# Multilevel and multisource data fusion approach for network intrusion detection system using machine learning techniques

**Harshitha Somashekar, Pramod Halebidu Basavaraju**
Department of Information Science and Engineering, Adichunchanagiri Institute of Technology affiliated to Visvesvaraya Technological University, Belagavi, India

**Article Info**

***Article history:***

Received Oct 7, 2024
Revised Mar 26, 2025
Accepted May 24, 2025

***Keywords:***

Data fusion
Dimensionality reduction
Network intrusion detection
Neural network
Prediction level fusion

## ABSTRACT

To enhance the performance of network intrusion detection systems (NIDS), this paper proposes a novel multilevel and multisource data fusion approach, applied to NSL-KDD and UNSW-NB15 datasets. The proposed approach includes three various levels of operations, which are feature level fusion, dimensionality reduction, and prediction level fusion. In the first stage features of NSL-KDD and UNSW-NB15 both datasets are fused by applying the inner join joint operation by selecting common features like protocol, service and label. Once the data sets are fused in the first level, linear discriminant analysis is applied for 12 feature columns which is reduced to a single feature column leading to dimensionality reduction at the second level. Finally, in the third level, the prediction level fusion technique is applied to two neural network models, where one neural network model has a single input node, two hidden nodes, and two output nodes, and another model having a single input node, three hidden nodes, and two output nodes. The outputs obtained from these two models are then fused using a prediction fusion technique. The proposed approach achieves a classification accuracy of 97.5%.

***Corresponding Author:***

Harshitha Somashekar
Department of Information Science and Engineering, Adichunchanagiri Institute of Technology, affiliated to Visvesvaraya Technological University
Belagavi 590018, Karnataka, India
Email: sh@mcehassan.ac.in

## 1. INTRODUCTION

Network intrusion detection systems (NIDS) have become one of the prime mechanisms for shielding or protecting computer networks from various types of malicious attacks [1]. Basically, it monitors network traffic for any suspicious activities from the hackers. In case of any remarkable or unrelated event, it informs or gives the signal to the administrators. Traditionally, intrusion detection systems have relied on single-source data, and like network traffic logs or system event logs. Although it was very effective in certain aspects, these methods usually fail to detect highly sophisticated and continuously evolving threats due to complex network structures, especially the ones that include several stages or different sources [2]. For example, an attack may consist of some unusual traffic patterns, unauthorized access attempts, or anomalous system behaviour. This would be pretty hard to detect in a single log source of data.

These limitations however controlled through data fusion techniques by the researchers which are discussed in the literature. Data fusion refers to a technique for merging information from several sources in order to improve the accuracy and robustness of intrusion detection. It becomes possible to create a more complete picture of network activities by integrating data from various logs and sensors like network traffic, system events, and application logs. Such a holistic approach can also boost an improved identification of known and unknown threats in the network by enabling the correlation of disparate data of evidence that might otherwise go unnoticed. NIDSs are one of the key aspects of a computer network for providing security and integrity, so NIDSs should be very accurate as advanced detection mechanisms against ever-increasing cyber threats. Researchers worked on different machine learning algorithms for building effective intrusion detection classifiers, which are neural networks, support vector machines (SVMs), and k-nearest neighbors (KNN) all having distinct advantages and challenges.

In many typical instances neural network will be able to learn very complex patterns, hence it is able to find anomalies which could be an intrusion. Since neural network models make use of a large amount of training data and high computational resources to analyses the sequence, it will be a drawback for the detection of intrusions. Another popular method in finding intrusion detection is SVM, it is more popular because they are capable of handling high-dimensional data and perform well even with a small training sample. In some cases, SVMs may be more sensitive while choosing the hyperparameters, which have to be carefully adjusted, else model will predict useful data as an intrusion. The KNN is a very simple algorithm that will work in some cases but they often lead to a high false-positive rate, mostly whenever the data have more noise.

The major drawback of using traditional machine learning methods are these models always leads to a high number of false positive rates and false alarms. In order to overcome these issues, researchers are focusing to create better updated methods which can improve detection rate. In order to create better models researchers are focusing on multi-source data fusion approaches for network intrusion detection. By studying the recent literature in multi-source data, it is concluded that the key concepts, techniques and applications of these newly developed approaches showed better results when compared to traditional machine learning (ML) models.

In order to build a strong intrusion detection system, the intrusion detection systems (IDS) should be trained on variety of network intrusions. These intrusions have limited scope in individual type of data sets. For example, the NSL-KDD dataset which we considered in this work focuses on traditional network attacks like DoS, Probe, R2L, and U2R, while another one UNSW-NB15 dataset has considered modern threats such as botnets, worms, and advanced persistent threats (APTs). By integrating both NSDL-KDD and UNSW-NB15 data sets a new fused data sets can be generated for training, by this we can train our model with the variety of intrusion scenarios, which makes the system more robust and ability to detect various intrusions effectively. However, the fusion process results in a large feature set of data which contains both NSL-KDD and UNSW-NB15 features, which increases computational complexity. Reducing the dimensionality of features plays a vital role, which can also contribute to optimize model performance. Further, handling a variety of input data sets remains a significant challenge, which can be handled by adopting prediction-level-fusion through diverse neural network architectures. The system's reliability can be improved and also ensures more accurate detection of new intrusions.

