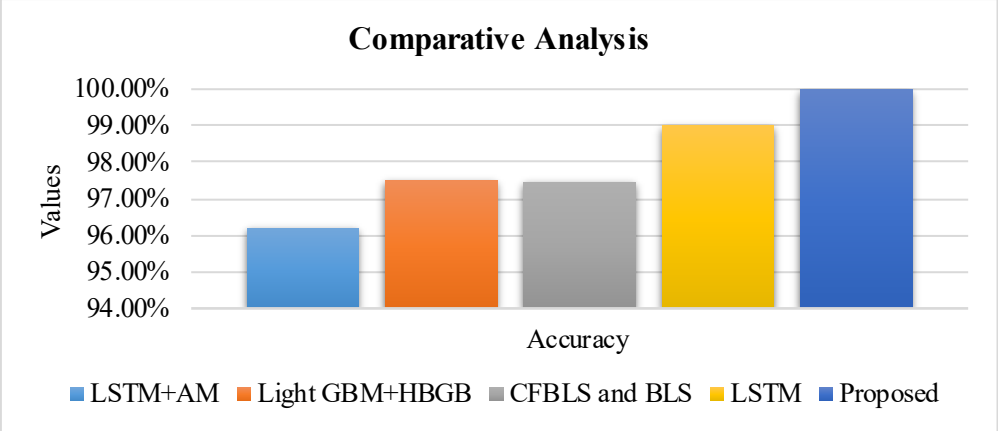

Figure.12 Graphical Representation on Comparison between Existing Models and Proposed Model [51]

Table-3 and figure.12 shows that the "Proposed" model yielded an overall accuracy of 100%, ranking ahead of all other Deep Learning (DL) as well as Machine Learning (ML) models considered. The LSTM model scored next to it, with a high accuracy of 99%. Light GBM+HBGB & CFBLS (then BLS) scored accuracies of 97.5% & 97.46% respectively. The LSTM+AM model yielded the next highest mention of accuracy, with a score of 96.2%. Overall, the results indicated that the proposed model is the most worthy for this specific purpose; 100% was attainable against other high-achieving models being just below their expected error range.

Table-4 Comparative Analysis between DNN and Proposed Model for CSE-CIC-IDS 2018 Dataset [56]

Model	Accuracy
DNN [57]	95%
DNN [56]	90.25%
Proposed	**100%**

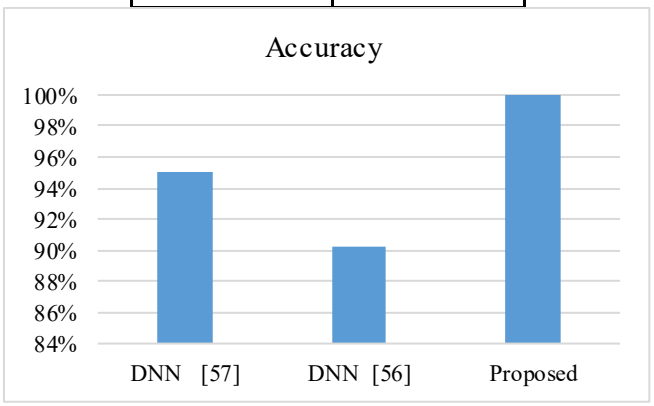

Figure.13 Comparative Analysis between DNN and Proposed Model for CSE-CIC-IDS 2018 Dataset [56]

Table-4 and Figure.13 demonstrates, the accuracy of the "Proposed" model has a perfect 100%, which is currently better than the other two DNN models. The DNN model from [11] had an accuracy of 95%, and the DNN model from [12] had the lowest accuracy of 90.75%. So, it can be concluded, that the proposed model is the best solution for the job and that the solution was correct and the other models would have a measurable error/uncertainty.

