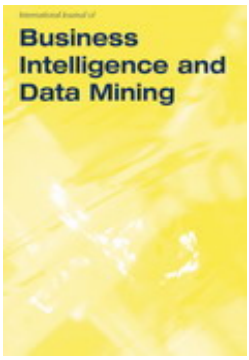


International Journal of Business Intelligence and Data Mining

ISSN online: 1743-8195 - ISSN print: 1743-8187
https://www.inderscience.com/ijbidm

DNN and BiGRU-based hierarchical attention network for intrusion detection

Hui Yan, Hupeng Liu, Ping Yu, Xiaoqing Xu, Mingxin Li, Yunxin Long, Hanlin Chen, Qi Wang, Duo Long

DOI: 10.1504/IJBIDM.2026.10079023

Article History:

Received:	11 March 2026
Last revised:	27 April 2026
Accepted:	28 April 2026
Published online:	07 September 2026

DNN and BiGRU-based hierarchical attention network for intrusion detection

Hui Yan

Information Engineering College,
Suqian University,
Suqian, 223800, China
Email: 17151@squ.edu.cn

Hupeng Liu, Ping Yu and Xiaoqing Xu

Jilin Province S&T Innovation Centre for Physical Simulation and Security of Water Resources and Electric Power Engineering,
Changchun Institute of Technology,
Changchun, 130012, China
Email: 2030969468@qq.com
Email: yuping19@mails.jlu.edu.cn
Email: 2625613463@qq.com

Mingxin Li

Information Engineering College,
Suqian University,
Suqian, 223800, China
Email: 106098997@qq.com

Yunxin Long

Changchun University of Chinese Medicine,
Changchun, 130117, China
Email: 1273112667@qq.com

Hanlin Chen

Jilin Province S&T Innovation Centre for Physical Simulation and Security of Water Resources and Electric Power Engineering,
Changchun Institute of Technology,
Changchun, 130012, China
Email: chenhanlin@ccit.edu.cn

Qi Wang and Duo Long*

Information Engineering College,
Suqian University,
Suqian, 223800, China
Email: 17198@squ.edu.cn
Email: 21129@squ.edu.cn
*Corresponding author

Abstract: To tackle the rising sophistication of cyberattacks and the critical need to enhance intrusion detection capabilities, this study presents a hybrid architecture combining a DNN with an attention mechanism and a BiGRU with an attention mechanism, BiGRU with attention, and an MLP classifier. The model integrates dual DNN-Attention and BiGRU-Attention submodules to capture high-dimensional static features and temporal dependencies, followed by feature fusion and classification via MLP. Evaluated on NSL-KDD and UNSW-NB15 datasets, the model achieves validation accuracies of 99.48% and 98.43%, respectively, with stable convergence and low loss. Comparative results demonstrate superior performance in accuracy, robustness, and generalisation, confirming its effectiveness for practical cybersecurity applications.

Keywords: intrusion detection; deep learning; bidirectional gated recurrent unit; attention mechanism; network security.

Reference to this paper should be made as follows: Yan, H., Liu, H., Yu, P., Xu, X., Li, M., Long, Y., Chen, H., Wang, Q. and Long, D. (2026) 'DNN and BiGRU-based hierarchical attention network for intrusion detection', *Int. J. Business Intelligence and Data Mining*, Vol. 28, No. 10, pp.1–24.

Biographical notes: Hui Yan is a Professor at the School of Information Engineering, Suqian University, Suqian City, China. She received the PhD degree in Agricultural Mechanisation Engineering from Jilin University, Changchun, China, in 2012. Her research interests include digital simulation, smart agriculture, big data, artificial intelligence.

Hupeng Liu is pursuing a Master's degree in Civil Engineering at Changchun Institute of Technology. His research focuses on deep learning-based network intrusion detection, primarily employing deep learning algorithms to enhance the accuracy and efficiency of intrusion detection. His work encompasses network traffic feature extraction, deep learning model design, and optimisation.

Ping Yu is an Associate Professor at the School of Computer Technology and Engineering, Changchun Institute of Technology, Changchun, China.She received the PhD degree in Computer Science and Technology from Jilin University, Changchun, China, in 2025.Her research interests include edge computing, artificial intelligence, big data, computer applications.

Xiaoqing Xu is a master's student in Civil and Hydraulic Engineering at Changchun Institute of Technology. Her research focuses on weed recognition for precision agriculture using convolutional neural network (CNN) models, and she also possesses additional expertise in path planning algorithms.

Mingxin Li is employed at the School of Information Engineering, Suqian College, and is currently pursuing a Master's degree in IBA at Dhurakij Pundit University. Their main research areas cover business administration, Homo sapiens resource management, strategic management, computer networks, Homo sapiens artificial intelligence, as well as big data analysis and applications.

Yunxin Long is a Doctoral student majoring in Acupuncture and Moxibustion at Changchun University of Chinese Medicine, Changchun, China. Her research interests are traditional Chinese medicine, acupuncture, artificial intelligence and big data.

Hanlin Chen is presently a Lecturer at the Changchun Institute of Technology. He obtained his Master's degree from the Changchun University of Technology in 2020. His research interests include grid computing, computer networks, and network security.

Qi Wang is a Teaching Assistant at the School of Information Engineering, Suqian University, Suqian city, China. She received the master's degree in Computer Application Technology from Nanjing University of Science and Technology, Nanjing, China, in 2021. Her research interests include computer networks, complex network theory, artificial intelligence and computer applications.

Duo Long is a Professor at the School of Economics and Management, Suqian University, Suqian city, China. He received his PhD degree in Agricultural Mechanisation Engineering from Jilin University, Changchun, China, in 2008, and the Post-Doctoral degree of Changchun University of Science and Technology, Changchun, China, in 2012. His research interests include smart agriculture, big data, soft science, artificial intelligence.

1 Introduction

The accelerating advancement of information technology has profoundly reshaped contemporary society, rendering digital communication remarkably convenient and fostering unprecedented global opportunities. Nevertheless, this rapid proliferation of networked systems has concurrently introduced formidable challenges to information assurance and data integrity. Security incidents encompassing unauthorised data exfiltration, system compromises, and malicious cyber intrusions have become increasingly prevalent, with certain breaches escalating to the level of national security crises. Consequently, safeguarding cyberspace has emerged as an imperative of global priority (Aslan et al. 2023). Intrusion detection systems (IDS), serving as a foundational defensive layer for preserving network integrity, have garnered substantial attention from both academic researchers and industrial practitioners (Mijwil et al. 2023; Zou and Wu 2023). Conventional machine learning paradigms including K-nearest neighbour (KNN) (Almomani et al. 2023), support vector machine (SVM) (Selvan 2024), Naive Bayes (NB) (Ige and Kiekintveld 2023), and decision tree (DT) (Chen et al. 2023) have historically contributed meaningfully to early-stage IDS development. However, such techniques encounter significant difficulties when confronted with contemporary network traffic characterised by massive volume, elevated dimensionality, and heterogeneous

structure. Moreover, their detection efficacy against sophisticated, continuously evolving attack vectors remains inadequate. By contrast, deep learning methodologies leverage their inherent strengths in hierarchical feature abstraction and nonlinear representation acquisition, thereby establishing themselves as a compelling alternative for attaining enhanced intrusion detection accuracy while concurrently suppressing false alarm rates.

To mitigate the inherent limitations of classical detection paradigms and bolster adaptability within complex operational environments, Introducing a hierarchical deep learning-based IDS framework, this paper unifies DNN, BiGRU, and self-attention within a cohesive architecture. The DNN component demonstrates proficiency in capturing high-dimensional, heterogeneous traffic features, whereas the BiGRU module specialises in extracting bidirectional temporal correlations embedded within sequential network data. Each traffic instance is characterised by 49 distinct features spanning multiple network-layer and flow-level attributes, including source and destination IP addresses, which collectively empower IDS to identify intricate and mutating cyber threats with greater efficacy. Such an architectural integration is simultaneously necessitated and rendered feasible by the demands of contemporary network infrastructures. Given the continuously mutating landscape of modern cyber threats ranging from stealthy intrusions to polymorphic malware deployments signature-based detection strategies have proven insufficient (Hussien 2023; Khraisat et al. 2019). Accordingly, this investigation focuses on constructing a multi-layered deep learning framework capable of robustly discriminating among diverse attack categories. We assess the proposed model on established benchmark datasets, specifically NSL-KDD and UNSW-NB15, evaluating its discriminative performance across multiple attack classes: denial-of-service (DoS) (Alrayes et al. 2024), user-to-root (U2R), remote-to-local (R2L) (El Asry et al. 2023), and probing attacks. The assessment adheres to a structured pipeline encompassing feature selection, model training, and performance validation. Comprehensive performance metrics including accuracy, precision, recall, and F1-score are derived from confusion matrix analysis, thereby affirming the practical utility of the proposed framework in reinforcing network protocol resilience and its robustness against adversarial perturbations.