The proposed model achieved 97.5% accuracy which outperformed several existing methods. All previous ML-based approaches are applied to single types of data sets. In this paper, we proposed a state-of-art work, that is the fusion of data sets with a multi-level fusion approach. The existing models like SVM, random forest (RF), and ensemble models reported 90%–95% accuracy. Deep learning-based methods, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) achieved 92%–96% accuracy, which also required high computational resources for execution. In some work, they utilized stacking models which reported 92%–95% accuracy, but these did not effectively incorporate multisource data fusion. In our proposed work feature-level fusion technique improved instruction detection accuracy by integrating two varieties of datasets (NSL-KDD and UNSW-NB15) at the initial stage, in the second stage the linear discriminant analysis (LDA)-based technique was used for dimensionality reduction which optimized the computational efficiency. Additionally, in the last stage, a prediction-level fusion of two different neural network models was used to further enhance the performance. The obtained results confirmed the effectiveness of using a multilevel fusion approach in cybersecurity to identify intrusion attacks. By highlighting its potential for real-time intrusion detection, the model overcomes the computational limitations of traditional ML and deep learning methods.

This study aligns with ongoing research in cybersecurity, machine learning, and data fusion, making it highly relevant to the scope of computer engineering or network security. By addressing the limitations of traditional IDS models and introducing a novel multilevel fusion strategy, this presented work provides significant contributions to the field of intelligent network security. The primary contributions of this research are: i) a novel multilevel fusion strategy that improves IDS detection accuracy, ii) the integration of two diverse datasets (NSL-KDD and UNSW-NB15) to cover a wide range of cyber threats, and iii) a hybrid neural network-based prediction-level fusion that enhances classification performance.

## 2. LITERATURE SURVEY

From recent years research on using multisensory data fusion in intrusion detection systems has been growing recently. Multisensory data fusion combines information from multiple sensors or data sources to obtain more accurate and complete information than using a single source of data. In intrusion detection, the data fusion method can utilize various data sources such as network traffic logs, system event logs, and user activity logs. Hall and Llinas [3] in their research work showed an overview of multisensory data fusion approach by covering key process models and techniques. Their research demonstrated the use of data fusion in practical applications like automated target recognition, battlefield surveillance, and complex machinery monitoring, which are closely related to network intrusion detection.

The limitations of traditional intrusion detection methods can be overcome using multi-sensor data fusion approaches, since traditional methods rely on single source of information. Rahul-Vigneswaran *et al.* [4] suggested that many existing intrusion detection systems suffers from high false detection rates and lack in decision support for the incident. By combining data from different sources, we can increase the strength of data sets and also features of intrusions for wide range of attacks can be considered.

Intrusion detection systems (IDS) play a key role in safeguarding modern networks against every day evolving various types of cyber threats. Traditional techniques fail to manage the today's complex network attacks. To address these issues, researchers started using advanced machine learning models, which have showed promising results by enhancing detection rates [5], [6]. One of the most promising techniques is to adopt data fusion approaches in NIDS. Some works are explored in this area [7], showed some promising results. Hence, researchers started using data fusion approaches to overcome limitations of traditional methods. These models excel at handling large-scale data and creating more effective representations which significantly improve intrusion detection rate [7], [8].

The primary goal of an IDS is to identify and prevent both known and unknown cyber threats in network environments. Traditional methods struggle to keep up with the fast-evolving threat landscape, often resulting in decreased accuracy as data sets grow very fast. By recognizing these shortcomings, researchers have focused on machine learning algorithms like neural networks, SVMs, and KNN to improve IDS [9]. However, these methods also face challenges with high false positive and alarm rates. To overcome these issues, researchers have explored deep learning models, particularly long short-term memory (LSTM) networks [10], which are better at processing and classifying complex data patterns.

Integrating data fusion techniques with neural network learning models can further enhance IDS performance. By fusing multiple data sources and manipulating the Neural network model for powerful feature extraction, IDS can more effectively detect and mitigate a wide range of network threats [2], [5], [7], [8], [11]. Unlike the previous studies that relied on single-source intrusion detection or traditional machine learning models, this research work demonstrates that the multisource data fusion approach significantly enhances intrusion detection accuracy. The proposed model bridges the gap between feature-level fusion and deep learning-based prediction fusion, setting a new benchmark for NIDS performance.

## 3. METHODOLOGY

Engineered approaches and techniques in the field of data fusion focus on effectively merging information from multiple sources of data. The key fusion architectures include centralized, decentralized, and distributed models, each involving various levels of abstraction in the fusion process like raw measurements, signals and characteristics, or decisions. At lower abstraction levels, the fusion techniques may work with raw sensor measurements or signal-level data, using statistical methods.