Table-5 Performance Metrics of Existing and Proposed Model for the BoTNeTIoT-L01-v2 dataset [58]

Algorithm	Accuracy	Recall	Precision	F1 score
DNN	89.70	85.83	94.95	84.68
CNN1	89.85	86.58	93.85	85.21
CNN2	90.60	88.08	95.95	86.64
GRU	89.88	86.14	94.79	85.04
LSTM	90.03	87.03	95	85.46

Hybrid	90.83	88.58	95.79	87.14
Proposed Model	**100**	**100**	**100**	**100**

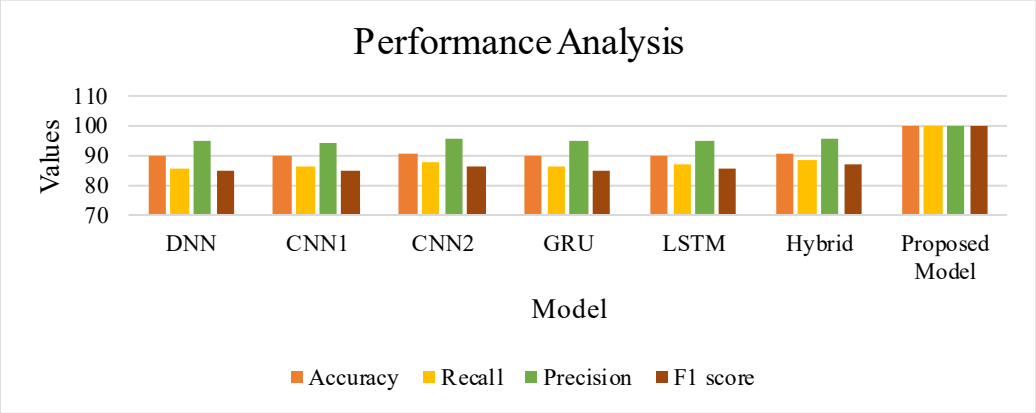

Figure.14 Performance Analysis of Existing Model with Proposed Model for BoTNeTIoT-L01-v2 dataset [58]

Table-5 and Figure.14 shows that the Proposed Model recorded the metrics with an accuracy, recall, precision, and F1 score of 100. The Hybrid model was next to demonstrate strong performance, recording an accuracy of 90.83, a recall of 88.58, a precision of 95.79, and an F1 score of 87.14. The CNN2 model was hot on its heels with an accuracy of 90.60, a recall of 88.08, precision of 95.95, and an F1 score of 86.64. The rest of the models, LSTM, GRU, CNN1, and DNN, also produced competitive outcomes. LSTM, in particular, had an accuracy of 90.03, recall of 87.03, precision of 95, and an F1 score of 85.46. GRU had an accuracy of 89.88, recall of 86.14, precision of 94.79, and an F1 score of 85.04. The CNN1 model had an accuracy of 89.85, recall of 86.58, precision of 93.85, and an F1 score of 85.21. Finally, the accuracy of the DNN model was 89.70, recall was 85.83, precision was 94.95, and the F1 score was 84.68.

Table-6 Performance Metrics of Existing and Proposed Model for the BoTNeTIoT-L01-v2 dataset [59]

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree (DT)	91.02	96.84	88.88	87.42
Extra tree (ET)	91.03	97.29	88.90	87.43
Gradient Boosting	90.54	96.21	87.81	86.44
Logistic Regression (LR)	75.06	69.77	73.98	70.31
Naïve bayes (NB)	56.17	60.38	63.70	56.10
Random Forest (RF)	90.99	97.11	88.85	87.36
Proposed Model	**100**	**100**	**100**	**100**

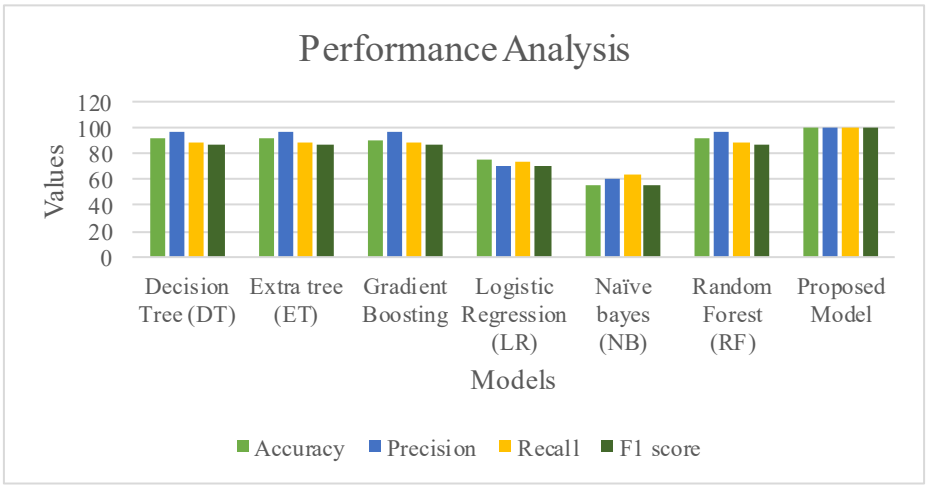

Figure.15 Comparison of Existing Model with Proposed Model for BoTNeTIoT-L01-v2 dataset [59]