2 Related work

As network environments become more complex and attack strategies evolve continuously, DNN-based IDS have displayed considerable potential to enhance detection performance. Despite this promise, considerable obstacles persist in adapting to unprecedented and highly sophisticated threat variants, processing voluminous and structurally diverse traffic streams, and achieving efficient deployment under computationally constrained conditions. A prominent limitation constraining DNN-based IDS adoption stems from their opaque ‘black-box’ operational nature, which severely impedes interpretability and thereby restricts their integration into mission-critical infrastructures where explainability and operational dependability are mandatory prerequisites. Additionally, the susceptibility of DNN architectures to adversarial perturbations has attracted intensifying scholarly focus, catalising numerous investigations into fortification of model robustness.

Notwithstanding these challenges, DNN-centric methodologies retain widespread utilisation within intrusion detection investigations. To ameliorate the deficiencies

associated with standalone implementations, researchers have increasingly pursued hybrid architectures that amalgamate DNNs with complementary deep learning modalities. Convolutional neural networks (CNNs), for instance, exhibit particular efficacy in encoding spatial and localised feature patterns, whereas recurrent neural networks (RNNs) demonstrate aptness for sequential dependency modelling; meanwhile, DNNs sustain their comparative advantage in deriving hierarchical representations from high-dimensional feature spaces (Wang et al. 2023). Illustratively, Fu et al. (2022) conceived the DLNID architecture, which reported superior accuracy and F1-score metrics on the NSL-KDD benchmark relative to traditional baseline approaches. Halbouni et al. (2022) proposed a CNN-LSTM composite structure for joint spatial-temporal feature extraction accompanied by network topology optimisation, a configuration that elevated detection accuracy while concomitantly attenuating false positive rates. Oboya et al. (2023) advanced a hybrid framework integrating DNN with radial basis function neural networks (RBFNN) to augment classification efficacy. Umair et al. (2024) implemented a multi-layer CNN paired with Softmax classification and deep DNN modules for comprehensive feature elucidation, demonstrating validated effectiveness across both NSL-KDD and KDDCup'99 datasets. Qazi et al. (2023) architected the HDLNIDS framework founded upon a convolutional recurrent neural network, resulting in a detection accuracy of 98.90% on the CICIDS-2018 dataset. Complementarily, Yin et al. (2023) deployed an integrated feature selection methodology designated IGRF-RF – synthesising information gain, random forest (RF), and recursive feature elimination (RFE) – which elevated MLP performance from 82.25% to 84.24% on the UNSW-NB15 dataset. For multi-class attack recognition within SDN-enabled IoT contexts, Chaganti et al. (2023) applied an LSTM-based methodology, achieving an accuracy of 0.971. Kasongo (2023) formulated an RNN-centric framework supplemented with XGBoost for feature refinement, whereas Aktar and Nur (2023) engineered a shrinkage autoencoder-oriented methodology targeting DDoS identification. Kimanzi et al. (2024) conducted a comprehensive survey of deep learning algorithms within IDS contexts. Mohamed and Ejbali (2023) investigated deep SARSA reinforcement learning for anomalous behaviour detection.

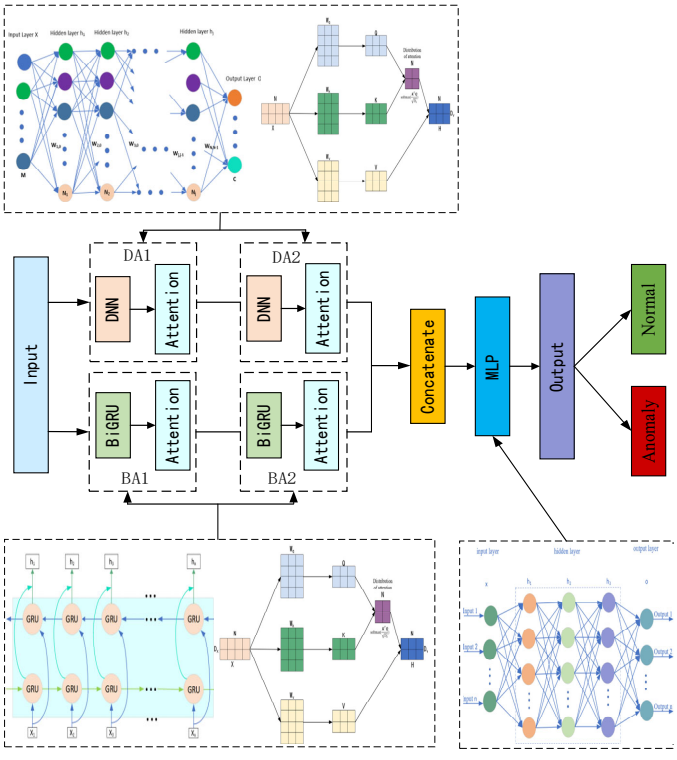
The principal contribution of this study resides in the conceptualisation and implementation of a novel hybrid IDS architecture that unifies DNN, BiGRU, and self-attention mechanisms within a hierarchical structure for multi-scale spatiotemporal feature extraction and integration. Specifically, the BiGRU module effectively encodes bidirectional temporal dependencies inherent in sequential network traces, while the attention mechanism dynamically reweights feature contributions, thereby amplifying the salience of discriminative attributes within the input distribution. Through the insertion of self-attention layers subsequent to both the DNN and BiGRU processing stages, the proposed model transcends the constraints associated with manual feature engineering and realises refined concentration upon pivotal dimensions in the derived feature manifold. Furthermore, the unified fusion of static representations originating from the DNN pathway with dynamic dependencies captured by the BiGRU pathway enables complementary multi-level learning. The architecture culminates in a multilayer perceptron (MLP) classifier augmented with Batch Normalisation and Dropout regularisation to enhance generalisation capacity, training stability, and overall detection performance. Empirical findings corroborate that the proposed framework exhibits pronounced recognition capability, environmental adaptability, and classification

accuracy across diverse and complex network scenarios, underscoring its pragmatic potential for operational intrusion detection deployments.

3 DNN-BiGRU-attention-MLP detection model construction

Deep neural networks (DNNs) have demonstrated remarkable capabilities in network intrusion detection, particularly regarding the extraction of deep-level features and the accurate classification of high-dimensional traffic data. Figure 1 illustrates a typical DNN architecture, comprising an input layer, several hidden layers, and an output layer. Through nonlinear activation functions that establish connections between neurons in adjacent layers, the model acquires the capacity to learn sophisticated feature representations. By strengthening the network's ability to capture intricate data patterns, this hierarchical structure leads to higher intrusion behaviour recognition accuracy.

Figure 1 Structure of hierarchical attention network model based on DNN-BiGRU (see online version for colours)



The DNN topology adopted in this study is informed by both theoretical principles and empirical experience in deep learning. Specifically, the adoption of deeper architectures facilitates hierarchical feature extraction, progressively enabling the model to derive increasingly abstract representations from network traffic data. A suitable selection of the neuron count for each layer achieves a balance between the model's expressive power and its computational cost. Furthermore, Batch Normalisation is integrated to ensure

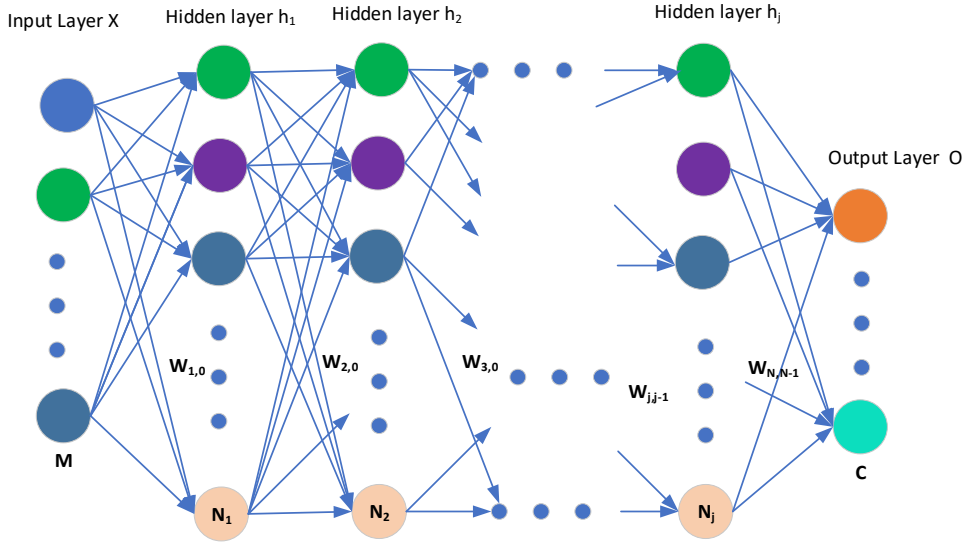
training stability and accelerate convergence, whereas Dropout is employed to alleviate overfitting and enhance generalisation performance. These architectural choices have been extensively validated for intrusion detection tasks, thereby contributing to both the robustness and efficacy of the proposed model.

