Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

TNDRM: Data Normalization Using Uncovering Complex Data Structures With Topological Nonlinear Dimensionality Reduction Using Manifold Learning

[1]Mrs.K.R.Prabha, [2[Dr. B. Srinivasan

[1] Research Scholar, Department of Computer Science, Gobi Arts & Science College, Gobichettipalyam, Tamilnadu, India
[2]Associate Professor, Department of Computer Science, Gobi Arts & Science College, Gobichettipalyam, Tamilnadu, India

Abstract: With the increasing complexity of Distributed Denial of Service (DDOS) attacks in Wireless Sensor Networks (WSNs), the accurate detection of these threats has become imperative. This research presents a robust preprocessing technique for DDOS attack detection, focusing on data normalization through the integration of topological nonlinear dimensionality reduction via manifold learning (TNDRM). Our methodology revolves around transforming the intricate high-dimensional feature space of WSN data into a lower-dimensional representation, all while preserving the intrinsic topology and geometry of the original data. Achieved through manifold learning techniques, this process enables a more meaningful understanding of complex data structures, essential for effective analysis. A pivotal step in our approach involves the normalization of the data within the reduced dimensional space. Leveraging statistical techniques, specifically z-score normalization and min-max scaling, we mitigate the impact of varying scales and outliers in the data. The accuracy of machine learning algorithms is improved with normalization because it guarantees uniformity and consistency by removing aberrations.
Keywords: Distributed Denial of Service, Intrusion Detection Systems, Preprocessing, wireless sensor network

1. Introduction

In today's Wireless Sensor Networks (WSNs) scenario, protecting the security and integrity of data transmission is critical [1]. With the proliferation of sophisticated cyber threats, notably Distributed Denial of Service (DDOS) attacks, protecting these networks from malicious infiltration has become an ongoing issue [2-3]. DDOS attacks, which are distinguished by their capacity to overload network resources and interrupt services, demand sophisticated detection techniques capable of navigating the complicated patterns within network traffic data [4-6]. When dealing with the large dimensionality and complicated structures inherent in network data, traditional techniques of DDOS detection in WSNs often confront problems [7-8]. Thus, there is a pressing want for cutting-edge methods that may reduce the dimensionality of this data while yet retaining its essential topology and geometry [9]. Furthermore, successful identification necessitates resolving the data's changing scales and outliers, ensuring that machine learning algorithms can function on a consistent and trustworthy input space [10-12].

This study tackles these issues by presenting a unique method for combining the strength of topological nonlinear dimensionality reduction via manifold learning with rigorous data normalization procedures [13-14]. We want to identify the complex and hidden structures inside the high-dime4421

Bibliography

[1] P. J. Lucas, J. Baird, L. Arai, C. Law, and H. M. Roberts, “Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews,” *BMC Medical Research Methodology*, vol. 7, no. 1, p. 4, Jan. 2007, doi: 10.1186/1471-2288-7-4.

[2] R. G. Vajrapu and S. Kothwar, “Software Requirements Prioritization Practices in Software Start-ups : A Qualitative research based on Start-ups in India,” *undefined*, 2018, Accessed: Dec. 09, 2021. [Online]. Available: https://www.semanticscholar.org/paper/Software-Requirements-Prioritization-Practices-in-%3A-Vajrapu-Kothwar/5605bd2a2dc93a0a997a7c68de30969b01312e5f

{3] M. Gutbrod and J. Münch, "Teaching lean startup principles:an empirical study on assumption prioritization," in *Software-intensive business: start-ups, ecosystems and platforms : proceedings of the International Workshop on Software-intensive Business: Start-ups, Ecosystems and Platforms (SiBW 2018) : Espoo, Finland, December 3, 2018. - (CEUR workshop proceedings ; 2305)*, 2018, pp. 245–253. [Online]. Available: http://ceur-ws.org/Vol-2305/ nsional feature space of WSN data by applying manifold learning. While preserving the data's essential characteristics, we transform it into a lower-dimensional space [15, 16]. Following that, the modified data is subjected to a thorough normalization procedure that employs statistical approaches such as z-score normalization and min-max scaling [17-18]. This phase is critical for removing the influence of various scales and outliers and guaranteeing consistent and uniform input for machine learning algorithms [19-20]. The fundamental goal of this study is to improve the detection of DDOS attacks in WSNs by establishing a robust preprocessing framework [21-22]. Our method not only advances anomaly detection techniques, but it also lays the door for enhanced network security tactics in the context of wireless sensor networks [23-24].