Table-6 and Figure,15 has demonstrated that the Proposed Model scored 100 in all measures, with accuracy, recall, precision, and an F1 score of 100. The Extra tree model also had a very good score, with accuracy at 91.03, recall at 88.90, precision at 97.29, and an F1 score at 87.43. The Decision Tree model also performed very well, with accuracy at 91.02, recall at 88.88, precision at 96.84, and an F1 score at 87.42. The Random Forest model also exhibited robust outcomes with an accuracy of 90.99, recall of 88.85, precision of 97.11, and an F1 score of 87.36. An accuracy of 90.54, recall of 87.81, precision of 96.21, and an F1 score of 86.44 was attained by the Gradient Boosting model. For comparison, the Logistic Regression model was at 75.06 accuracy, 73.98 recall, 69.77 precision, and an F1 of 70.31, while the Naïve bayes model was at accuracy of 56.17, 63.70 recall, 60.38 precision, and an F1 of 56.10.

Correspondingly, in the proposed model, SSA-VMD is designed to reduce noise, select features relevant to the application, and provide the classifier with clear, informative features. Next comes BiGRU, which captures temporal dependencies and contextual information in the data sequences so that the model may perceive complex pattern evolutions across time. The selective feature extraction of SSA-VMD and the temporal modeling ability of BiGRU together give a boost to classification accuracy with cleaner input from SSA-VMD and BiGRU utilizing the inputs to model sequential dynamics more precisely.

5. Conclusion and Future Direction

Traditional approaches in intelligent Intrusion Detection Systems (IDS) had limitations, including high false-positive rates, poor feature extraction, and inability to adapt to evolving threats, as these approaches utilized shallow learning models or manually selected features, which limited their capacity to detect complex and intrusions and reduced the security effectiveness overall. These issues were resolved by developing a new approach that combined advanced feature selection techniques with a representative DL architecture. More specifically the proposed approach utilized a Multi-Strategy SSA-VMD for the feature selection process that improved the validity and robustness of the features selected by merging the two processes. It would be combined with a Hierarchical Bidirectional Gated Recurrent Unit (BiGRU) that introduced residual gates and time-aware attention mechanisms that would enhance the ability of the model to learn long-term dependencies, temporo-spatial dynamics, and contextual representations within network traffic data. The proposed model was validated using a wide-reaching intrusion detection dataset resulted in greater performance in terms of accuracy, precision, recall, and F1-score than conventional and current DL processes with the value of 1.00 or 100%. The results indicated that the addition of advanced feature selection to the hierarchical BiGRU architecture significantly improved detection performance with lower probability of false positives. The benefits of the proposed system included high detection accuracy, relevance of features and temporal modeling. The limitations of this work included a more complex architecture and tuning of parameters. Future work highlighted recommendations to increase efficiency, examine features of real-time detection, and extend the work to adaption to new threats, perhaps by a continual learning approach. The model is reconfigured to detect zero-day attacks by way of continual learning methods so it evolves to be in sync with newer attack patterns. Moreover, it can be applied in streamlining the real-time traffic data. Then, the evaluation can be done in various datasets. Overall, the proposed research was a considerable advancement to intelligent and adaptive intrusion detection systems.

References

[1] M. A. I. Mallick and R. Nath, "Navigating the cyber security landscape: A comprehensive review of cyber-attacks, emerging trends, and recent developments," *World Scientific News,* vol. 190, no. 1, pp. 1-69, 2024.

[2] F. Alamri, "Comprehensive study on object detection for security and surveillance: A concise review," *Multimedia Tools and Applications,* pp. 1-32, 2025.

[3] L. Diana, P. Dini, and D. Paolini, "Overview on intrusion detection systems for computers networking security," *Computers,* vol. 14, no. 3, p. 87, 2025.

[4] N. Jeffrey, Q. Tan, and J. R. Villar, "A review of anomaly detection strategies to detect threats to cyber-physical systems," *Electronics,* vol. 12, no. 15, p. 3283, 2023.