3.1 DNN deep neural network

Within the field of network intrusion detection, DNNs have shown strong abilities for deep feature extraction and the effective classification of high-dimensional traffic data. The architecture of a typical DNN, as illustrated in Figure 2, consists of an input layer, a number of hidden layers, and an output layer. Through nonlinear activation functions that interconnect neurons across successive layers, the model is enabled to learn complex feature representations. By enhancing the network's capacity to model intricate patterns in the data, this hierarchical structure consequently boosts the accuracy of intrusion behaviour recognition.

The design of the DNN topology in this study is guided by both theoretical considerations and empirical practices in deep learning. Specifically, The use of deeper network structures enables hierarchical feature extraction, which in turn allows the model to learn representations of network traffic data that are increasingly abstract. Appropriate selection of the neuron count per layer balances model expressiveness against computational efficiency. In addition, Batch Normalisation is incorporated to stabilise the training process and accelerate convergence, while Dropout is utilised to lessen overfitting and strengthen generalisation performance. In intrusion detection tasks, these design choices have received wide validation. They also enhance the robustness and effectiveness of the proposed model.

Figure 2 DNN model architecture (see online version for colours)



The input layer initially receives an input vector $x = (x_1, x_2, \ldots, x_m)$ of dimension m, x_i denotes an individual feature. This input vector is then fed into the first hidden layer of

the DNN, where it undergoes a linear transformation defined by the weight matrix W_1 and bias vector b_1. The transformation is computed as follows:

$$z_k^{(1)} = \sum_{i=1}^{m} w_{ik}^{(1)} \cdot x_i + b_k^{(1)} \tag{1}$$

Then, applying a nonlinear activation function – for instance, ReLU or Sigmoid – to the weighted sum produces the first hidden layer's output, The formula is as follows:

$$h_k^{(1)} = f\left(z_k^{(1)}\right) \tag{2}$$

In the j^{th} hidden layer ($j \in \{2, ..., n\}$) , all neurons in the $j-1^{\text{th}}$ layer provide inputs to each neuron. The latter then performs a weighted summation over those inputs to calculate its output, adding a bias term, and applying an activation function, the formula is as follows:

$$z_k^{(j)} = \sum_{i=1}^{n_{j-1}} w_{ik}^{(j)} \cdot h_i^{(j-1)} + b_k^{(j)} \tag{3}$$

$$h_k^{(j)} = f\left(z_k^{(j)}\right) \tag{4}$$

In these expressions, n_{j-1} indicates the neurons of $j-1^{\text{th}}$ layer, $w_{ik}^{(j)}$, denotes the weight value, and $b_k^{(j)}$ denotes its bias term.

Located after the final hidden layer is the output layer, whose neuron count is configured in accordance with the specific task's requirements. The activations from the last hidden layer are transformed by the output layer via a linear operation and an output activation function to yield the model's predicted outputs. The formula is presented below:

$$z_l^{(o)} = \sum_{k=1}^{n} w_{kl}^{(o)} \cdot h_k^{(n)} + b_l^{(o)} \tag{5}$$

$$o_l = f\left(z_l^{(o)}\right) \tag{6}$$

Let $w^{(o)}$ and $b^{(o)}$ be the output layer's weight matrix and bias vector, respectively, and let f be its activation function. MSE serves as a typical loss function. The formula is as follows:

$$L = \frac{1}{T} \sum_{i=1}^{T} \sum_{l=1}^{p} \left(y_l^{(t)} - o_l^{(t)}\right)^2 \tag{7}$$

where T is the number of training samples, p is the number of output classes, For the t^{th} sample, $y_l^{(t)}$ denotes the true label, while $o_l^{(t)}$ represents the corresponding predicted output.

To decrease the loss and upgrade model performance, using backpropagation, the model's parameters undergo iterative updates. Taking the update of the weight $w_{ik}^{(j)}$ connecting the $j-1^{\text{th}}$ and j^{th} layers as an example, its gradient is computed as:

$$\frac{\partial L}{\partial w_{ik}^{(j)}} = \frac{\partial L}{\partial z_{ik}^{(j)}} \cdot \frac{\partial z_{ik}^{(j)}}{\partial w_{ik}^{(j)}} = \delta_k^{(j)} \cdot h_i^{(j-1)} \tag{8}$$

where $\delta_k^{(j)}$ represents the error term for neuron k in layer j, and $h_i^{(j-1)}$ is the activation from the previous layer. This formulation follows the chain rule of calculus in gradient computation.

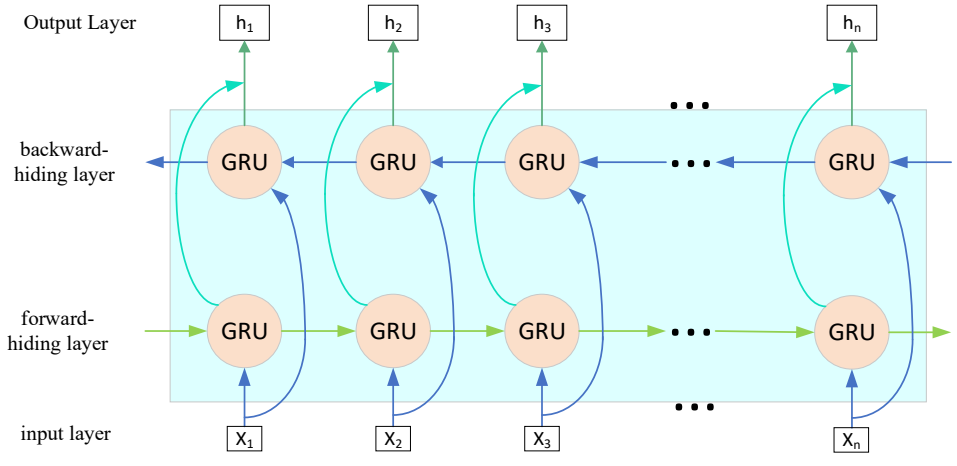
The parameters are then updated using the following equations:

$$w_{ik(new)}^{(j)} = w_{ik(old)}^{(j)} - \alpha \cdot \frac{\partial L}{\partial w_{ik}^{(j)}} \tag{9}$$

$$b_{k(new)}^{(j)} = b_{k(old)}^{(j)} - \alpha \cdot \delta_k^{(j)} \tag{10}$$

By calculating weighted combinations of features across all positions in the input sequence, the attention mechanism adaptively modulates the significance of individual feature dimensions. This enhancement of the model's ability to pinpoint critical information leads to improved accuracy in detecting intrusion behaviours. While the DNN architecture demonstrates exceptional capability in encoding high-dimensional features, the self-attention mechanism augments this by capturing global dependency relationships. The fusion of these two components facilitates a smooth progression from localised feature perception to comprehensive global comprehension. Through the construction of multiple parallel DNN-Attention submodules, the model achieves simultaneous extraction of deep semantic features, suppression of redundant information, and emphasis on salient patterns. Regulated by the learning rate α, the model iteratively refines its network weights via forward and backward propagation, progressively enhancing its discriminative power for complex samples.