The primary contributions and objectives of this manuscript may be summarized as follows.

- Statistical Normalization using z-score normalization and min-max scaling
- Dataset preprocessing using topological nonlinear dimensionality reduction using manifold learning

From here on out, the structure of this article will look like this. In Section 2, multiple authors address various methods of data normalization. In Section 3, we show the suggested model. We provide our research results in Section 4. Our results are summed up and recommendations for further study are offered in the last section.

1.1 Motivation of the paper

Because the complexity of DDOS attacks in WSN is increasing, accurate detection techniques are required. Topological nonlinear dimensionality reduction via manifold learning is combined with data normalization approaches in this study. Our technique enables precise DDOS attack detection by retaining key data structures. The objective is to improve WSN security by resolving extensive data nuances, maintaining consistency via normalization, and exceeding conventional approaches' constraints. The validation of the research on the WS-NDS dataset demonstrates its superiority, providing a substantial development in Intrusion Detection Systems (IDS) for WSNs.

2. Background Study

Abidoye, A. P., & Obagbuwa, I. C. [1] Because of their versatility, WSNs have been gaining in popularity in recent years. The communication styles and methods of deploying sensor networks make them vulnerable to many different kinds of attacks. Using the pre-shared keys, the MAS encrypts the message before sending it to the recipient node, ensuring the authenticity and integrity of the data being sent over the network. For data packet authentication, MAS uses a nonce and a hash value. The MAS has been shown to be able to identify and fight against DDoS attacks in WSNs in simulations. To determine whether the suggested method satisfies the resource limitations of WSNs, it will be implemented in a real test bed in a subsequent research.

Edlund, A. et al. [5] The sequencing of 16S rRNA genes and other metagenomic methods have provided the bulk of these authors present understanding of the intricate human microbiome by revealing both the presence and absence of microorganisms and the relative abundance of genes. In this work, the author used a metatranscriptomic technique to learn more about how bacteria behave and how an oral biofilm community develops and grows by analyzing the mRNA produced by active bacterial genes and genomes.

Guo, W., & Banerjee, A. G. [9] the author apply a powerful TDA instrument, the Mapper algo-rithm, to estimate yield and find errors in data sets from the chemical manufacturing process and the semiconductor etch process. The author demonstrate that the Mapper algorithm provides additional insights into the complicated data beyond the standard methods of feature selection. To better comprehend the tangential connections between the features and manufacturing system outputs, the author use direct visualisation to build an abstract representation of the data.

Islam, M. T., & Xing, L. [11] For this reason, the author suggest a three-stage technique, GSE, for nonlinear dimensionality reduction, with the first stage being concerned with maintaining the scalability and integrity of the data geometry. To correct the first short-circuiting issue, the second stage integrates neighborhood-specific data. The third phase of GSE keeps the data's statistics intact. Multiple instances have shown that GSE performs better than other dimensionality reduction methods at simultaneously retaining the data's geometric and statistical features while shrinking their size.

Puschmann, D. et al. [17] the author provides a new method for discovering patterns and connections across disparate datasets. By taking an unsupervised approach, these authors solution was able to examine multiple streams of sensor data, extract patterns therein, and then translate those patterns into higher-level abstractions that can be understood by both humans and machines. Any kind of Internet of Things (IoT) data stream may benefit from the suggested approach.

Ramesh, S. et al. [19] the aforementioned deep neural network-based DoS detection approach has been shown to be effective in all tests. The suggested method's detection accuracy and low time investment result from the use of a neural network with several layers of neurons. The experimental findings were predicated on the idea that optimization methods may boost learning efficiency. In addition, it has been shown that feature selection decreases the dataset's dimensionality.

2.1 Problem definition

The increasing complexity of DDOS attacks on WSNs offers a significant challenge to network security. Due to the complicated and high-dimensional structure of WSN data, traditional approaches fail to identify these sophisticated threats correctly. The problem is in converting this complicated data into a format that can be effectively analyzed. The data's varied sizes and outliers impede the efficiency of machine learning algorithms, lowering the accuracy of DDOS attack detection. In order to address these difficulties, we are developing a robust preprocessing approach that combines topological nonlinear dimensionality reduction via manifold learning and data normalization.