This paper describes a multilevel and multisource data fusion approach designed to enhance performance in network intrusion detection systems and address these challenges. In this paper, a novel method was proposed where data fusion was applied at various stages and also data dimensionality reduction was applied to improve the performance of classification accuracy. The proposed model with multi-level multi-source data fusion is shown in Figure 1. The model we built used data fusion approaches, because the traditional IDS struggle with limited data sources, hence combining multiple sources of data improves detection accuracy. We used the dimensionality reduction technique since high-dimensional data leads to computational inefficiencies, so LDA optimizes feature selection and reduces the data size. We applied a prediction-level fusion approach because Single-model approaches have limitations, but combining multiple neural network outputs improves robustness. This paper provides mathematical formulations (1)-(8) for each execution step and also the Algorithmic descriptions provided for the proposed multilevel and multisource data fusion approach to ensure that researchers can replicate the method.

### 3.1. Level 1: Feature-level fusion

In the first level, features from datasets NSL-KDD and UNSW-NB15 are fused using an inner join technique based on common features such as protocol, service, and label. The fusion of features from these sources makes a detailed perception of data happening in the network under observation, which will enhance better detection of both known and unknown attacks or threats. The proposed methods for feature-level fusion in level 1 are shown in mathematical form.

a. Feature selection on the NSL-KDD dataset:

$$F_{NSL-KDD}=\text{SelectFeatures}\ (D_{NSL-KDD},15) \tag{1}$$

where $D_{NSL-KDD}$ is the NSL-KDD dataset, and $F_{NSL-KDD}$ are the selected feature.

b. Selecting features from the UNSW-NB15 dataset:

$$F_{UNSW-NB15} = \text{SelectFeatures}\ (D_{UNSW-NB15},15) \tag{2}$$

where $D_{UNSW-NB15}$ is the UNSW-NB15 dataset, and $F_{UNSW-NB15}$ are the selected features.

c. Dataset fusion inner join the selected features:

$$F_{fused} = F_{NSL-KDD} \cap F_{UNSW-NB15} \tag{3}$$

where the intersection is performed on the columns protocol, service, and label.

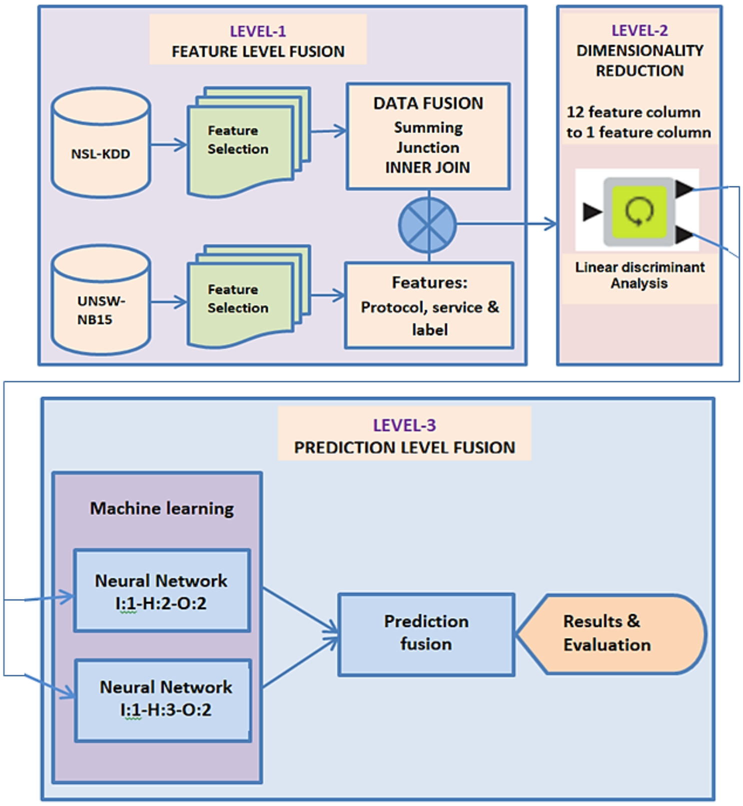


Figure 1. The proposed approach involves three levels of data fusion

### 3.2. Level 2: Dimensionality reduction

In the proposed model the second level of the approach involves dimensionality reduction [12]. The 12 feature columns from the fused dataset are reduced to a single feature column using LDA. This dimensionality reduction step helps to mitigate the curse of dimensionality and improve the efficiency of the subsequent prediction models. The proposed methods for dimensionality reduction in level 2 in mathematical form.

a. Dimensionality reduction using LDA apply LDA:

$$X_{LDA} = LDA(F_{fused}, n = 12, m = 1) \tag{4}$$

### 3.3. Level 3: Prediction-level fusion

The last level in the proposed approach is prediction-level fusion [13]. Here, two neural network models have been used, out of which one has a single input node, two hidden nodes, and two output nodes, while the second model has a single input node, three hidden nodes, and two output nodes. These outputs from the two neural network models are then fused using a prediction fusion technique. In prediction fusion, multiple models fuse to upgrade the strengths of the different neural models for overall improved performance in classification accuracy. The level 3 prediction level fusion is shown below in mathematical form.

a. Neural network model 1

$$O_1 = NN_1(X_{LDA},[1], [2], [2],sigmoid) \tag{5}$$

where O1 is the output of the first neural network model.

b. Neural network model 2

$$O_2 = NN_2(X_{LDA},[1],[3],[2],softplus) \tag{6}$$

where O2 is the output of the second neural network model.

c. Output fusion: Fuse outputs from the two models:

$$Ofused = \frac{O1 + O2}{2} \tag{7}$$

Apply the dense network threshold:

$$O_{final} = \{ \text{ 1 if } O_{fused} > T, \text{ O otherwise} \tag{8}$$

where T is the dense network threshold.

d. Evaluation of the fused output:

$$R = Evaluate(O_{final})$$

where R is the evaluation result.