Figure 3 Bi-directional gated cycle unit architecture (see online version for colours)



3.2 BiGRU module

Extracting features from sequential data constitutes a significant hurdle for traditional unidirectional GRUs. Restricted to leveraging merely past information, such units are

inherently hindered in comprehending full contextual dependencies. To remedy this deficiency and fortify the model's grasp of global sequence relationships, the present study deploys a BiGRU architecture (Figure 3). Rather than processing the input sequence in a single direction like conventional GRUs, the BiGRU operates simultaneously along both forward and backward pathways. As a result, semantic information from preceding as well as succeeding contexts is captured at every individual time step. The hidden representations yielded by the two directional traversals are subsequently merged to generate a more informative latent encoding. Such a bidirectional scheme considerably enhances how the model acquires time-sensitive and context-dependent characteristics, thereby yielding improved performance in intrusion detection and analogous tasks. At a given time step t, the hidden state draws upon not merely the instantaneous input x_t, but also the forward hidden state h_{t-1} (from the prior step) and the backward hidden state h_{t+1} (from the upcoming step).

The forward derivation equation is as follows.

$$\vec{r}_t = \sigma\left(\vec{W}_r \vec{x}_t + \vec{U}_r \vec{h}_{t-1} + \vec{b}_r\right) \tag{11}$$

$$\vec{z}_t = \sigma\left(\vec{W}_z \vec{x}_t + \vec{U}_z \vec{h}_{t-1} + \vec{b}_z\right) \tag{12}$$

$$\vec{h}_t = \tanh tanh\left(\vec{W}_h \vec{x}_t + \vec{U}_h \left(\vec{r}_t \odot \vec{h}_{t-1} \vec{b}_h\right)\right) \tag{13}$$

$$\vec{h}_t = \vec{z}_t \odot \vec{h}_{t-1} + \left(1 - \vec{z}_t\right) \odot \tilde{h}_t \tag{14}$$

The backward derived equations are as follows.

$$\overleftarrow{r}_t = \sigma\left(\overleftarrow{W}_r \overleftarrow{x}_t + \overleftarrow{U}_r \overleftarrow{h}_{t-1} + \overleftarrow{b}_r\right) \tag{15}$$

$$\overleftarrow{z}_t = \sigma\left(\overleftarrow{W}_z \overleftarrow{x}_t + \overleftarrow{U}_z \overleftarrow{h}_{t-1} + \overleftarrow{b}_z\right) \tag{16}$$

$$\overleftarrow{h}_t = \tanh tanh\left(\overleftarrow{W}_h \overleftarrow{x}_t + \overleftarrow{U}_h \left(\overleftarrow{r}_t \odot \overleftarrow{h}_{t-1} \overleftarrow{b}_h\right)\right) \tag{17}$$

$$\overleftarrow{h}_t = \overleftarrow{z}_t \odot \overleftarrow{h}_{t-1} + \left(1 - \overleftarrow{z}_t\right) \odot \tilde{h}_t \tag{18}$$

The linear fusion of the results from both the forward and reverse directions is given by.

$$h_t^{combined} = \left[\vec{h}_t ; \overleftarrow{h}_t\right] \tag{19}$$

In the above equations:

$\vec{r}_t$ and $\overleftarrow{r}_t$ denote the reset gate vectors for the forward and backward directions, respectively, determining the contribution of the prior hidden state to the candidate hidden state.

$\vec{z}_t$ and $\overleftarrow{z}_t$ act as the update gate vectors, governing the balance between the prior hidden state and the candidate state.

$\vec{h}_t$ refers to the candidate hidden state, which is constructed using current input data, preceding hidden state and reset gate parameters.

$h_t^{combined}$ fuses the bidirectional hidden features by concatenation, and is adopted as the ultimate context feature representation of time step *t*.

σ refers to the sigmoid activation function, while the symbol $\odot$ indicates the Hadamard product operation between elements.

By integrating information from both temporal directions, The BiGRU enriches per-time-step semantic representation and enhances the model's complex sequence dynamics capture ability. This characteristic makes it highly suitable for network intrusion detection, where strong temporal dependencies are vital. Building upon this, The inclusion of a self-attention mechanism further enhances the model's focusing capability on salient temporal segments. Through dynamic allocation of attention weights, the model is able to automatically highlight time steps most relevant to the current classification task, which in turn improves feature selection precision and decision-making efficiency.

Apart from enabling productive bidirectional temporal dependency modelling, the BiGRU-attention pairing also intensifies the model's competence in encoding global temporal characteristics. Via the parallel instantiation of multiple BiGRU-Attention submodules, the model achieves more exhaustive extraction of deep behavioural features, thereby furnishing a firm basis for ensuing global feature fusion and classification.

3.3 Self-attention mechanism

The self-attention mechanism, likewise named intra-attention, dates back to neural network research in the early 1970s. Although the conceptual foundation has existed for decades, self-attention has only recently attracted significant attention and rapid development, particularly with the advent of deep learning. This mechanism excels at capturing intricate internal dependencies within sequences by computing pairwise similarities among elements. Moreover, Inherent parallel computation within its architecture greatly improves the model's processing capability for long texts and sequences. This makes self-attention a critical component in the construction of high-performance deep learning models. The computational structure is illustrated in Figure 4.

The core idea behind the self-attention mechanism is to dynamically adjust the contribution of each element in a sequence based on their contextual relevance, thereby producing more informative and expressive representations. The computation adopts query-key-value (QKV) structure. The detailed process is as follows.

Consider an input sequence $Y = [y_1, y_2, ..., y_n] R^{D_y \times N}$, with N being the sequence length and D_y representing the dimension of each input feature. This input is linearly transformed into three distinct feature spaces, yielding the query (Q), key (K), and value (V) matrices:

$$Q = W_Q X \in R^{D_k \times N} \tag{20}$$

$$K = W_K X \in R^{D_k \times N} \tag{21}$$

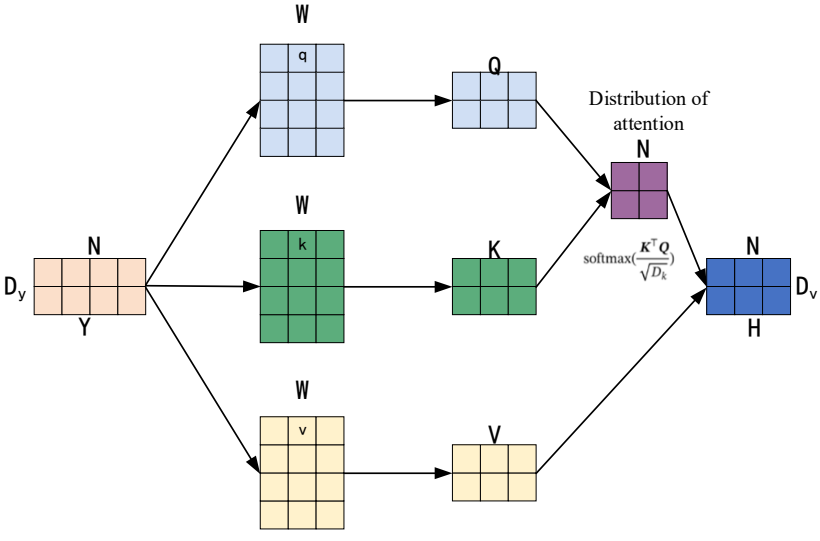
$$V = W_V X \in R^{D_v \times N} \tag{22}$$

Here, W_Q, W_K, and W_V are learnable projection matrices that map the input sequence into query, key, and value spaces, respectively. To compute the attention distribution for a given query vector $q_i \in Q$, its similarity with each key vector $k_j \in K$ is calculated using scaled dot-product attention, defined as:

$$att(Q, K, V) = \text{softmax}\left(\frac{Q^T K}{\sqrt{D_k}}\right)V \tag{23}$$

where Q^T indicates the dot-product similarity between the query and the key, thereby measuring their association; the scaling factor $\sqrt{D_k}$ alleviates the issue of gradient instability that could stem from an excessively large dot product; and the softmax function normalises the similarities across all positions to generate the attention weights; the final output H is the result of weighted summation after applying these weights to the value after the weighted summation result, indicating the updated representation of each position in the sequence. This mechanism allows the model to flexibly emphasise relevant contextual information at each time step, thus enabling effective long-range dependency modelling in various sequence learning tasks, such as machine translation, language modelling, and intrusion detection.