3. Materials and methods

In this part, we describe the materials utilized and the methods employed in our study to create a robust preprocessing mechanism to improve the detection accuracy of DDOS attacks in WSNs. Our method combines topological nonlinear dimensionality reduction with rigorous data normalization approaches. The data normalization using uncovering complex data structures with topological nonlinear dimensionality reduction using manifold learning model flowchart has represented at figure 1.

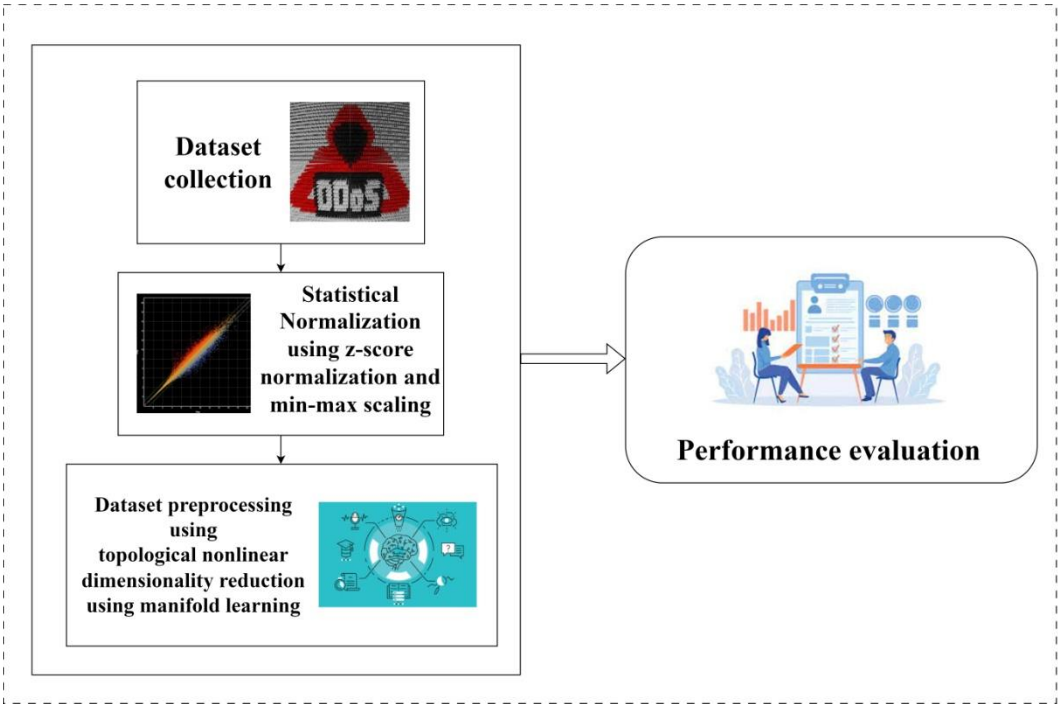

Fig 1: TNDRM Flow chart

3.1 Dataset collection

https://www.kaggle.com/datasets/bassamkasasbeh1/wsnds, The WS-NDS dataset, available on Kaggle, is a collection of network traffic data for intrusion detection system (IDS) evaluation. The dataset contains a total of 56,422 network traffic instances, which are labeled as either normal or attack traffic.

3.2 Statistical Normalization using z-score normalization and min-max scaling

3.2.1 Z-score normalization

Z-score normalization, also known as standardization, is a common statistical procedure for cleaning and organizing data. From the above information, a normal distribution is derived, with mean 0 and standard deviation 1. This technique is very helpful for standardizing a dataset containing characteristics that utilize a variety of units and scales. To calculate a z-score, we first remove the dataset's mean from each data point and then divide that number by its standard deviation. The mathematical formula for determining the z-score (Z) for a set of features XX is as follows:

$Z = \frac{(X-\mu)}{\sigma}$ ------ (1)

- X is the feature's starting point value.
- μ represents an average of all feature values.
- σ represents the dispersion of feature values as a whole.

In the preprocessing phase, normalization decomposes data with numerical properties so that the values in the data may be transformed into a specified range. Min-max normalization, z-score normalization, and decimal scaling are the most prevalent approaches to normalizing data. Z-score normalization, as shown by Equation 1, assigns an attribute E value to a new range.