[5] F. Sharif, "The role of ensemble learning in strengthening intrusion detection systems: A machine learning perspective," *Int. J. Comput. Eng. Technol,* 2024.

[6] U. I. Okoli, O. C. Obi, A. O. Adewusi, and T. O. Abrahams, "Machine learning in cybersecurity: A review of threat detection and defense mechanisms," *World Journal of Advanced Research and Reviews,* vol. 21, no. 1, pp. 2286-2295, 2024.

[7] S. Ali, S. U. Rehman, A. Imran, G. Adeem, Z. Iqbal, and K.-I. Kim, "Comparative evaluation of ai-based techniques for zero-day attacks detection," *Electronics,* vol. 11, no. 23, p. 3934, 2022.

[8] B. Nawaal, U. Haider, I. U. Khan, and M. Fayaz, "Signature-based intrusion detection system for IoT," in *Cyber security for next-generation computing technologies*: CRC Press, 2024, pp. 141-158.

[9] A. I. Blessing, A. Chole, and Y. Themmah, "Signature-Based vs. Anomaly-Based Detection," 2023.

[10] E. Tufan, C. Tezcan, and C. Acartürk, "Anomaly-based intrusion detection by machine learning: A case study on probing attacks to an institutional network," *IEEE Access,* vol. 9, pp. 50078-50092, 2021.

[11] S. Alem, D. Espes, L. Nana, E. Martin, and F. De Lamotte, "A novel bi-anomaly-based intrusion detection system approach for industry 4.0," *Future Generation Computer Systems,* vol. 145, pp. 267-283, 2023.

[12] T. Sowmya and E. M. Anita, "A comprehensive review of AI based intrusion detection system," *Measurement: Sensors,* vol. 28, p. 100827, 2023.

[13] S. K. Parisa and S. Banerjee, "AI-Enabled Cloud Security Solutions: A Comparative Review of Traditional vs. Next-Generation Approaches," *International Journal of Statistical Computation and Simulation,* vol. 16, no. 1, 2024.

[14] P. Singh, P. Pranav, and S. Dutta, "Bi-GAN-LDA for cybersecurity: a hybrid deep learning framework for advanced network anomaly detection," *Engineering Research Express,* vol. 7, no. 2, p. 025238, 2025.

[15] M. Aminu, A. Akinsanya, D. A. Dako, and O. Oyedokun, "Enhancing cyber threat detection through real-time threat intelligence and adaptive defense mechanisms," *International Journal of Computer Applications Technology and Research,* vol. 13, no. 8, pp. 11-27, 2024.

[16] B. Cao, C. Li, Y. Song, and X. Fan, "Network intrusion detection technology based on convolutional neural network and BiGRU," *Computational Intelligence and Neuroscience,* vol. 2022, no. 1, p. 1942847, 2022.

[17] K. Yang, J. Wang, and M. Li, "An improved intrusion detection method for IIoT using attention mechanisms, BiGRU, and Inception-CNN," *Scientific Reports,* vol. 14, no. 1, p. 19339, 2024.

[18] R. Lalduhsaka and A. K. Khan, "Enhanced Intrusion Detection System Using a Two-Staged Feature Selection Method," *Security and Privacy,* vol. 8, no. 3, p. e70025, 2025.

[19] Z. Xia, S. He, C. Liu, Y. Liu, X. Yang, and H. Bu, "PSO-GA Hyperparameter Optimized ResNet-BiGRU Based Intrusion Detection Method," *IEEE Access,* 2024.

[20] F. Alamri and A. Dutta, "Implicit and explicit attention mechanisms for zero-shot learning," *Neurocomputing,* vol. 534, pp. 55-66, 2023.

[21] C. Liu, Y. Liu, Y. Yan, and J. Wang, "An intrusion detection model with hierarchical attention mechanism," *IEEE Access,* vol. 8, pp. 67542-67554, 2020.

[22] J. Sun, C. Li, and Y. Song, "A network security situation prediction approach based on MAML and BiGRU," *Journal of Intelligent & Fuzzy Systems,* vol. 47, no. 3-4, pp. 307-319, 2024.

[23] D. Guo and Y. Xie, "Research on Network Intrusion Detection Model Based on Hybrid Sampling and Deep Learning," *Sensors,* vol. 25, no. 5, p. 1578, 2025.

