Simulation Analysis of Optical Sensor-Based Intrusion Detection Using Machine Learning Algorithms

Kavitha Thandapani, Akshit Kasanagottu, Jayasurya Pasupula and Sriharsha Daggubati

May 12, 2023

SIMULATION ANALYSIS OF OPTICAL SENSOR-BASED INTRUSION DETECTION USING MACHINE LEARNING ALGORITHMS

DR. T KAVITHA
PROFESSOR/ECE. VEL TECH UNIVERSITY
VEL TECH RANGARAJAN DR SAGUNTHALA R and D INSTITUTE OF SCIENCE AND TECHNOLOGY
CHENNAI, INDIA
kavithaecephd@gmail.com

KASANAGOTTU AKSHIT
ECE.VEL TECH UNIVERSITY
VEL TECH UNIVERSITY
CHENNAI, INDIA
vtu15415@veltech.edu.in

Pasupula Jaya Surya
ECE.VEL TECH UNIVERSITY
VEL TECH UNIVERSITY
CHENNAI, INDIA
vtu15442@veltech.edu.in

Daggubati Sri Harsha,
ECE.VEL TECH UNIVERSITY
VEL TECH UNIVERSITY
CHENNAI, INDIA
vtu14802@veltech.edu.in

Abstract—-- **The maintenance of security in varied situations depends on intrusion detection. In this study, we assess three machine-learning systems' abilities to identify intrusions using data from optical sensors. The algorithms tested are Ridge Classifier, k-Nearest Neighbor (KNN), and a neural network. The system uses data collected from Optical Time Domain Reflectometer (OTDR) machines, which receive data from optical fiber sensors laid on the ground or walls/fences. The difference in the amplitude between the OTDR data traces that result from an intruder's movement disrupting the optical fiber signals is utilized to identify intrusions. The system preprocesses the data, and the three machine learning models are trained on the preprocessed data. Our study shows that ANN outperforms Ridge Classifier and the ANN in terms of accuracy, achieving 93% accuracy compared to Ridge Classifier's 92.5% and the neural network's 91%. These results indicate that KNN is a promising algorithm for intrusion detection using optical sensors.**

Index Terms— **Intrusion detection, Optical sensors, KNN, Ridge classifier, ANN**

I. INTRODUCTION

Intrusion detection is critical to security in various environments including military bases, industrial sites, and government buildings. Traditional intrusion detection systems (IDS) rely on physical sensors like motion detectors, acoustic sensors, and pressure sensors [2][3].

However, these systems have limitations, such as high false alarm rates, low accuracy, and vulnerability to tampering.

The assessment of intrusion detection systems is a difficult undertaking that calls for a full understanding of methodologies from various disciplines, including intrusion detection, attack tactics, networks, systems, technical testing, and evaluation. [1], [3]

In recent years, optical sensors have emerged as a promising technology for intrusion detection due to their sensitivity and reliability. Optical sensors work by detecting changes in the intensity of light caused by disturbances like vibrations or movement[5].

An optical sensor known as an optical Time Domain Reflectometer (OTDR)[3] may analyze the backscattered light from an optical fiber in order to passively identify intrusions. The OTDR generates a time-domain graph of the backscattered light, and any disturbance in the fiber optic cable caused by an intruder will be detected as a spike in the graph.

Optical sensor-based intrusion detection systems have seen their accuracy increase because of the introduction of machine learning techniques [6]. The OTDR data can be examined by machine learning techniques to find patterns that match to invasions. Several machine learning algorithms have been proposed for intrusion detection using optical sensors, including K-Nearest Neighbor (KNN), Ridge Classifier, and Neural Network.

II. LITERATURE SURVEY

The use of intrusion detection systems utilizing optical fiber has been gaining increasing attention in recent years due to their high sensitivity and accuracy in detecting intrusion events. Numerous research projects have been conducted to investigate the effectiveness of various machine learning algorithms in detecting intrusions in optical fiber-based systems.