Equation (1) to (8) gives the formation of data at each level in the mathematical form for analysis.

## 4. PROPOSED ALGORITHM: MULTILEVEL AND MULTISOURCE DATA FUSION (MMDF) APPROACH FOR NIDS

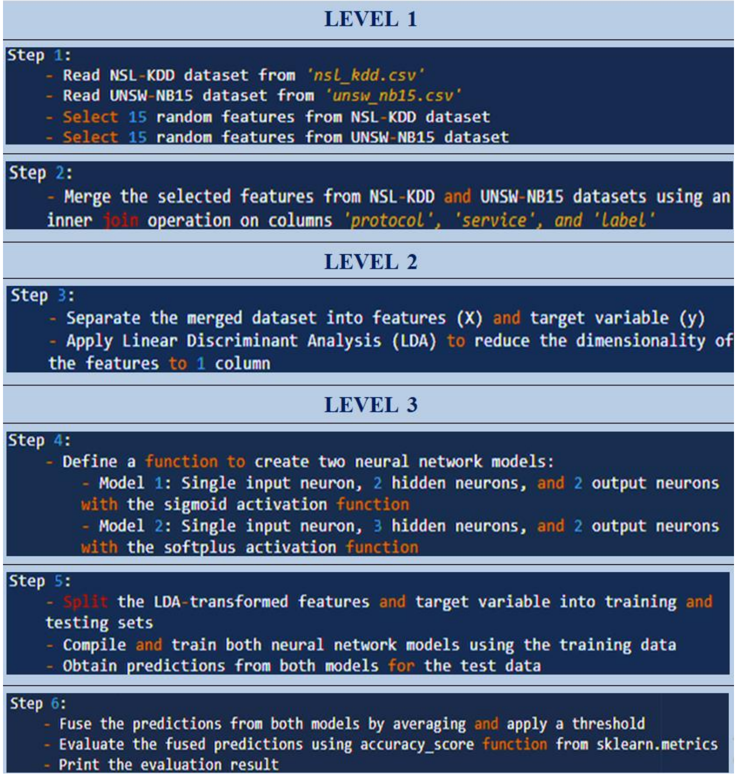
The MMDF methodology was implemented in a NIDS utilizing the following algorithm. The method itself is broken down into three main stages intended to improve detection relevancy and calculation time. In the initial stage, data preprocessing and feature selection are the focus. For this example, we load the NSL-KDD and UNSW-NB15 datasets that represent different types of network traffic flows. To obtain a rich and informative feature set, 15 features are extracted from each dataset. The resultant features are then joined using an inner join based on common attributes (i.e., protocol, service, and label). It is an essential step in preparing the final dataset, retaining only relevant data from both original datasets.

This second stage relates to feature transformation designed to make calculations more efficient, as well as to improve model accuracy. At this stage, the merged dataset is split into independent variables (X) and target variables (y), ensuring structured data organization for training. Since the feature space is high dimensional, this uses LDA to condense multiple features into a single, optimized representation. This step reduces dimensions but retains the necessary information for maintaining separability among different classes while considerably reducing complexity, thus making the dataset more amenable for classification using machine learning-based techniques.

The last stage is focused on the development, training & evaluation of the model. We define two different neural networks, with different complexities. In the first model, we have one input layer - two hidden layers - and two output nodes using a sigmoid activation function. The second model has one input layer, three hidden layers, two output nodes, and a soft plus activation function. We then split our dataset into train and test datasets and train both models separately. Following that, their predictions on test data are ensembled using an average technique with a defined threshold to improve the classification ratio. Finally, the accuracy_score function checks the final predictions, providing a comprehensive performance evaluation. Using this technique, we can mitigate the weaknesses of single models effectively, and improve the capabilities of the intrusion detection process in terms of accuracy and robustness by integrating multiple models and data sources.

The step-by-step pseudo code of multilevel and multisource data fusion approach for NIDS is shown in the Figure 2. The general overview of each steps are:

- At the beginning the both NSL-KDD and UNSW-NB15 datasets were normalized. The features protocol, service and labels are used as key columns for fusion.
- Performed an inner join operation on selected features to combine features by applying feature level fusion technique. Which helped in keeping all types of attacks from both the data sets in a single fused data set.
- By applying LDA the feature column 12 is reduced to 1, which increases the performance of the model.
- Neural network architectures: Model 1: 1 input layer, 2 hidden, 2 output, activation = sigmoid Note: Replaced activation with a less used activation function soft plus.
- Fusion: The outputs of both models were averaged for effective decision-making. Model performance & Evaluation: accuracy, precision, recall and F1-scores were calculated to compare performance.