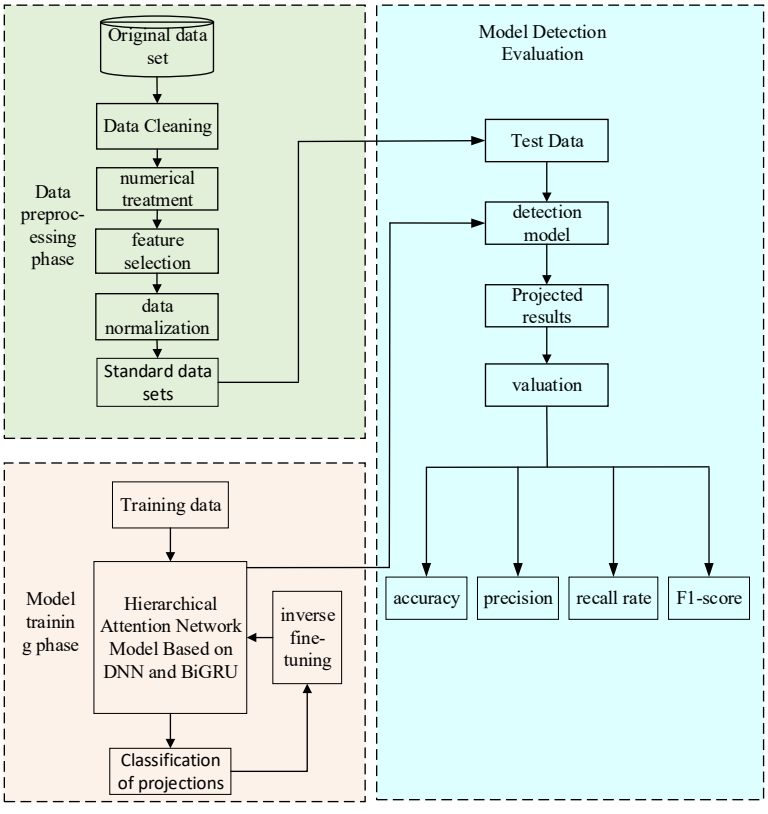
Figure 4 Self-attention mechanism computational process (see online version for colours)



4 Model solution from testing process

The intrusion detection approach introduced in the present work implements an exhaustive detection pipeline utilising two publicly accessible benchmark datasets, NSL-KDD and UNSW-NB15. This pipeline comprises three fundamental stages, specifically data pre-processing, model training, and assessment of detection effectiveness. The complete detection architecture is presented in Figure 5.

Figure 5 Overall detection framework (see online version for colours)



4.1 Data pre-processing

As a vital prerequisite for deep learning model training, data pre-processing guarantees that input data satisfies requirements pertaining to both quality and uniformity before architectural construction commences. Within this investigation, a methodical pre-processing regimen is implemented upon the raw datasets, incorporating data cleansing, encoding and numerical conversion of features, feature subset selection, and data scaling, as elaborated subsequently:

- Data cleaning: raw datasets frequently encompass absent values or inappropriate data formats capable of impeding productive model training. To remedy this deficiency, a column-oriented examination is executed to detect and correct irregularities. As an illustration, within the service feature column, invalid entries such as hyphens (–) are substituted with NaN, followed by the elimination of all records containing missing values, thereby preserving data completeness and uniformity.
- Feature encoding and numericalisation: given that deep learning architectures mandate numerical inputs, categorical variables necessitate transformation into numeric representations. Categorical attributes including protocol_type, service, flag, and state undergo one-hot encoding to be converted into an appropriate vectorised format, consequently augmenting the model's capacity for learning.

- Feature selection: with the objective of enhancing training efficiency and model effectiveness, Pearson correlation coefficients are calculated between individual features and the target label. Attributes demonstrating pronounced correlations (exceeding 0.3) are preserved, whereas irrelevant or superfluous features are discarded. This procedure facilitates the construction of a concise yet informative feature subset for subsequent model ingestion.
- Data normalisation: to counteract the challenge posed by features functioning across disparate scales, normalisation becomes essential to avert attributes with elevated magnitudes from disproportionately influencing the learning procedure. All numerical features are standardised to the [0, 1] range through application of the min-max normalisation methodology, which is formally expressed via the equation below:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{24}$$

In this expression, X denotes the original feature value, with X_{min} and X_{max} representing its minimum and maximum, and X' indicating the normalised result. This transformation accelerates model convergence and enhances generalisation performance.

4.2 Experimental setup

The pre-processed training data is supplied to a hierarchical attention-driven network architecture. Using the fused architecture of a DNN and a BiGRU, simultaneous modelling of static characteristics and sequential patterns is accomplished. Additionally, integrating a self-attention mechanism supports the dynamic prioritisation of important feature dimensions and temporal steps, thereby elevating the model's representational potential. Leveraging this architectural foundation, the model executes predictive classification upon input samples, with parameters being progressively refined through backpropagation coupled with fine-tuning. This establishes a cyclical process encompassing prediction and optimisation throughout training, permitting the model to progressively enhance its classification effectiveness and attain precise, generalisable identification of network intrusion activities.

4.3 Model testing and evaluation

In evaluating network IDS, the metrics most commonly adopted comprise accuracy, precision, recall, and F1-score, each crucial for delivering a thorough performance evaluation from multiple viewpoints. These criteria mirror the model's proficiency in discriminating between legitimate and malicious traffic, simultaneously offering valuable guidance for refining training methodologies and enhancing generalisation capacity.

1 Accuracy (ACC) quantifies the proportion of correctly classified instances across the entire sample set. This metric serves as a basic indicator of the model's overall classification ability. A higher accuracy value signifies better performance for both positive and negative classes. The formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{25}$$

2 Precision (P) evaluates the proportion of true positives among all instances predicted to be positive, thereby demonstrating the model's trustworthiness in intrusion detection and its capacity to sidestep false alarms. A greater precision suggests a reduced rate of false positives. The calculation is:

$$Precision = \frac{TP}{TP + FP} \tag{26}$$

3 Recall (Recall, R) synonymous with sensitivity, quantifies the fraction of true positive instances that the model correctly recognises, thereby reflecting the model's ability to detect intrusions without omissions. An increase in recall indicates better coverage of the positive class. The formula is:

$$Recall = \frac{TP}{TP + FN} \tag{27}$$

4 F1-score is obtained as the harmonic mean of precision and recall. This metric yields an equilibrium by jointly assessing how accurately positive cases are predicted and how comprehensively true instances are identified. A higher F1-Score indicates that the model achieves a better trade-off between reducing false positives and minimising missed detections. It is computed as:

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{28}$$

5 Experimentation and analysis

5.1 Experimental environment

The experiments were run on Windows 10 operating system with code based on Python 3.9.19, using TensorFlow and PyTorch as deep learning frameworks, Intel(R) Core(TM) i5-8500 @3.00GHZ CPU, 8GB of RAM, and NVIDIA GeForce GT 730 graphics card.

With the aim of further evaluating the proposed model's practical applicability, this study analyses its computational complexity and runtime performance.

From a theoretical perspective, the computational complexity of the proposed DNN-BiGRU-Attention-MLP model mainly arises from three components: DNN layers, BiGRU layers, and the self-attention mechanism. The DNN modules have a time complexity of $O(n \cdot d \cdot h)$, where n is the number of samples, d is the input feature dimension, and his the number of hidden units. The BiGRU modules introduce sequential processing, with complexity approximately $O(n \cdot T \cdot h^2)$, where T represents the sequence length. Additionally, the self-attention mechanism incurs a computational cost of $O(T^2 \cdot d)$, which may increase significantly with longer sequences.

Despite the increased architectural complexity, the model benefits from parallel computation in DNN and attention layers, which improves computational efficiency. In practical experiments, training was conducted on a system equipped with an Intel i5-8500 CPU and NVIDIA GeForce GT 730 GPU. Training took an average of 1.8–2.5 seconds

per epoch on the NSL-KDD dataset, and 2.5–3.2 seconds per epoch on the UNSW-NB15 dataset. The overall training time for 100 epochs was within an acceptable range, thereby demonstrating that the model achieves a balance between detection performance and computational cost.

Furthermore, compared with traditional deep models such as CNN-LSTM, the proposed architecture avoids excessively deep stacking and reduces redundant computations through feature fusion, making it more suitable for real-world deployment scenarios with limited computational resources.