Due to their great sensitivity and accuracy in detecting intrusions, optical fiber-based intrusion detection systems have seen an increase in popularity in recent years[7]. These systems utilize optical fibers as sensors and measure the changes in the optical signals caused by disturbances or vibrations, such as those triggered by an intruder. The location and type of the intrusion can be ascertained by detecting and analyzing the changes in the optical signal.

Several studies have been conducted to develop machine learning-based algorithms for intrusion detection using optical fiber sensors.

One approach is to use supervised learning algorithms such as K-nearest neighbors (KNN) to classify the signals and detect the presence of intrusions. For instance, in a study [8] intrusion detection is done using the KNN algorithm.

Utilizing deep learning methods like convolutional neural networks (CNNs) is another strategy for intrusion detection. CNNs have demonstrated promising results in identifying network anomalies and are widely employed in a variety of applications, including computer vision and natural language processing. For instance, in a study [9], a CNN-based algorithm was proposed for network anomaly detection and achieved good results.

Moreover, other machine learning algorithms such as the introduction of semi-supervised learning for the identification of road intrusion signals using optical fiber sensors [11].

In terms of the preprocessing of the optical signal data, several methods have been proposed [10]. For digital logic and signal regeneration, optical signal processing techniques are provided here.

The first model for intrusion detection was produced in 1987, following initial research in the field in 1980 [12]. Intrusion detection technology is still in its infancy and is hence ineffectual despite significant over the past few decades, there have been extensive research and commercial investments.

The commercial success of anomaly-based network IDS has not been as great as that of signature-based network IDS, which has been widely adopted by technology-based organizations worldwide.

There are still some significant issues that need to be solved since anomaly detection functionality-enabled security products are just starting to become more prevalent. This is true in spite of the wide array of anomaly-based network intrusion detection approaches that have lately been reported in the literature [13].

Some of the anomaly-based techniques that have been developed include Decision Tree, Gaussian Mixture Model, KNN, Support Vector Machines (SVM), Genetic Algorithm, and Linear Regression [14,15].

These days, artificial neural networks (ANN) are widely trained using the backpropagation method, which has been around since 1970 and is utilized as the reverse mode of automatic differentiation [16].

Overall, these investigations show how machine learning techniques have the potential to enhance the precision and efficiency of intrusion detection in optical fiber-based sensors. The comparison of performances of various algorithms such as KNN, Ridge Classifier, and neural networks can provide insights into which algorithm works best for a specific application.

Overall, the use of machine learning techniques for optical fiber sensor-based intrusion detection is a promising field of research, and additional research is required to create algorithms that are effective.

III. DATASET

The dataset used in this research paper consists of optical sensor data collected from an OTDR (Optical Time Domain Reflectometer) machine. The data is generated using a mathematical model of OTDR, and intrusion is introduced at random locations using sinusoidal noise.

The dataset includes 97 zones, with each zone containing approximately 41 meters of optical fiber data. The length of the cable used for testing is 4,000 meters. The main parameter used for detecting intrusion in the dataset is the difference in amplitude of the optical signal. The data is preprocessed and converted to CSV format for use in Python.

For the purpose of developing and assessing the machine-learning models, the dataset is divided into training and testing sets. The goal is to accurately detect intrusion and its location based on the changes in the optical signal caused by vibrations from the intruder. The dataset is used to assess the performance of various machine learning methods, including KNN, Ridge Classifier, and Neural Network.