LEVEL 1

```
Step 1:
    - Read NSL-KDD dataset from 'nsl_kdd.csv'
    - Read UNSW-NB15 dataset from 'unsw_nb15.csv'
    - Select 15 random features from NSL-KDD dataset
    - Select 15 random features from UNSW-NB15 dataset
```

```
Step 2:
    - Merge the selected features from NSL-KDD and UNSW-NB15 datasets using an
    inner join operation on columns 'protocol', 'service', and 'label'
```

LEVEL 2

```
Step 3:
    - Separate the merged dataset into features (X) and target variable (y)
    - Apply Linear Discriminant Analysis (LDA) to reduce the dimensionality of
    the features to 1 column
```

LEVEL 3

```
Step 4:
    - Define a function to create two neural network models:
        - Model 1: Single input neuron, 2 hidden neurons, and 2 output neurons
        with the sigmoid activation function
        - Model 2: Single input neuron, 3 hidden neurons, and 2 output neurons
        with the softplus activation function
```

```
Step 5:
    - Split the LDA-transformed features and target variable into training and
    testing sets
    - Compile and train both neural network models using the training data
    - Obtain predictions from both models for the test data
```

```
Step 6:
    - Fuse the predictions from both models by averaging and apply a threshold
    - Evaluate the fused predictions using accuracy_score function from sklearn.metrics
    - Print the evaluation result
```

Figure 2. Algorithm MMDF- multilevel and multisource data fusion approach for NIDS

## 5. EXPERIMENTAL RESULTS

The proposed multilevel and multisource data fusion approach achieved a classification accuracy of 97.5%. The most efficient accuracy results achieved are because of different fusion techniques used at various stages. Feature-level fusion combined useful features from both NSL-KDD and UNSW-NB15 datasets, which provided a more complete view of network activity, which in turn helped in improving the detection of both known and unknown attacks. The LDA is used for dimensionality reduction, which helped to overcome the issues with large datasets and made the prediction models more efficient. Additionally, combining the two different neural network models at the prediction level took advantage of their unique strengths for improving classification accuracy.

Previous research has shown the benefits of utilizing information fusion in areas like fault diagnosis and identifying fault detection. This research work demonstrates the effectiveness of a multilevel and multisource data fusion approach for network intrusion detection, which is pivotal for cybersecurity. The simulation model is shown in Figure 3.

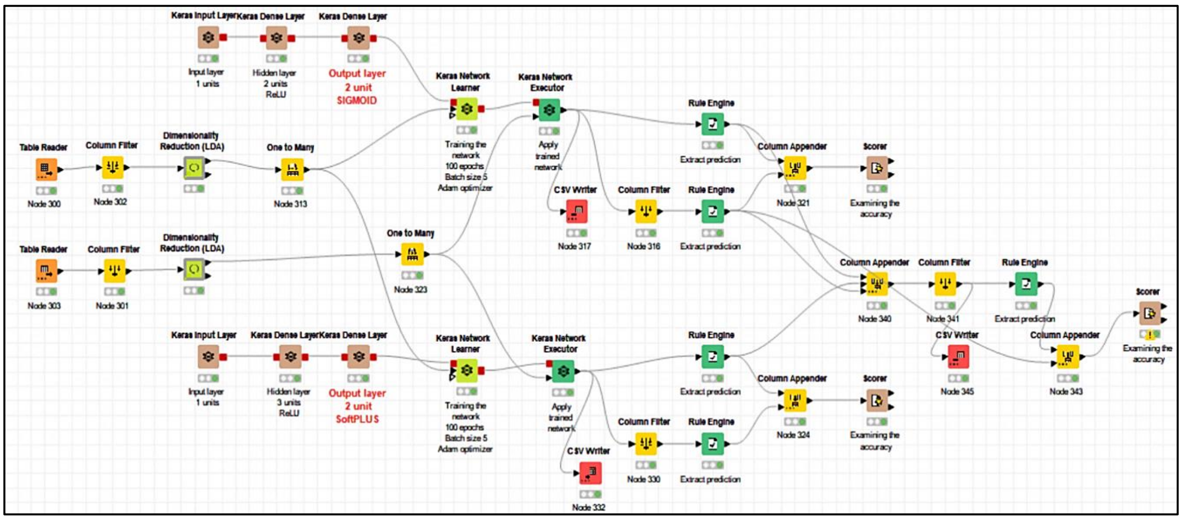


Figure 3. The proposed approach simulation model using KNIME

The KNIME workflow in the Figure 3 shows the proposed approach Simulation model. The simulation steps are listed below, all bold words in steps represents each node in KNIME.

a. Data preprocessing:
- Two table readers load datasets.
- Column filters remove unnecessary features.
- Dimensionality reduction (LDA) reduces feature dimensions.
- One to many nodes split data for multiple models.

b. Neural network training:
- Keras input layers define input structures.
- Keras dense layers create hidden layers with activation functions (ReLU and Sigmoid/SoftPLUS).
- Keras network learners train models using backpropagation and Adam optimizer.

c. Model execution and prediction:
- Keras network executors apply trained models.
- CSV writers store results.

d. Prediction Processing:
- Rule engines extract predictions.
- Column filters refine outputs.
- Column appenders combine predictions.

e. Evaluation:
- Scorers assess model accuracy.
- CSV writers store evaluation metrics.

The proposed workflow effectively handles input data, feature fusion, deep learning training, prediction fusion and model validation in KNIME.