$v' = \frac{v_i - E_i}{std(E)}$ ------- (2)

Description:

v' = value obtained after normalization.

v_i = the property value that has to be normalized

E_i = mean attribute value

$std(E)$ = the E-attribute of the standard deviation.

3.2.2. Min-max scaling

To normalize characteristics to a specified range, often [0, 1], min-max scaling is a preprocessing method used in data analysis. When working with data that fluctuates greatly in size, it shines. This technique uses a linear transformation to scale the data such that the feature's lowest and maximum values are represented by 0 and 1, respectively. For a given feature XX, the min-max scaling formula is:

$X_{scaled} = \frac{X - x_{min}}{x_{max} - x_{min}}$ ------ (3)

Where:

- X is the original worth of the component.
- x_{min} is the lowest value of this attribute that appears in the data set.
- x_{max} is the highest possible value of the characteristic found in the data set.

3.3 Dataset preprocessing using topological nonlinear dimensionality reduction using manifold learning

During the preprocessing phase, our method uses topological nonlinear dimensionality reduction techniques like t-Distributed Stochastic Neighbor Embedding (t-SNE) and Isometric Mapping (Isomap) to transform raw network traffic data from Wireless Sensor Networks (WSNs) into a lower-dimensional space while preserving the underlying topology and geometry of the original data. These techniques, which are especially useful in high-dimensional environments, provide a more accurate representation of the data essential for in-depth analysis by capturing intricate patterns and nonlinear interactions. In addition to preserving essential topological and geometric information, this converted, lower-dimensional data makes it possible for future machine learning algorithms to work on a more relevant and intelligible input space. The effectiveness of DDOS attack detection models in WSNs may be greatly improved with the help of this preprocessing phase, which ensures that the detection system can accurately identify even the most subtle patterns and abnormalities in the network data.

$V = R^D = \left\{ x = \begin{pmatrix} x_1 \\ \vdots \\ x_D \end{pmatrix} : x_j \in R \right\}$ ------ (4)

Supposing V is a vector space, we may define W^T as a subspace of V with a finite basis.

$W^T = \{x \in R^D : < x, w >= 0 \forall_w \in W\}$ --------- (5)

The RD subspace whose orthogonal counterpart is RD is denoted as W^T. W is a subspace of RD iff and only iff RD, hence W is a subspace of RD.

$dim(W^T) = D - dim(w)$ ------- (6)

From the example, it's easy to see that picking neighbors based on distance alone would result in x3 being selected as one of x0's neighbors. First, we'll pretend that the two vectors that connect x0 to its immediate neighbors both independently go via W^T. The W may then be calculated using the vectors that form its basis. The angle between W^T and x3 is less than 90 degrees if x3 is not on the same surface as x1 and x2.

Incorrectly establishing neighboring relationships is a major cause of topological instability. The suggested approach is utilized to create the neighborhood link between the data points and then build the weighted graph G over the data. Our manifold reconstruction is significantly easier to compute than Freedman's, which requires an expensive optimization for convex hulls.

When working with an unstructured data set as a starting point, the reconstruction difficulty boils down to reestablishing local edge connections. The set of K points in M that surround a single point p is referred to as its neighborhood, abbreviated $NBD(p)$.

$EP(p) = \{q \in NBD(p) | < p - r, q - r > \geq 0, any\ r \in NBD\ (p)\}$ ------- (7)

Acute or right angles may be shown to exist between any two adjacent edges, but obtuse angles cannot. This characteristic enables the production of well-shaped simplices, which are fundamental for erecting the desired simplicial complex. It's often held that (b)'s 1D reconstruction is superior than (c)'s 2D version.

Algorithm 1: TNDRM
Input: X: the input data matrix with shape (n,d), where n is the number of data points, and d is the number of features. $n_{components}$: count of low-dimensional spatial dimensions Perplexity: The degree to which local and global structure are maintained may be adjusted by adjusting this hyperparameter. **Steps:** **Compute pairwise similarities (similarity matrix)** Gaussian kernel distances are used to calculate pairwise similarities between data points in the high-dimensional space. Construct a similarity matrix P where P_{ij} represents the conditional probability that point *I* would pick point *j* as its neighbor if neighbors were picked proportionally to their similarity. **Compute perplexity-Adjusted probabilities** Adjust the conditional probability in the similarity matrix P to achieve the target perplexity. Perplexity is a measure of the effective number of neighbors for each data points. The algorithm aims to match the desired perplexity value specified as input. **Initialize Low-Dimensional embedding** Initialize the low-dimensional representation Y randomly or with PCA (Principal component analysis) **Define similarity distributions in low-dimensional Space** Compute pairwise similarities in the low-dimensional space using the students t-distribution with one degree of freedom (Cauchy distribution). Construct a similarity matrix Q for the low-dimensional representation. **Minimize the kullback-leibler divergence**