5.2 Experimental setup

The final training dataset is input into the proposed network model for training. The model comprises multiple modules, including fully connected (DNN) and bidirectional GRU (BiGRU) layers, each configured with 32 neurons. The DNN modules utilise the ReLU activation function, Batch Normalisation for feature scaling, and we use a Dropout rate of 0.5 to reduce overfitting. For the BiGRU modules, BiGRUs with 32 units are employed and set to return full sequences. To boost the model's ability to focus on salient features, a self-attention mechanism is introduced. An MLP module of 60 neurons conducts feature fusion and nonlinear transformation, and is further optimised via ReLU activation, Batch Normalisation, and Dropout. The entire model is trained for 100 epochs using the Adam optimiser (batch size 32), with binary cross-entropy employed as the loss function.

5.3 Introduction to the dataset

5.3.1 NSL-KDD dataset

The KDD CUP 99 dataset was initially generated by emulating a U.S. Air Force local area network, integrating network connection information with system audit data, and has been instrumental in the progress of NIDS (Tavallaee et al., 2009). Nonetheless, it contains a substantial number of redundant samples – around 78% in training and 75% in testing – which can result in biased learning by models. To overcome these limitations, researchers developed an improved version known as the NSL-KDD dataset. This version eliminates duplicate samples, reconstructs training and testing data, and addresses issues of data imbalance and missing values, thereby significantly improving overall data quality. NSL-KDD has since become a widely used benchmark dataset in intrusion detection research. It consists of four subsets (KDDTrain+, KDDTest+, KDDTrain+_20Percent, and KDDTest-21), with 43 features in total – 41 numerical and 2 label attributes. The data includes normal traffic and four categories of attacks: DoS, Probe, U2R, and remote to local (R2L). The training set (KDDTrain+) contains 125,973 records covering 22 attack types, while the test set (KDDTest+) includes 22,544 records covering 39 attack types. Among the features, three categorical variables (ICMP, UDP, and TCP) represent network protocols, and the remaining numerical features reflect comprehensive characteristics of network traffic. Together, Owing to these properties, the NSL-KDD dataset serves as a reliable foundation for assessing intrusion detection model performance.

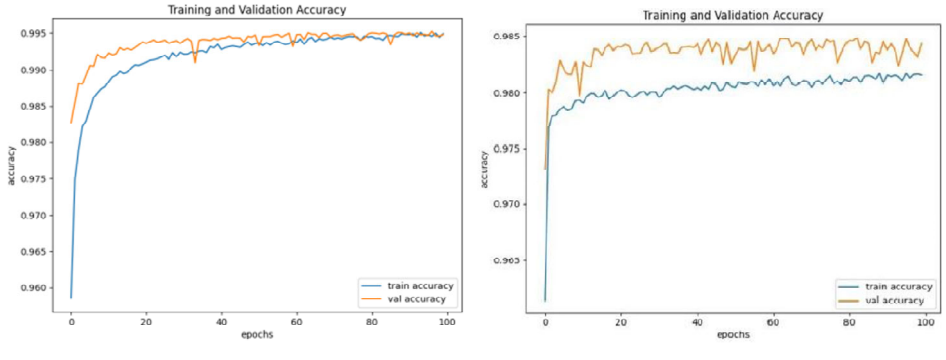
5.3.2 UNSW-NB15 dataset

Widely viewed as a benchmark in network intrusion detection, the UNSW-NB15 dataset was developed and publicly released by the University of New South Wales (UNSW) in Australia. Unlike earlier synthetic datasets, UNSW-NB15 is constructed from real network traffic captured in a cyber range environment, exhibiting a high degree of authenticity, diversity, and complexity (Moustafa and Slay, 2015). These attributes make the dataset highly suitable for IDS performance evaluation under realistic, dynamic conditions. In this study, For experimental analysis, the dataset's training and testing subsets are employed. The dataset consists of approximately 80,000 network connection records, covering nine types of malicious activities and one category of normal traffic. Specifically, it includes 45,743 training records, 22,361 testing records, and 11,089 validation records. Forty-nine features per sample describe a variety of network-layer and flow-related attributes, such as source and destination IP addresses, port numbers, protocol types, and connection states. The comprehensive representation of modern network behaviours and traffic patterns provided by these features establishes UNSW-NB15 as a valuable resource for assessing the effectiveness, robustness, and generalisability of intrusion detection models in real-world environments.

5.4 Analysis of experimental results

Experiments carried out on two representative benchmark datasets – NSL-KDD and UNSW-NB15 – served to confirm the effectiveness of the proposed network intrusion detection method. Figure 6(a) demonstrates that the model's training accuracy and validation accuracy on the NSL-KDD dataset both exhibit consistent improvement as the training epoch count grows. The performance begins to converge after approximately 80 epochs, ultimately achieving a training accuracy of around 99.20% and a validation accuracy of approximately 99.48%. For the UNSW-NB15 dataset, the training results depicted in Figure 6(b) show a similar upward trend in accuracy, although the convergence occurs at a comparatively slower pace, stabilising after roughly 100 epochs. Approximately 98.0% is obtained as the final training accuracy, while the validation accuracy levels off at around 98.43%. Overall, the model exhibits strong fitting capability and robustness across both datasets, confirming its reliability and generalisation performance in different intrusion detection scenarios.

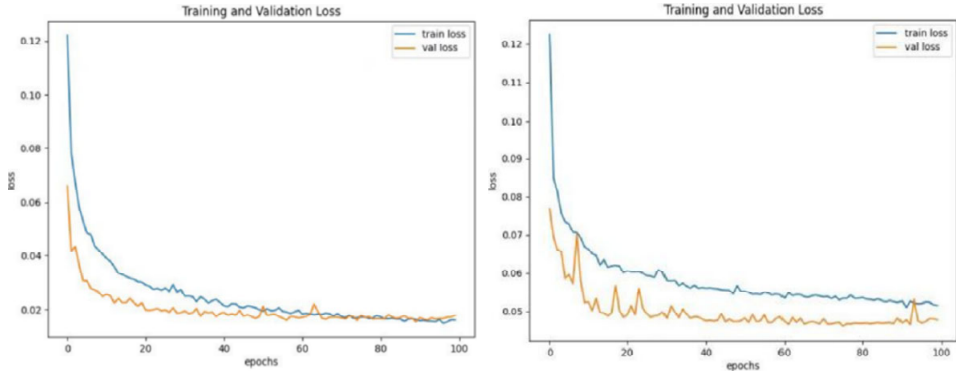
Figure 6 Training and verification accuracy performance graphs (see online version for colours)



(a) (b)

During the training phase, The change in the loss function further verifies the model's convergence and generalisation abilities. On the NSL-KDD dataset, the starting training loss is comparatively high, around 0.12, but it rapidly decreases to around 0.04 by the 20th epoch and gradually stabilises at a lower level near 0.02. The validation loss exhibits a consistent downward trend, ultimately stabilising slightly below the training loss at approximately 0.018, indicating no significant signs of overfitting, as illustrated in Figure 7(a). For the UNSW-NB15 dataset, the training loss starts at approximately 0.11 and similarly decreases rapidly, reaching around 0.07 at epoch 20. It then follows a fluctuating downward pattern and eventually converges to approximately 0.05. The validation loss follows a trend consistent with that of the training loss and stabilises at around 0.045, slightly lower than the final training loss. These results, as depicted in Figure 7(b), verify that the proposed model also attains good convergence and generalisation performance on the UNSW-NB15 dataset.

Figure 7 Training and validation loss performance plots (see online version for colours)



(a) (b)

Aiming to validate the effectiveness and superiority of the proposed DNN-BiGRU-Attention-MLP intrusion detection model when applied to real-world network environments, a systematic series of comparative experiments was conducted on two widely recognised benchmark datasets: NSL-KDD and UNSW-NB15. The experiment design was split into four major components, with the goal of evaluating the proposed model's detection performance and generalisation ability in a comprehensive and multi-dimensional manner. The first experimental group pairs the proposed DNN-BiGRU-Attention feature extractor with a range of mainstream classifiers – such as NB, SVM, DT, K-NN, and RF – to assess how different classification strategies affect overall detection performance, KNN, and RF – to explore how different classification strategies affect the overall detection performance. The second group of experiments focuses on feature selection, comparing the effectiveness of techniques including linear discriminant analysis, independent component analysis, RFE and RF-based feature importance are the methods adopted. The combination of each feature processing approach with an MLP classifier is used to analyse the impact of such approaches on model performance improvement. The third experiment series centres on model architecture optimisation. By progressively incorporating BiGRU, attention mechanisms, and MLP layers, multiple variants of deep learning models with structural differences are constructed, revealing the contributions of each component in modelling

temporal dependencies, enhancing the weighting of critical features, and improving classification accuracy. The final set of experiments introduces representative baseline and state-of-the-art algorithms – such as SDAE-ELM, BDNN, BCNN, and PV-DM – for performance comparison. These experiments evaluate the proposed model from both traditional and deep learning perspectives, demonstrating its overall superiority. The comparative experimental results on the NSL-KDD dataset are shown in Table 1.