IV. SYSTEM ARCHITECTURE

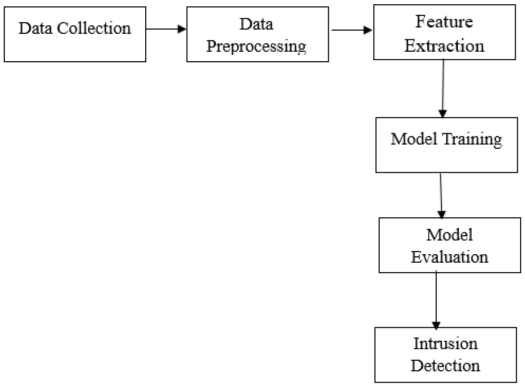


Figure 1. Block Diagram of the proposed system

The above architecture is designed to provide a real-time intrusion detection system that can monitor large areas and quickly detect and locate any intrusion. The system can be integrated with existing security infrastructure to provide an added layer of protection.

V. DATAPREPROCSSNG

The optical time domain reflectometer (OTDR) mathematical model was used to generate data as part of the data preparation for the optical fiber intrusion detection system. Sinusoidal noise was used to introduce intrusions at random points. To improve data extraction and detection, a 4000-meter wire length was cut into 97 zones, each measuring about 41 meters, for testing reasons. The difference in amplitude was the primary parameter utilized to provide data into the machine learning model. To utilize Python, all data was first created in MATLAB and then converted to CSV format.

A deep learning or machine learning model was applied to data traces, and the model was trained to utilize the constructed loss as a parameter. By running traces through the trained autoencoder model, a mean signal was produced. After that, the mean signal was employed as a dynamic average, and an envelope representing the typical range of data was made around it using the constructional loss standard deviation. The data was regarded as an intrusion if it went beyond this bond.

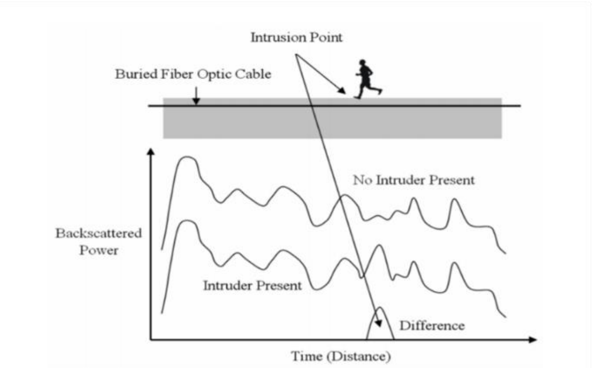


Figure 2. Intrusion detection system

Figure 2 shows how the intrusion is detected based on the difference in the amplitude of the OTDR-generated data.

VI. MODEL ALGORITHM

Here we have used 3 machine-learning algorithms which are ANN, Ridge classifier, and KNN.

A. ANN

The artificial neural network is known as ANN. It is a kind of artificial intelligence algorithm that draws inspiration from the design and operation of biological neural networks. Layered networks of interconnected nodes or neurons make up ANNs.

The network's predictions are generated by the output layer once the input layer has received the data. One or more hidden layers that convert the input into a format that the output layer may use may be present in between.

ANNs are employed in a variety of disciplines, including prediction modeling, speech and picture recognition, natural language processing, and others. They are potent tools for resolving difficult issues due to their capacity for learning from big datasets and spotting intricate patterns.

To obtain great accuracy, ANNs need a lot of training data, which can be computationally expensive.

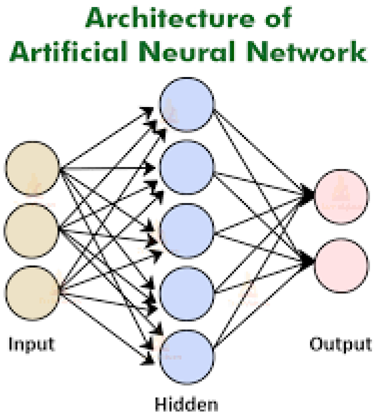


Figure 3. ANN Architecture

Figure 3 shows the architecture of ANN which consists of one input layer, one hidden layer, and one output layer.