[24] B. Cao, C. Li, Y. Song, Y. Qin, and C. Chen, "Network intrusion detection model based on CNN and GRU," *Applied Sciences,* vol. 12, no. 9, p. 4184, 2022.

[25] F. Alamri, S. Kalkan, and N. Pugeault, "Transformer-encoder detector module: Using context to improve robustness to adversarial attacks on object detection," in *2020 25th international conference on pattern recognition (ICPR)*, 2021: IEEE, pp. 9577-9584.

[26] Y. Xiang, D. Li, X. Meng, C. Dong, and G. Qin, "ResNeSt-biGRU: An Intrusion Detection Model Based on Internet of Things," *Computers, Materials & Continua,* vol. 79, no. 1, 2024.

[27] M. A. Talukder *et al.*, "Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction," *Journal of big data,* vol. 11, no. 1, p. 33, 2024.
[28] S. S. Bamber, A. V. R. Katkuri, S. Sharma, and M. Angurala, "A hybrid CNN-LSTM approach for intelligent cyber intrusion detection system," *Computers & Security,* vol. 148, p. 104146, 2025.
[29] B. Susilo, A. Muis, and R. F. Sari, "Intelligent intrusion detection system against various attacks based on a hybrid deep learning algorithm," *Sensors,* vol. 25, no. 2, p. 580, 2025.
[30] R. Yao, N. Wang, Z. Liu, P. Chen, and X. Sheng, "Intrusion detection system in the advanced metering infrastructure: A cross-layer feature-fusion CNN-LSTM-based approach," *Sensors,* vol. 21, no. 2, p. 626, 2021.
[31] L. R. Vuyyuru, N. R. Purimetla, K. Y. Reddy, S. S. Vellela, S. K. Basha, and R. Vatambeti, "Advancing automated street crime detection: a drone-based system integrating CNN models and enhanced feature selection techniques," *International Journal of Machine Learning and Cybernetics,* vol. 16, no. 2, pp. 959-981, 2025.
[32] F. Alamri and A. Dutta, "Implicit and explicit attention for zero-shot learning," in *DAGM German conference on pattern recognition*, 2021: Springer, pp. 467-483.
[33] R. Jablaoui and N. Liouane, "Network security based combined CNN-RNN models for IoT intrusion detection system," *Peer-to-Peer Networking and Applications,* vol. 18, no. 3, p. 129, 2025.
[34] S. Islam, D. Javeed, M. S. Saeed, P. Kumar, A. Jolfaei, and A. N. Islam, "Generative AI and cognitive computing-driven intrusion detection system in industrial CPS," *Cognitive Computation,* vol. 16, no. 5, pp. 2611-2625, 2024.
[35] F. Alamri and A. Dutta, "Multi-head self-attention via vision transformer for zero-shot learning," *arXiv preprint arXiv:2108.00045,* 2021.
[36] J. Mao, X. Yang, B. Hu, Y. Lu, and G. Yin, "Intrusion Detection System Based on Multi-Level Feature Extraction and Inductive Network," *Electronics,* vol. 14, no. 1, p. 189, 2025.
[37] W. Yao, H. Peng, Q. Li, and X. Shen, "Modeling Realistic Adversarial Traffic Against Deep Learning-Based Intrusion Detection System in Industrial IoT," *IEEE Internet of Things Journal,* 2025.
[38] N. Javaid, A. Almogren, M. Adil, M. U. Javed, and M. Zuair, "RFE based feature selection and KNNOR based data balancing for electricity theft detection using BiLSTM-LogitBoost stacking ensemble model," *IEEE Access,* vol. 10, pp. 112948-112963, 2022.
[39] H. Yu, C. Kang, Y. Xiao, and Y. Yang, "Network intrusion detection method based on hybrid improved residual network blocks and bidirectional gated recurrent units," *IEEE Access,* vol. 11, pp. 68961-68971, 2023.
[40] *IDS 2018 Intrusion CSVs (CSE-CIC-IDS2018)*. [Online]. Available: https://www.kaggle.com/datasets/solarmainframe/ids-intrusion-csv
[41] *IoT dataset for Intrusion Detection Systems (IDS)*. [Online]. Available: https://www.kaggle.com/datasets/azalhowaide/iot-dataset-for-intrusion-detection-systems-ids
[42] H. Najafi Mohsenabad and M. A. Tut, "Optimizing cybersecurity attack detection in computer networks: A comparative analysis of bio-inspired optimization algorithms using the CSE-CIC-IDS 2018 dataset," *Applied Sciences,* vol. 14, no. 3, p. 1044, 2024.
[43] M. A. Khan, "HCRNNIDS: Hybrid convolutional recurrent neural network-based network intrusion detection system," *Processes,* vol. 9, no. 5, p. 834, 2021.
[44] F. Zhao, H. Zhang, J. Peng, X. Zhuang, and S.-G. Na, "A semi-self-taught network intrusion detection system," *Neural Computing and Applications,* vol. 32, no. 23, pp. 17169-17179, 2020.
[45] L. Liu, P. Wang, J. Lin, and L. Liu, "Intrusion detection of imbalanced network traffic based on machine learning and deep learning," *IEEE access,* vol. 9, pp. 7550-7563, 2020.
[46] S. Gamage and J. Samarabandu, "Deep learning methods in network intrusion detection: A survey and an objective comparison," *Journal of Network and Computer Applications,* vol. 169, p. 102767, 2020.
[47] Y. N. Kunang, S. Nurmaini, D. Stiawan, and B. Y. Suprapto, "Attack classification of an intrusion detection system using deep learning and hyperparameter optimization," *Journal of Information Security and Applications,* vol. 58, p. 102804, 2021.
[48] A. Mezina, R. Burget, and C. M. Travieso-González, "Network anomaly detection with temporal convolutional network and U-Net model," *IEEE Access,* vol. 9, pp. 143608-143622, 2021.
[49] M. Al-Fawa'reh, M. Al-Fayoumi, S. Nashwan, and S. Fraihat, "Cyber threat intelligence using PCA-DNN model to detect abnormal network behavior," *Egyptian Informatics Journal,* vol. 23, no. 2, pp. 173-185, 2022.
[50] S. Alzughaibi and S. El Khediri, "A cloud intrusion detection systems based on dnn using backpropagation and pso on the cse-cic-ids2018 dataset," *Applied Sciences,* vol. 13, no. 4, p. 2276, 2023.