Table 1 shows the results of experimentation, where the proposed model upholds the accuracy. The proposed model is able to find the network attacks of both data sets used for experimentation. Figure 4 shows the comparison graph, where the proposed method showed a prominent result.

Several studies have explored machine learning-based intrusion detection, with experiment models such as support vector machines (SVM), k-nearest neighbors (KNN), decision trees (DT), and random forests (RF). Table 2 shows the comparison of results with the existing models. Our multilevel and multisource data fusion approach surpasses deep learning-based IDS in three key ways: Better accuracy (97.5%) than CNNs and RNNs, due to multisource data integration. Lower false positives (2.8%), as prediction-level fusion refines classification results. Reduced dimensionality, as LDA decreases the number of predictor variables, while deep learning techniques often work with such high-dimensional input. Deep learning-based methods like LSTMs and CNNs require significant computational resources. However, our fusion-based approach offers a good balance between accuracy and efficiency, making it a practical choice for real-time cybersecurity applications. Key benefits of our approach include:

– Scalability: By integrating multiple datasets, it enables broader attack detection.
– Reduced processing overhead: Techniques like feature-level fusion and dimensionality reduction help minimize memory and processing power requirements.
– Enhanced adaptability: The fused model can quickly adjust to evolving cyber threats.

Overall, our method improves upon previous models by combining feature fusion, dimensionality reduction, and neural network prediction fusion.

Table 1. Results of classification models

| Sl.No | Classification model | Classification accuracy |
|---|---|---|
| 1 | ANN-10 hidden layers | 95.74 |
| 2 | ANN2-12 hidden layers | 80.49 |
| 3 | ANN-2 hidden layer with LDR data | 97.01 |
| 4 | ANN-3 hidden layer with LDR data | 97.06 |
| 5 | **Proposed fused model** | **97.49** |

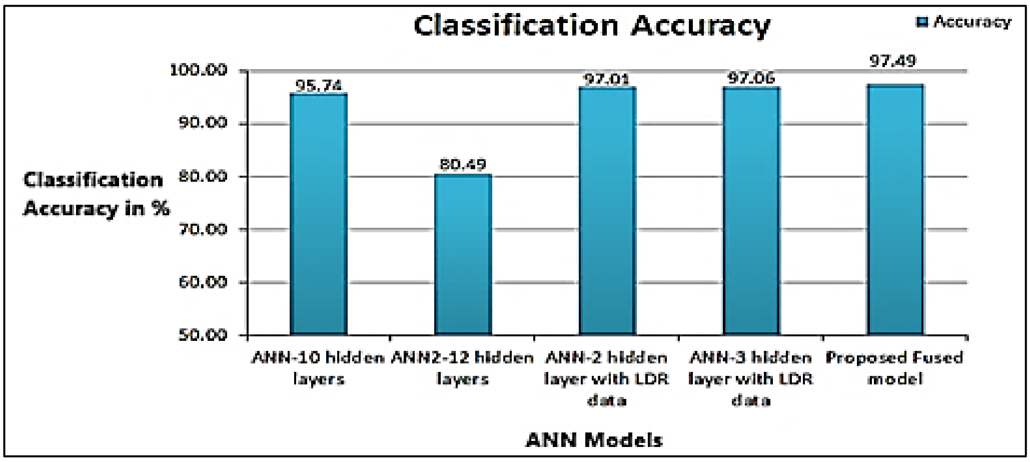


Figure 4. Comparison of classification models

Table 2. Comparison with existing methods

| Method | Data sets used/fused | Accuracy (%) | False positives (%) | Computational cost |
|---|---|---|---|---|
| SVM | Single | 91.2 | 5.8 | Medium |
| CNN | Single | 93.4 | 4.5 | High |
| RNN | Single | 94.1 | 4.2 | Very High |
| Ensemble | Single | 95.0 | 5.0 | Very High |
| Proposed Model | Multiple/Fused | 97.5 | 2.8 | Medium |

## 6. COMPARISON OF PROPOSED WORK WITH OTHER RESEARCH WORK

The proposed approach performed better than other research works on the NSL-KDD and UNSW-NB15 datasets. Previous works achieved classification accuracies ranging from 90% to 95% on these datasets [14]. Prior studies have achieved promising results with models employing traditional machine learning algorithms such as ensemble models [15], SVM, and RFs, but frequently underperformed due to the complexity and diversity of network data [16]. In [17], a feature selection method for network intrusion detection achieved 88 to 91% accuracy. The researchers showed that in some cases the deep learning approaches, CNN and RNN achieved accuracy 92% to 96% [5], [18]. But these approaches required more computation time and space particularly during the implementation of multisource data [19]–[21]. In the recent approach, IDS takes on Big data by applying ML methods [22]. Some work showed that feature-level fusion techniques can integrate complementary information from each dataset [23], and they also improved the detection rate of both known and unknown attacks. The researcher worked on MANET by using KDD data sets which showed an accuracy of about 75% [24]. By considering all the recent approaches, in our work we used feature-level fusion followed by dimensionality reduction and additionally, prediction-level fusion enhances the performance of neural network models by leveraging their strengths. The experimental results show that network intrusion detection is a valuable application in cybersecurity, providing effective solutions and insights for future improvements. This approach serves as a strong foundation for further research, as multilevel and multisource methods generally outperform traditional single-source techniques in cybersecurity. The proposed method achieved a classification accuracy of 97.5%. There is no model which is tested on both type of datasets [25].