B. RDGE CLASSIFIER

The Ridge Classifier is a linear model for binary classification. In order to prevent overfitting, a penalty term is added to the loss function of the basic linear model (such as logistic regression). When working with datasets where the number of features is high relative to the number of samples, the Ridge Classifier is quite helpful.

Each feature in the input is given a coefficient-weighted linear combination, which is then computed by the Ridge Classifier. By minimizing a loss function that has a regularisation term, the coefficients are obtained. Overfitting is prevented by this regularisation term's penalization of the coefficient sizes.

In practice, the Ridge Classifier is often used in conjunction with cross-validation to tune the regularization parameter. This parameter controls the strength of the penalty term, with larger values leading to more regularization.

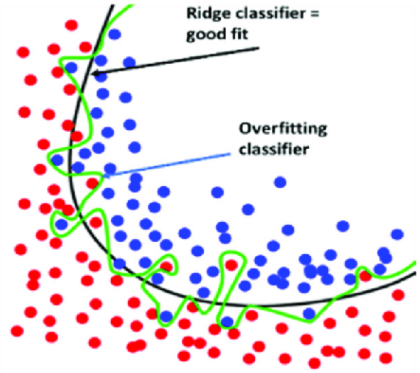


Figure 4. Ridge classifier Architecture

Figure 4 shows the architecture of the Ridge classifier

C. KNN

Non-parametric machine learning algorithms like K-Nearest Neighbours (KNN) are utilized for both classification and regression. The KNN technique uses the majority class of a new data point's K nearest neighbors in the feature space to determine what class it belongs to when classifying data. A hyperparameter called K must have a value before the algorithm can be used on the data.

A higher value of K can aid in reducing noise and enhancing the model's capacity to generalize, but it may also cause over-smoothing and the loss of crucial data points. A lower value of K, on the other hand, can result in overfitting and worse performance on new data. KNN is an easy-to-understand method that performs well in low-dimensional feature spaces but may struggle with high-dimensional data or classes that are unbalanced.

Additionally, KNN can be computationally expensive, particularly as the dataset and dimension sizes grow.

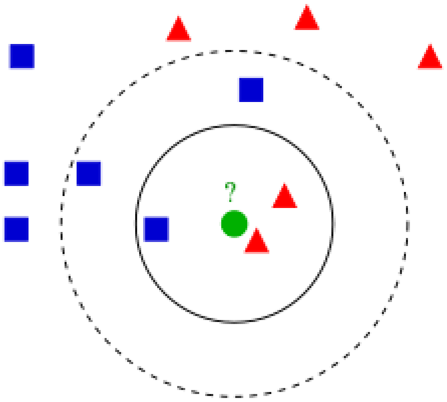


Figure 5. KNN Architecture

Figure 5 shows the architecture of KNN

The above-mentioned three machine-learning algorithms are used in our analysis and the accuracies are compared.

VII. METHODOLOGY

Here are the steps which we have taken for the analysis using the machine learning algorithms.

1. Data collection: In this study, optical fiber data is collected from the OTDR machine. The data is generated with a mathematical model of OTDR and intrusion is introduced at a random location using sinusoidal noise.

2. Data preprocessing: The collected data is preprocessed to extract useful features and remove noise. For this purpose, a deep learning model, autoencoder, is used. The model is trained to remove anomaly present in the signal while encoding and after decoding the traces, the output trace acts as the mean signal. The difference in amplitude is used as the main parameter for feeding in the model and detection.

3. Model selection: In this study, three machine learning algorithms are used for intrusion detection, namely K-Nearest Neighbor (KNN), Ridge Classifier, and Artificial Neural Network (ANN). The performance of each model is evaluated and the best model is selected based on accuracy, precision, and recall.

4. Model training: After selecting the best model, it is trained on the preprocessed data. The dataset is divided into training and testing sets. The model is trained on the training set and evaluated on the testing set. The training process involves adjusting the model parameters to minimize the loss function.

5. Model evaluation: The testing set is used to evaluate the performance of the trained model by measuring its accuracy, precision, recall, and F1 score. The confusion matrix is employed to compute these performance metrics.