[51] A. Ghosh, H. M. Ibrahim, W. Mohammad, F. C. Nova, A. Hasan, and R. Rab, "CoWrap: an approach of feature selection for network anomaly detection," in *International Conference on Advanced Information Networking and Applications*, 2022: Springer, pp. 547-559.

[52] P. Lin, K. Ye, and C.-Z. Xu, "Dynamic network anomaly detection system by using deep learning techniques," in *International conference on cloud computing*, 2019: Springer, pp. 161-176.

[53] S. Seth, K. K. Chahal, and G. Singh, "A novel ensemble framework for an intelligent intrusion detection system," *IEEE access,* vol. 9, pp. 138451-138467, 2021.

[54] A. L. G. Rios, Z. Li, K. Bekshentayeva, and L. Trajković, "Detection of denial of service attacks in communication networks," in *2020 IEEE international symposium on circuits and systems (ISCAS)*, 2020: IEEE, pp. 1-5.

[55] B. I. Farhan and A. D. Jasim, "Performance analysis of intrusion detection for deep learning model based on CSE-CIC-IDS2018 dataset," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 26, no. 2, pp. 1165-1172, 2022.

[56] R. I. Farhan, A. T. Maolood, and N. F. Hassan, "Optimized deep learning with binary PSO for intrusion detection on CSE-CIC-IDS2018 dataset," *Journal of Al-Qadisiyah for computer science and mathematics,* vol. 12, no. 3, pp. Page 16-27, 2020.

[57] A. Aatresh, K. Alabhya, S. Lal, J. Kini, and P. Saxena, "LiverNet: efficient and robust deep learning model for automatic diagnosis of sub-types of liver hepatocellular carcinoma cancer from H&E stained liver histopathology images," *International Journal of Computer Assisted Radiology and Surgery,* vol. 16, no. 9, pp. 1549-1563, 2021.

[58] B. A. kadheem Hammood and A. T. Sadiq, "Deep Learning Approaches for IoT Intrusion Detection Systems," *Iraqi Journal of Science,* pp. 6631-6646, 2024.

[59] B. A. K. Hammood and A. T. Sadiq, "ENSEMBLE MACHINE LEARNING APPROACH FOR IOT INTRUSION DETECTION SYSTEMS," *Iraqi Journal for Computers & Informatics Ijci,* vol. 49, no. 2, 2023.