Comparison with previous work: Conventional ML models (SVM and RF) gave 90-95% accuracy with an increased false positive ratio. Several deep learning-based IDS models, such as CNN and RNN achieved 92% to 96% accuracy but needed substantial computational resources. The proposed multilevel data fusion model takes advantage of multiple features and achieves 97.5% accuracy, which is better than the previous models. The datasets NSL-KDD & UNSW-NB15 for attack detection showed various forms of attacks, which are integrated with multiple threat types through feature-level fusion. Dimensionality reduction technique LDA reduces computational complexity by focusing on relevant features. At prediction-level fusion, the fusion of two neural nets increases the robustness and accuracy of classification.

Even with the effective outcome, the proposed approach faces key challenges. Its recursive fusion structure increases data dimensionality proportionally and require powerful LDA for data reduction, since in real-time deployment LDA demands high processing power. The model tested on NSL-KDD and UNSW-NB15 datasets needs further validation for real-world networks, example enterprise and cloud environments. Additionally, it remains vulnerable to adversarial attacks such as data poisoning and evasion. To improve its applicability, the future research must focus on testing the model with live network traffic in real-time cyber security environment. Implement algorithms that learn and refine themselves as new patterns of malicious activity emerge. Improve prediction-level fusion-integrate multiple AI models such as graph neural networks (GNNs) or LSTM networks for sequentially attacking analysis. By addressing these areas, multisource data fusion-based NIDS can evolve into a more robust, adaptive, and scalable solution for modern cybersecurity challenges.

## 7. CONCLUSION

This study demonstrates the effectiveness of a using multilevel and multisource data fusion approach in enhancing network intrusion detection systems. The proposed methodology combines features from the NSL-KDD and UNSW-NB15 datasets and reduces the dimensionality of the datasets through linear discriminant analysis technique, and further prediction-level fusion of two neural network models is applied to achieve a classification accuracy of 97.5%. The success of the proposed approach is due to the extensive representation of network activities through feature-level fusion in the first stage, the efficiency obtained from dimensionality reduction in the second stage, and further improved performance resulting from prediction-level fusion at the last stage. The results clearly show the advantages of applying data fusion techniques and highlight their potential to improve cyber security schemes against known and unknown types of threats. The proposed multilevel multisource fusion approach holds promise for future research and development in NIDS. This research presents a novel data fusion approach that significantly enhances network intrusion detection accuracy. Given its contributions to cybersecurity and artificial intelligence, this study is highly relevant to the scope of the computer science field and can serve as a foundation for future research in intelligent threat detection.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Harshitha Somashekar | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |  |  | ✓ |  |
| Pramod Halebidu Basavaraju |  |  |  |  |  | ✓ |  |  |  | ✓ | ✓ | ✓ | ✓ |  |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis

I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting

Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## REFERENCES

[1] T. R. Devi and S. Badugu, "A review on network intrusion detection system using machine learning," in *Advances in Decision Sciences, Image Processing, Security and Computer Vision*, 2020, pp. 598–607.

[2] N. S. Bhati, M. Khari, V. García-Díaz, and E. Verdú, "A review on intrusion detection systems and techniques," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 28, pp. 65–91, 2020, doi: 10.1142/S0218488520400140.

[3] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997, doi: 10.1109/5.554205.

[4] Kunal and M. Dua, "Machine learning approach to IDS: A comprehensive review," in *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2019, pp. 117–121, doi: 10.1109/ICECA.2019.8822120.

[5] L. Ashiku and C. Dagli, "Network intrusion detection system using deep learning," *Procedia Computer Science*, vol. 185, pp. 239–247, 2021, doi: 10.1016/j.procs.2021.05.025.

[6] B. P. L. Lau *et al.*, "A survey of data fusion in smart city applications," *Information Fusion*, vol. 52, no. January, pp. 357–374, 2019, doi: 10.1016/j.inffus.2019.05.004.

[7] S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, and H. Karimipour, "Cyber intrusion detection by combined feature selection algorithm," *Journal of Information Security and Applications*, vol. 44, pp. 80–88, 2019, doi: 10.1016/j.jisa.2018.11.007.

[8] R. Ahmad and I. Alsmadi, "Data fusion and network intrusion detection systems," *Cluster Computing*, vol. 27, no. 6, pp. 7493–7519, Sep. 2024, doi: 10.1007/s10586-024-04365-y.

[9] I. F. Kilincer, F. Ertam, and A. Sengur, "Machine learning methods for cyber security intrusion detection: Datasets and comparative study," *Computer Networks*, vol. 188, no. January, p. 107840, 2021, doi: 10.1016/j.comnet.2021.107840.

[10] F. Laghrissi, S. Douzi, K. Douzi, and B. Hssina, "Intrusion detection systems using long short-term memory (LSTM)," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00448-4.

[11] A. Khraisat and A. Alazab, "A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges," *Cybersecurity*, vol. 4, no. 1, 2021, doi: 10.1186/s42400-021-00077-7.