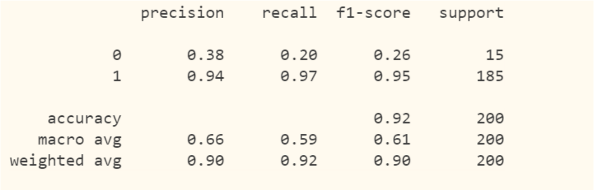
6. Result analysis: The results obtained from the selected model are analyzed and compared with other models to validate the effectiveness of the proposed approach.

VIII. RESULTS AND DSCUSSON

The performance of the proposed optical sensor-based intrusion detection system using machine learning algorithms was evaluated and compared using three different classifiers, namely Artificial Neural Network (ANN), Ridge Classifier, and K-Nearest Neighbor (k-NN). The results were analyzed based on the metrics of accuracy, precision, recall, and F1-score.

A. ANN

Here firstly ANN algorithm is used for training and testing the data and here are the results that we have achieved.



```
              precision    recall  f1-score   support

           0       0.38      0.20      0.26        15
           1       0.94      0.97      0.95       185

    accuracy                           0.92       200
   macro avg       0.66      0.59      0.61       200
weighted avg       0.90      0.92      0.90       200
```

Figure 6. Classification report of ANN

Here Figure 6 represents the classification report of the ANN algorithm where there are two cases and the values of precision, recall, f1-score, and support are mentioned.

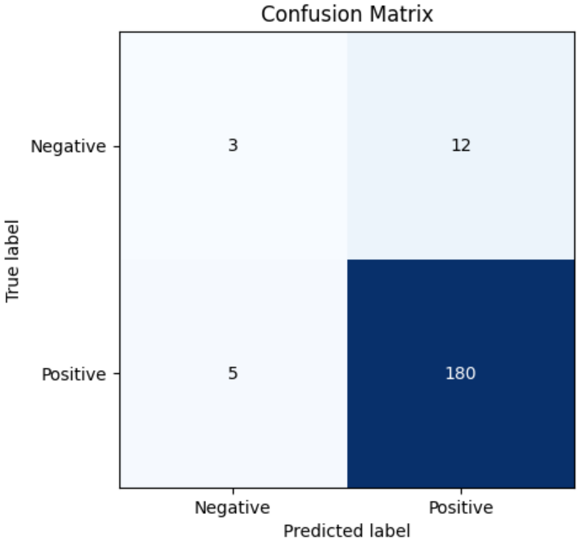


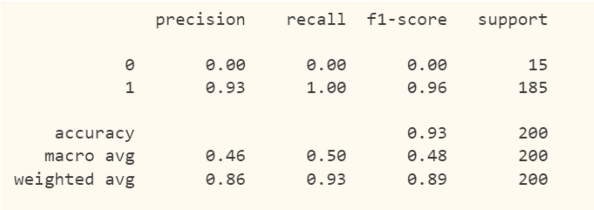
Figure 7. Confusion matrix of ANN

Here Figure 7 represents the confusion matrix of the ANN algorithm where there are 3 true negatives, 12 false positives, 5 false negatives, and 180 true positives values.

The ANN model's accuracy is calculated using the confusion matrix mentioned above, and the result is 91.5%.

B. *RIDGE CLASSIFIER*

Here, we utilized the Ridge Classifier algorithm for both training and testing the data, and the outcomes are shown below.



```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        15
           1       0.93      1.00      0.96       185

    accuracy                           0.93       200
   macro avg       0.46      0.50      0.48       200
weighted avg       0.86      0.93      0.89       200
```

Figure 8. Classification Report of Ridge Classifier

Here Figure 8 represents the classification report of the ANN algorithm where there are two cases and the values of precision, recall, f1-score, and support are mentioned.