Table 1 Comparative experimental results of NSL-KDD

Comparative classification	*Model*	*Metric*			
		Accuracy	*Precision*	*Recall rate*	*F1-score*
Different classification methods	DNN-BiGRU-Attention-NB	80.56%	81.25%	77.73%	79.45%
	DNN-BiGRU-Attention-SVM	94.10%	97.33%	90.28%	93.67%
	DNN-BiGRU-Attention-DT	98.23%	98.15%	98.19%	98.17%
	DNN-BiGRU-Attention-KNN	98.49%	98.51%	98.37%	98.44%
	DNN-BiGRU-Attention-RF	98.73%	98.96%	98.42%	98.69%
Different feature extraction methods	LDA-MLP	93.06%	94.28%	91.17%	92.70%
	ICA – MLP	94.00%	96.80%	90.58%	93.59%
	REF – MLP	93.53%	96.33%	90.05%	93.08%
	RF – MLP	95.17%	97.57%	92.30%	94.86%
Different model	DNN	94.29%	98.09%	89.85%	93.79%
	DNN-BiGRU	99.03%	98.75%	99.24%	99.00%
	DNN-BiGRU-Attention	98.70%	99.53%	97.78%	98.65%
	DNN-BiGRU – MLP	99.14%	99.20%	99.03%	99.11%
Different algorithms	SDAE-ELM	78.04%	95.99%	64.12%	76.87%
	BDNN	84.70%	79.45%	87.00%	83.05%
	BCNN	88.81%	89.00%	89.00%	89.00%
	PV-DM	98.92%	98.92%	95.44%	96.77%
Paper model		*99.48%*	*99.44%*	*99.43%*	*99.43%*

The proposed DNN-BiGRU-Attention-MLP model was comprehensively evaluated by integrating various mainstream classifiers, feature selection methods, and architectural configurations to assess its performance in intrusion detection tasks. Experimental results demonstrate that classifier selection significantly impacts overall model performance. Among the tested classifiers, the highest accuracy (98.73%) and F1-score (98.69%) were achieved by the RF classifier, highlighting its superior capability in modelling high-dimensional and nonlinear data. the KNN and DT classifiers achieved nearly matching performance, also exhibiting strong classification capabilities. Although SVM achieved relatively high precision, its low recall reduced the overall F1-score, indicating limitations in identifying minority attack classes. In contrast, the NB classifier, constrained by its assumption of feature independence, exhibited inferior performance in capturing complex feature relationships. In terms of feature selection, the RF-based method combined with an MLP classifier yielded the best results, with an accuracy of 95.17% and an F1-score of 94.86%. This indicates that RF-based feature importance

ranking effectively preserves critical features while eliminating redundant information. Although ICA, RFE, and LDA demonstrated slightly lower performance, they still maintained stable detection outcomes and acceptable generalisation.

As for model architecture, the progressive incorporation of BiGRU, attention mechanisms, and MLP modules significantly enhanced the model's overall capability. The F1-score increased from 93.79% with the base DNN model to 99.11% in the final integrated architecture. This validates the effectiveness of multi-module collaboration in improving temporal sequence modelling, enhancing feature attention, and boosting classification accuracy. Lastly, comparative experiments with representative existing methods – such as SDAE-ELM, BDNN, BCNN, and PV-DM – further confirmed the superiority of the proposed DNN-BiGRU-Attention-MLP model. It outperformed these models in both accuracy (99.48%) and F1-score (99.43%), particularly excelling in the detection of minority attack types, robustness to data variations, and adaptability to complex network environments. These findings firmly confirm the proposed model's real-world viability for network intrusion detection applications. In spite of the strong performance attained by the proposed model, a number of limitations should be pointed out. First, due to the integration of multiple deep learning modules, including DNN, BiGRU, and attention mechanisms, the model introduces increased computational complexity, which may present limitations for deployment in real-time or resource-constrained scenarios. Second, although the model demonstrates strong generalisation ability on benchmark datasets, its performance in highly dynamic and large-scale real-world networks still requires further validation. These limitations provide important directions for future improvement.

To further explore the model's effectiveness across diverse attack categories, a class-wise performance evaluation was undertaken, particularly focusing on minority classes such as U2R and R2L in the NSL-KDD dataset. These attack types are typically underrepresented and more difficult to detect due to their subtle behavioural patterns and limited training samples.

High detection performance is maintained by the proposed model, even on minority classes, as indicated by the experimental results. Specifically, the recall rates for U2R and R2L attacks are substantially elevated when compared with conventional machine learning methods, demonstrating the advantage of the BiGRU and attention mechanisms in capturing subtle temporal and contextual features. The attention mechanism assists the model in highlighting discriminative features, while the BiGRU effectively models sequence dependencies, thereby reducing false negatives in minority attack detection.

Moreover, compared with baseline models such as SVM and NB, which exhibit noticeable performance degradation on imbalanced data, the proposed model achieves more balanced precision-recall trade-offs across all categories. This discovery validates both the model's resilience to class imbalance and its fitness for real-world intrusion detection, a context that demands the accurate identification of rare but critical attacks.

Overall, from classifier selection and feature extraction to architectural optimisation and benchmarking against existing approaches, the proposed model consistently demonstrated outstanding performance on the NSL-KDD dataset. Its modular design ensures complementary strengths across individual components, enabling a synergistic enhancement effect. Whether applied to traditional classification tasks, feature modelling, or real-world network intrusion detection, the model exhibits superior capability in recognising diverse attack behaviours and holds substantial practical value.

The comparative experimental results on the UNSW-NB15 dataset are shown in Table 2.

To comprehensively assess the performance of the proposed DNN-BiGRU-Attention-MLP model, a systematic set of experiments was performed across four dimensions: classifier selection, feature extraction methods, architectural optimisation, and comparison with mainstream intrusion detection algorithms. Results showed that classifier choice significantly affected performance, with the RF classifier yielding the best F1-score at 97.16%, indicating its strength in modelling high-dimensional, nonlinear data. When using a fixed MLP classifier, RF-based feature selection also outperformed other methods (LDA, ICA, RFE), reaching an F1-score of 98.89% by effectively preserving key features and filtering redundancy. Architectural optimisation further improved model performance as BiGRU, attention mechanisms, and MLP layers were progressively introduced, increasing overall accuracy and F1-score to 98.41% and 96.64%, respectively. Comparative experiments with existing models – including SDAE-ELM, BDNN, BCNN, and PV-DM – it was the proposed model that demonstrated superiority, recording the highest accuracy (98.43%) and F1-score (96.68%). These findings attest to the model's strong robustness and generalisation ability, and effectiveness in detecting complex, high-dimensional network intrusions, highlighting its potential as a reliable solution in practical cybersecurity applications.