[12] S. Suhana, S. Karthic, and N. Yuvaraj, "Ensemble based dimensionality reduction for intrusion detection using random forest in wireless networks," in *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Jan. 2023, pp. 704–708, doi: 10.1109/ICSSIT55814.2023.10060929.

[13] H. Somashekar and R. Boraiah, "Network intrusion detection and classification using machine learning predictions fusion," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 2, pp. 1147–1153, Aug. 2023, doi: 10.11591/ijeecs.v31.i2.pp1147-1153.

[14] S. Bhosale and S. M. Kamalapurkar, "A novel approach for network intrusion detection system using machine learning algorithms," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 8, no. 4, pp. 1–6, 2018.

[15] M. M. Otoom, K. N. A. Sattar, and M. Al Sadig, "Ensemble model for network intrusion detection system based on bagging using J48," *Advances in Science and Technology Research Journal*, vol. 17, no. 2, pp. 322–329, 2023, doi: 10.12913/22998624/161820.

[16] Y. Chang, W. Li, and Z. Yang, "Network intrusion detection based on random forest and support vector machine," in *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 2017, vol. 1, pp. 635–638, doi: 10.1109/CSE-EUC.2017.118.

[17] S. M. Kasongo and Y. Sun, "Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset," *Journal of Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00379-6.

[18] I. Idrissi, M. Boukabous, M. Azizi, O. Moussaoui, and H. El Fadili, "Toward a deep learning-based intrusion detection system for IoT against botnet attacks," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 1, pp. 110–120, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp110-120.

[19] J. Kim and A. Benton, "Evaluating deep learning approaches to intrusion detection in wireless networks," in *Proceedings of the 2016 IEEE Conference on Network Function Virtualization and Software Defined Networks, Palo Alto, CA, USA*, 2016, pp. 1–6.

[20] H. H. Jazi, H. Gonzalez, N. Stakhanova, and A. A. Ghorbani, "Detecting HTTP-based application layer DoS attacks on web servers in the presence of sampling," *Computer Networks*, vol. 121, pp. 25–36, Jul. 2017, doi: 10.1016/J.COMNET.2017.03.018.

[21] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Jul. 2009, pp. 1–6, doi: 10.1109/CISDA.2009.5356528.

[22] M. A. Alaketu *et al.*, "Comparative analysis of intrusion detection models using big data analytics and machine learning techniques," *The International Arab Journal of Information Technology*, vol. 21, no. 2, 2024, doi: 10.34028/iajit/21/2/14.

[23] A. Ayantayo *et al.*, "Network intrusion detection using feature fusion with deep learning," *Journal of Big Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-023-00834-0.

[24] M. Sasikumar and K. Rohini, "Expedient intrusion detection system in MANET using robust dragonfly-optimized enhanced naive Bayes (RDO-ENB)," *International Journal of Computer Networks and Applications*, vol. 11, no. 1, pp. 46–60, 2024, doi: 10.22247/ijcna/2024/224435.

[25] A. R. Al-Ghuwairi, Y. Sharrab, D. Al-Fraihat, M. AlElaimat, A. Alsarhan, and A. Algarni, "Intrusion detection in cloud computing based on time series anomalies utilizing machine learning," *Journal of Cloud Computing*, vol. 12, no. 1, 2023, doi: 10.1186/s13677-023-00491-x.

## BIOGRAPHIES OF AUTHORS

**Harshitha Somashekar** received a degree in Bachelor of Information Science and Engineering from Visvesvaraya Technological University, Belgaum, Karnataka, India, and M.Tech in Computer Networks Engineering from Visvesvaraya Technological University Belgaum, Karnataka. India. Currently, she is pursuing a PhD in Computer Science and Engineering at Adichunchanagiri Institute of Technology, Chikkamagaluru, Affiliated to Visvesvaraya Technological University, Belgaum Karnataka, India. She is currently working as an Assistant Professor in the Department of Computer Science & Engineering at Malnad College of Engineering, Hassan Karnataka, India. She has 9 years of teaching experience. Her research interest includes cyber security, artificial intelligence, artificial neural network, deep learning, and machine learning. She has published papers in international conferences and journals. She can be contacted at: sh@mcehassan.ac.in. Other sosial media: https://www.researchgate.net/profile/Harshitha-Somashekar.

**Pramod Halebidu Basavaraju** has experience of 12 and above years as an academician, currently working as an associate professor in the Department of Information Science and Engineering, Adichunchanagiri Institute of Technology, Chikkamagaluru, Affiliated to Visvesvaraya Technological University, Belgaum Karnataka, India. She has published papers in international journals and conferences. He received a Bachelor of Engineering degree in Computer Science and Engineering from the VTU in the year of 2007, and a Master of Technology degree in Computer Science from the University of Mysore, in the year 2012. He received a doctorate degree, PhD in the field of wireless sensor networks from the Department of CSE, Shri Venkateshwara University, Uttar Pradesh in the year of 2019. His research area includes wireless sensor networks, network security, and data analytics. He can be contacted at: hbpramod@aitckm.in. Other sosial media: https://www.researchgate.net/profile/Pramod-H-B.