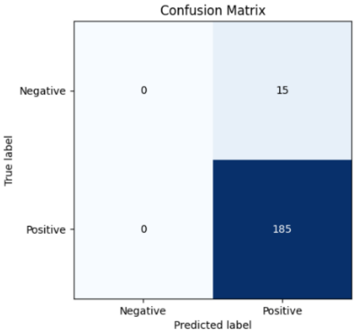


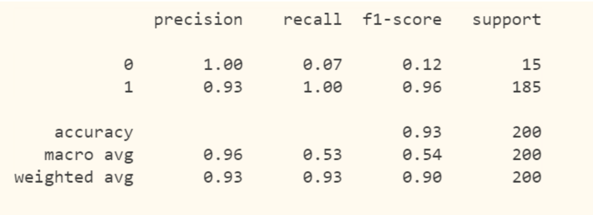
Figure 9. Confusion matrix of Ridge Classifier

Here Figure 9 represents the confusion matrix of the Ridge Classifier algorithm where there are 0 true negatives, 15 false positives, 0 false negatives, and 185 true positives values.

The Ridge Classifier model's accuracy is calculated using the confusion matrix mentioned above, and the result is 92.5%.

C. *KNN*

Lastly, we used the KNN algorithm is used for training and testing the data, and here are the results that we have achieved.



```
              precision    recall  f1-score   support

           0       1.00      0.07      0.12        15
           1       0.93      1.00      0.96       185

    accuracy                           0.93       200
   macro avg       0.96      0.53      0.54       200
weighted avg       0.93      0.93      0.90       200
```

Figure 10. Classification report of KNN

Here Figure 10 represents the classification report of the KNN algorithm where there are two cases and the values of precision, recall, f1-score, and support are mentioned.

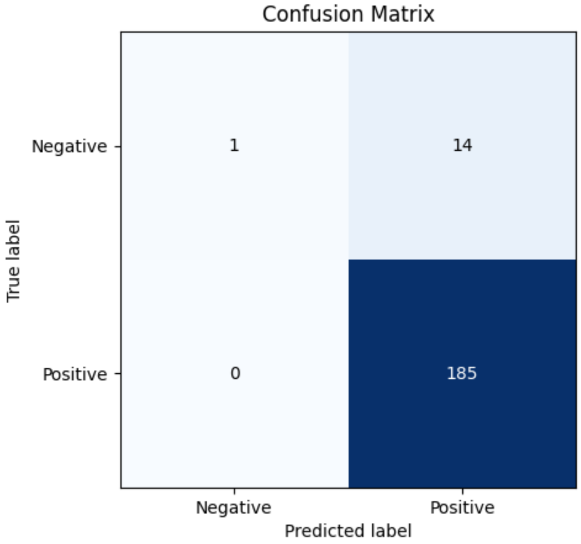


Figure 11. Confusion matrix of KNN

Here Figure 11 represents the confusion matrix of the KNN algorithm where there are 1 true negative, 14 false positives, 0 false negatives, and 185 true positives values.

The KNN model's accuracy is calculated using the confusion matrix mentioned above, and the result is 93%.

MACHINE LEARNING ALGORITHM	ACCURACY
ANN	91.5%
Ridge Classifier	92.5%
KNN	93.0%

Table 1. Comparison of Accuracies

The above table represents the comparison of the machine-learning model accuracies.

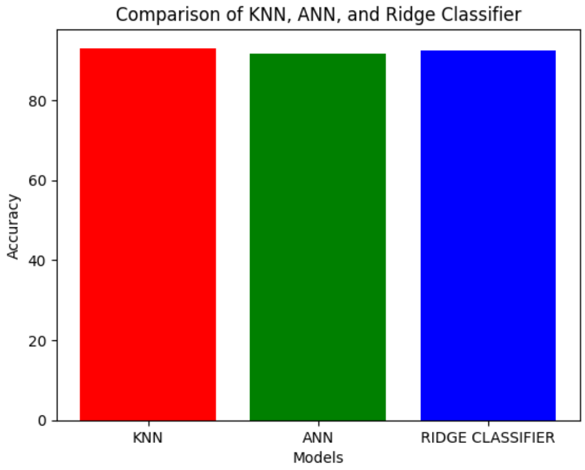


Figure 12. Comparison of all three models

Figure 12 shows the comparative accuracies of all three machine-learning models which shows clearly that the performance of KNN is better than the other two algorithms.