Table 2 Comparative experimental results of UNSW-NB15

Comparative classification	*Model*	*Metric*			
		Accuracy	*Precision*	*Recall rate*	*F1-score*
Different classification methods	DNN-BiGRU-Attention-NB	74.54%	48.58%	98.31%	65.03%
	DNN-BiGRU-Attention-SVM	97.94%	99.95%	91.46%	95.54%
	DNN-BiGRU-Attention-DT	98.11%	95.80%	96.37%	96.08%
	DNN-BiGRU-Attention-KNN	98.31%	97.34%	95.57%	96.45%
	DNN-BiGRU-Attention-RF	98.14%	97.90%	96.44%	97.16%
Different feature extraction methods	LDA-MLP	97.86%	99.58%	91.51%	95.37%
	ICA-MLP	97.98%	99.81%	91.79%	95.63%
	REF-MLP	97.96%	98.85%	92.61%	95.61%
	RF-MLP	98.09%	99.56%	92.48%	98.89%
Different model	DNN	96.85%	96.13%	86.93%	91.30%
	DNN-BiGRU	98.37%	97.72%	95.47%	96.58%
	DNN-BiGRU-Attention	98.21%	99.62%	92.91%	96.15%
	DNN-BiGRU - MLP	98.41%	98.51%	94.81%	96.64%
Different algorithms	SDAE-ELM	72.20%	66.93%	97.85%	79.49%
	BDNN	80.63%	86.00%	81.00%	79.00%
	BCNN	90.25%	91.00%	90.00%	90.45%
	PV-DM	82.86%	82.86%	77.70%	80.20%
Paper model		*98.43%*	*98.54%*	*94.88%*	*96.68%*

Overall, experimental results on the NSL-KDD and UNSW-NB15 datasets highlight the model's adaptability and effectiveness across varying levels of data complexity. While

the NSL-KDD dataset features a relatively simple structure and well-defined attack types – enabling the model to achieve higher accuracy and F1-score due to its ability to capture structured patterns – the UNSW-NB15 dataset reflects more realistic network conditions, with higher feature dimensionality, diverse attack behaviours, and class imbalance. Despite the increased challenge, the model maintained excellent classification performance, demonstrating strong generalisation and robustness. In summary, the DNN-BiGRU-Attention-MLP model proposed in this study effectively integrates spatial and temporal feature representation, global perception, and adaptive feature focusing capabilities. It consistently delivers stable and high-performance results across different datasets, offering a powerful and practical solution for constructing efficient network IDS.

6 Conclusions

In order to overcome the weaknesses of traditional intrusion detection concerning high-dimensional feature modelling and cross-modal fusion, this paper puts forward a modular detection architecture that unifies DNN, BiGRU, and a self-attention mechanism. The model builds dual DA substructures (two DNN modules with self-attention) and dual BA substructures (two BiGRU modules with self-attention), forming a hierarchical architecture that jointly captures spatial and temporal characteristics of network traffic. Specifically, DNN extracts static high-dimensional features, BiGRU models bidirectional temporal dependencies, and self-attention emphasises critical feature dimensions and time steps, thereby improving anomaly recognition accuracy.

The deep features from all substructures are concatenated into a global representation and classified by a MLP. Compared with conventional models, the proposed approach advances both structural design and feature modelling: the layered fusion of DA and BA substructures addresses the limitations of single-network architectures, while the embedded attention mechanism reduces redundancy and enhances fine-grained feature selection. Experiments on NSL-KDD and UNSW-NB15 datasets verify the effectiveness of the method. Furthermore, Additional investigation into model limitations and future optimisation directions suggests that the proposed approach not only secures high detection accuracy but also furnishes a solid foundation for further improvements in computational efficiency and real-world deployment.

Future Work. Although the proposed method achieves high performance, several directions remain open for further research. First, extending the model to real-time environments would enhance its applicability in practical deployment scenarios. Second, The inclusion of online learning or incremental training mechanisms could enable system adaptation to evolving attack patterns while avoiding full retraining.

Acknowledgements

This work was supported by the China University Industry-University-Research Innovation Fund Project of the Ministry of Education:' Research on Deep Learning Network Threat Detection Based on Large Language Models' and the 'Suqian Yingcai'

Xiongying Plan Project 'Innovation and Practice of Artificial Intelligence Technology in Modern Agriculture'(SQXY202432).

Declarations

All authors declare that they have no conflicts of interest.

References

Aktar, S. and Nur, A.Y. (2023) 'Towards DDoS attack detection using deep learning approach', *Computers and Security*, Vol. 129, No. 7, p.103251.

Almomani, A., Akour, I., Manasrah, A.M. et al. (2023) 'Ensemble-based approach for efficient intrusion detection in network traffic', *Intelligent Automation and Soft Computing*, Vol. 37, No. 2, pp.137–145.

Alrayes, F.S., Zakariah, M., Amin, S.U. et al. (2024) 'CNN channel attention intrusion detection system using NSL-KDD dataset', *Computers, Materials and Continua*, Vol. 79, No. 3, pp.521–531.

Aslan, Ö., Aktuğ, S.S., Ozkan-Okay, M. et al. (2023) 'A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions', *Electronics*, Vol. 12, No. 6, p.1333.

Chaganti, R., Suliman, W., Ravi, V. et al. (2023) 'Deep learning approach for SDN-enabled intrusion detection system in IoT networks', *Information*, Vol. 14, No. 1, p.41.

Chen, D., Song, Q., Zhang, Y. et al. (2023) 'Identification of network traffic intrusion using decision tree', *Journal of Sensors*, Vol. 2023, No. 1, p.5997304.

El Asry, C., Douzi, S. and El Ouahidi, B. (2023) 'Intrusion detection system, a new approach to R2L and U2R attack classification', *The International Conference on Artificial Intelligence and Smart Environment*, Springer Nature Switzerland, Cham, pp.297–304.

Fu, Y., Du, Y., Cao, Z. et al. (2022) 'A deep learning model for network intrusion detection with imbalanced data', *Electronics*, Vol. 11, No. 6, p.898.

Halbouni, A., Gunawan, T.S., Habaebi, M.H. et al. (2022) 'CNN-LSTM: hybrid deep neural network for network intrusion detection system', *IEEE Access*, Vol. 10, No. 11, pp.99837–99849.

Hussien, M. (2023) 'Enhancing network security with a deep learning-based intrusion detection system for 5G networks', *RS Open J. Innov. Commun. Technol.*, October, Vol. 9, No. 3, pp.1–10.

Ige, T. and Kiekintveld, C. (2023) 'Performance comparison and implementation of bayesian variants for network intrusion detection', *2023 IEEE International Conference on Artificial Intelligence, Blockchain, and Internet of Things* (*AIBThings*), IEEE, pp.1–5.

Kasongo, S.M. (2023) 'A deep learning technique for intrusion detection system using a recurrent neural networks based framework', *Computer Communications*, Vol. 199, no. 18, pp.113–125.

Khraisat, A., Gondal, I., Vamplew, P. and Kamruzzaman, J. (2019) 'Survey of intrusion detection systems: techniques, datasets and challenges', *Cybersecurity*, December, Vol. 2, No. 1, p.20.

Kimanzi, R., Kimanga, P., Cherori, D. et al. (2024) *Deep Learning Algorithms Used in Intrusion Detection Systems – A Review*, arXiv preprint arXiv:2402.17020.

Mijwil, M.M., Salem, I.E. and Ismaeel, M.M. (2023) 'The significance of machine learning and deep learning techniques in cybersecurity: a comprehensive review', *Iraqi Journal For Computer Science and Mathematics*, Vol. 4, No. 1, p.10.

Mohamed, S. and Ejbali, R. (2023) 'Deep SARSA-based reinforcement learning approach for anomaly network intrusion detection system', *International Journal of Information Security*, Vol. 22, No. 1, pp.235–247.

Moustafa, N. and Slay, J. (2015) 'UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)', *2015 Military Communications and Information Systems Conference* (*MilCIS*), *IEEE*, pp.1–6.

Oboya, W.M., Gichuhi, A.W. and Wanjoya, A.A. (2023) 'Hybrid DNN-RBFNN model for intrusion detection system', *Journal of Data Analysis and Information Processing*, Vol. 11, No. 4, pp.371–387.

Qazi, E.U.H., Faheem, M.H. and Zia, T. (2023) 'HDLNIDS: hybrid deep-learning-based network intrusion detection system', *Applied Sciences*, Vol. 13, No. 8, p.4921.

Selvan, M.A. (2024) 'SVM-enhanced intrusion detection system for effective cyber attack identification and mitigation', *Journal of Big Data*, Vol. 21, No. 9, pp.443–451.

Tavallaee, M., Bagheri, E., Lu, W. et al. (2009) 'A detailed analysis of the KDD CUP 99 data set', *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, IEEE, pp.1–6.

Umair, M.B., Iqbal, Z., Faraz, M.A. et al. (2024) 'A network intrusion detection system using hybrid multilayer deep learning model', *Big Data*, Vol. 12, No. 5, pp.367–376.

Wang, Y.C., Houng, Y.C., Chen, H.X. et al. (2023) 'Network anomaly intrusion detection based on deep learning approach', *Sensors*, Vol. 23, No. 4, p.2171.

Yin, Y., Jang-Jaccard, J., Xu, W. et al. (2023) 'IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset', *Journal of Big Data*, Vol. 10, No. 1, p.15.

Zou, B. and Wu, Y. (2023) 'A review of network intrusion detection techniques based on deep learning', *Network Security Technology and Application*, Vol. 2023, No. 8, pp.3–8.s