Reduce the p-Q Kullback-Leibler divergence by shifting data points around in the low-dimensional space.
Gradient descent
Optimize data-point locations in the low-dimensional space by gradient descent.
Update the positions iteratively until convergence
Output low-dimensional representation
Return the low-dimensional representation Y.
Output
Y:the low-dimensional representation of the input data with shape (n,$n_{components}$)

4. Results and discussion

In this part, we provide the results of our studies and talk in length about what these results mean. Our findings demonstrate that combining topological nonlinear dimensionality reduction via manifold learning with meticulous data normalization techniques can significantly enhance the precision with which Wireless Sensor Networks (WSNs) can detect Distributed Denial of Service (DDOS) attacks.

Table 1: Data frame

	Non-Null count	Data type
Id	374661	Int64
Time	374661	Int64
Is_CH	374661	Int64
Who CH	374661	Int64
DIST_TO_CH	374661	Float64
ADV_S	374661	Int64
ADV_R	374661	Int64
JOIN_S	374661	Int64
JOIN_R	374661	Int64
SCH_S	374661	Int64
SCH_R	374661	Int64
Rank	374661	Int64
DATA_S	374661	Int64
DATA_R	374661	Int64
DATA_SENT_TO_BS	374661	Int64
DIST_CH_TO_BS	374661	Float64
SEND_CODE	374661	Int64
Expaned Energy	374661	Float64
Attack type	374661	Object

The table 1 shows the dataset comprises 374,661 entries with multiple attributes. 'Id' represents a unique identifier, 'Time' indicates a timestamp, and 'Is_CH' and 'Who CH' are binary indicators denoting specific network configurations. 'Dist_to_CH' is a numerical value capturing distances. Additionally, the dataset includes various counts related to network activities: 'ADV_S' and 'ADV_R' denote advertisement activities, while 'JOIN_S' and 'JOIN_R' represent joining activities. 'SCH_S' signifies scheduling activities. These attributes, ranging from binary flags to numerical distances and activity counts, provide a diverse dataset for analysis. The 'Id' column serves as a primary identifier, 'Time' offers temporal context, and the other attributes indicate network states and activities, forming a comprehensive dataset for detailed exploration and analysis in the context of Wireless Sensor Networks (WSNs).

Table 2: Statistical calculation

	Mean	**STD**	**Min**
Id	274969.325879	389898.554898	101000.0
Time	1064.748712	899.646164	50.0
Is_CH	0.115766	0.319945	0.0
Who CH	274980.411108	389911.221734	101000.0
Dist_to_CH	22.599380	21.955794	0.0
ADV_S	0.267698	2.061148	0.0
ADV_R	6.940562	7.044319	0.0
JOIN_S	0.779905	0.414311	0.0
JOIN_R	0.737493	4.691498	0.0
SCH_S	0.288984	2.754746	0.0
SCH_R	0.747452	0.434475	0.0
Rank	9.687104	14.681901	0.0
DATA_S	44.857925	42.574464	0.0
DATA_R	73.890045	230.246335	0.0
Data_Sent_To_BS	4.569448	19.679195	0.0
Dist_CH_To_BS	22.562735	50.261604	0.0
Send_Code	2.497957	2.407337	0.0
Expaned Energy	0.305661	0.669462	0.0