DISCUSSION

The results of this study indicate that machine learning algorithms, specifically KNN, ANN, and Ridge Classifier, can effectively detect intrusions using optical sensor data The KNN algorithm had the best accuracy (93%), which was followed by the Ridge Classifier and ANN, which had accuracies of 92.5% and 91.5%, respectively.

This result is in line with earlier research that demonstrated the efficacy of machine learning algorithms for intrusion detection employing a range of sensors, including optical sensors. For instance, a study by [17] used ANN machine learning algorithms to check the performance of the dataset, achieving an accuracy of 81.2%. Similarly, a study by [18] used a comparison study of intrusion databases based on SOM, achieving an accuracy of 77.23%.

The superior performance of KNN in this study is likely due to its ability of robustness to noisy data and outliers which does not assume any underlying distribution of data which is particularly useful for high-dimensional datasets like the optical sensor data used in this study. And also it is less prone to overfitting than the other two algorithms.

This is because KNN uses the surrounding data points to make decisions, which means that it doesn't get biased by a small subset of the data.

However, it is worth noting that the Ridge Classifier and ANN algorithms also performed well in this study, achieving accuracies of 92.5% and 91.5%, respectively. These algorithms have been widely used in intrusion detection systems and have been shown to perform well for various types of data [17][18].

Overall, this study's findings indicate that optical sensor data can be used to detect intrusions using machine learning algorithms, particularly KNN. Future research could explore the use of other machine learning algorithms or combinations of algorithms to further improve the accuracy of intrusion detection systems.

IX. CONCLUSION

In this study, we have presented a simulation analysis of optical sensor-based intrusion detection using machine learning algorithms. We have collected optical fiber data from an OTDR machine and fed it to different machine learning models, including ANN, Ridge Classifier, and KNN. The results show that KNN achieves the highest accuracy of 93%, followed by Ridge Classifier and ANN with 92.5% and 91.5% accuracy, respectively. The KNN model is able to detect intrusion with high accuracy and locate it accurately, which makes it a promising candidate for practical intrusion detection systems.

Our study contributes to the research on optical sensor-based intrusion detection, which has important applications in various domains, such as perimeter security for critical infrastructure, military bases, and border control. The effectiveness and efficiency of intrusion detection systems can be increased, and the incidence of false alarms can be decreased, through the use of machine learning algorithms.

Future studies will need to address a few remaining issues and problems, though. For instance, the complexity and variety of the environment, such as the weather, vegetation, and terrain, may have an impact on the performance of the machine-learning models. The models' performance may also be impacted by the availability and caliber of the optical fiber data.

In conclusion, our study demonstrates the feasibility and effectiveness of using machine learning algorithms for optical sensor-based intrusion detection and provides useful insights for designing and optimizing practical intrusion detection systems.

X. FUTURE SCOPE

One of the key areas for future work in this research is to improve the accuracy and robustness of the intrusion detection system by using a larger dataset. The current study used a limited dataset due to the constraints of acquiring and preprocessing OTDR data. Expanding the dataset by collecting data from multiple fibers, different types of fibers, and different levels of signal-to-noise ratio would increase the generalizability of the proposed system. Another direction for future work could be to explore the use of deep learning techniques, such as CNN, to automatically learn relevant features from the OTDR data.

Furthermore, the proposed system can be extended to detect intrusions at multiple points along the fiber by using a distributed sensing technique. Lastly, the proposed system can be integrated with an automated alerting mechanism to notify the network administrator in real time when an intrusion is detected.

REFERENCES

[1] Khorkov, D. A. ”Methods for testing network-intrusion detection systems”. Scientific and Technical Information Processing, 2012, vol.39, no2, p. 120-126. DOI=10.3103/S014768821202012

[2] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, “Modeling and implementation approach to evaluate the intrusion detection system,” in International Conference on Networked Systems, 2015, pp. 513–517.

[3] R. Berthier, W. H. Sanders, and H. Khurana, ”Intrusion Detection for Advanced Metering Infrastructures: Requirements and Architectural Directions,” 2010 First IEEE International Conference on Smart Grid Communications, Gaithersburg, MD, 2010, pp. 350-355. doi:10.1109/SMART-GRID.2010.5622068

[4] Shi, Y.; Wang, Y.; Zhao, L.; Fan, Z. An Event Recognition Method for Φ-OTDR Sensing System Based on Deep Learning. *Sensors* **2019**, *19*, 3421. https://doi.org/10.3390/s19153421

[5] Sabri, Naseer & Aljunid, S. & Salim, Muhammed & Ahmad, R.Badlishah & Kamaruddin, Rosliha. (2013). Toward Optical Sensors: Review and Applications. Journal of Physics Conference Series. 423. 2064-. 10.1088/1742-6596/423/1/012064.

[6] Liu, H.; Lang, B. Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. *Appl. Sci.* **2019**, *9*, 4396. https://doi.org/10.3390/app9204396

[7] Allwood, G. & Wild, Graham & Hinckley, Steven. (2016). Optical Fiber Sensors in Physical Intrusion Detection Systems: A Review. IEEE Sensors Journal. 16. 1-1. 10.1109/JSEN.2016.2535465.

[8] Liu, G.; Zhao, H.; Fan, F.; Liu, G.; Xu, Q.; Nazir, S. An Enhanced Intrusion Detection Model Based on Improved kNN in WSNs. Sensors 2022, 22, 1407. https://doi.org/10.3390/s22041407

[9] Wang, Y.-C.; Houng, Y.-C.; Chen, H.-X.; Tseng, S.-M. Network Anomaly Intrusion Detection Based on Deep Learning Approach. Sensors 2023, 23, 2171. https://doi.org/10.3390/s23042171

[10] Ennser, K. et al. (2009). Optical Signal Processing Techniques for Signal Regeneration and Digital Logic. In: Tomkos, I., Spyropoulou, M., Ennser, K., Köhn, M., Mikac, B. (eds) Towards Digital Optical Networks. Lecture Notes in Computer Science, vol 5412. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-01524-3_4

[11] Jun He, Xing Hu, Dawei Zhang, Yong Kong, Jing Cheng, and Wenzhe Xiao, "Semi-supervised learning for optical fiber sensor road intrusion signal detection," Appl. Opt. 61, C65-C72 (2022)

[12] N. Chakraborty, “Intrusion detection system and intrusion prevention system: A comparative study,” International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.

[13] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, “Anomaly-based network intrusion detection: Techniques, systems and challenges,” computers & security, vol. 28, no. 1–2, pp. 18–28, 2009.

[14] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, “Modeling and implementation approach to evaluate the intrusion detection system,” in International Conference on Networked Systems, 2015, pp. 513–517.

[15] A. S. Ashoor and S. Gore, “Importance of intrusion detection system (IDS),” International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[16] J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural networks, vol. 61, pp. 85–117, 2015.

[17] B. Ingre and A. Yadav, “Performance analysis of NSL-KDD dataset using ANN,” in Signal Processing And Communication Engineering Systems (SPACES), 2015 International Conference on, 2015, pp. 92–96.

[18] L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod, “A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self organization map (SOM) artificial neural network,” Journal of Engineering Science and Technology, vol. 8, no. 1, pp. 107–119, 2013.