The table 2 provided statistics offer valuable insights into the dataset attributes. The 'Mean' values provide an average measure across the dataset: notably, the average 'Id' and 'Who CH' values indicate mid-range identifiers, while 'Time' averages around 1065, suggesting a moderate timestamp value. Binary features such as 'Is_CH' demonstrate a low average, around 0.12, indicating infrequent occurrences. 'Dist_to_CH' averages at approximately 22.6, suggesting a moderate distance metric. Activity counts like 'ADV_S,' 'ADV_R,' 'JOIN_S,' 'JOIN_R,' and 'SCH_S' have relatively low averages, highlighting limited activities in general. 'DATA_S,' 'DATA_R,' and 'Data_Sent_To_BS' indicate moderate data transmission values, with 'DATA_R' displaying a notably higher mean due to potential outliers. Distance-related features like 'Dist_CH_To_BS' average around 22.56, indicating a moderate distance between nodes and base stations. 'Send_Code' suggests an average value of 2.5, pointing to moderate coding activities. 'Rank' displays an average of approximately 9.7, reflecting a moderate rank level. Energy-related feature 'Expanded Energy' averages at 0.31, indicating relatively low energy expansion. The 'STD' values highlight the data dispersion around the mean, and the 'Min' values indicate the minimum observed values for each attribute. These statistics provide a comprehensive overview of the dataset, aiding in understanding the distribution and variability of the features in the context of Wireless Sensor Networks analysis.

Table 3: Unique data

	Data
Id	11120
Time	196
Is_CH	2
Who CH	7088
Dist_to_CH	13956
ADV_S	85
ADV_R	31
JOIN_S	2
JOIN_R	101
SCH_S	95
SCH_R	2
Rank	100
DATA_S	192
DATA_R	1345
Data_Sent_To_BS	237
Dist_CH_To_BS	305
Send_Code	16
Expaned Energy	69352

The table 3 shows specific attributes within the dataset. 'Id' with a value of 11120 indicates a unique identifier for the data entry, while 'Time' at 196 represents the corresponding timestamp. 'Is_CH' and 'Who CH' with values 2 and 7088 respectively denote specific network configurations. 'Dist_to_CH' stands at 13956, indicating a considerable distance metric. Activity counts include 'ADV_S' at 85, 'ADV_R' at 31, 'JOIN_S' at 2, 'JOIN_R' at 101, 'SCH_S' at 95, and 'SCH_R' at 2, showcasing varied activities within the network. 'Rank' is at its maximum value of 100, suggesting a high ranking level. Data transmission metrics include 'DATA_S' at 192, 'DATA_R' at 1345, and 'Data_Sent_To_BS' at 237, reflecting substantial data exchange. Distance-related features include 'Dist_CH_To_BS' at 305, indicating the distance between nodes and base stations. 'Send_Code' at 16 denotes specific coding activities. 'Expaned Energy' stands significantly high at 69352, suggesting an extensive energy expansion. These values offer a glimpse into the diverse and dynamic nature of the data, providing a specific instance that contributes to the overall dataset's complexity in the context of Wireless Sensor Networks analysis.

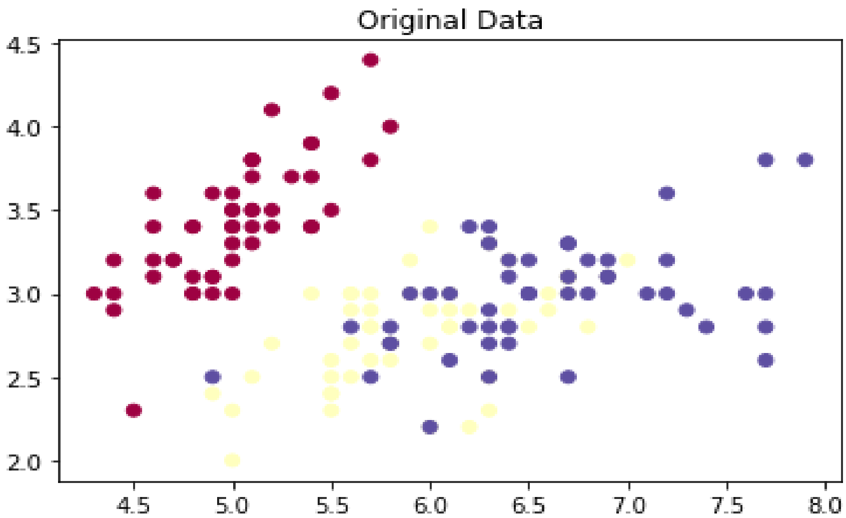

Fig 2: Original data

Figure 3 represents the original dataset in a visual format, providing a graphical overview of the data attributes and their relationships. Each axis likely represents a specific feature from the dataset, and the points or patterns on the graph illustrate the data points' distribution and clustering within this multidimensional space.

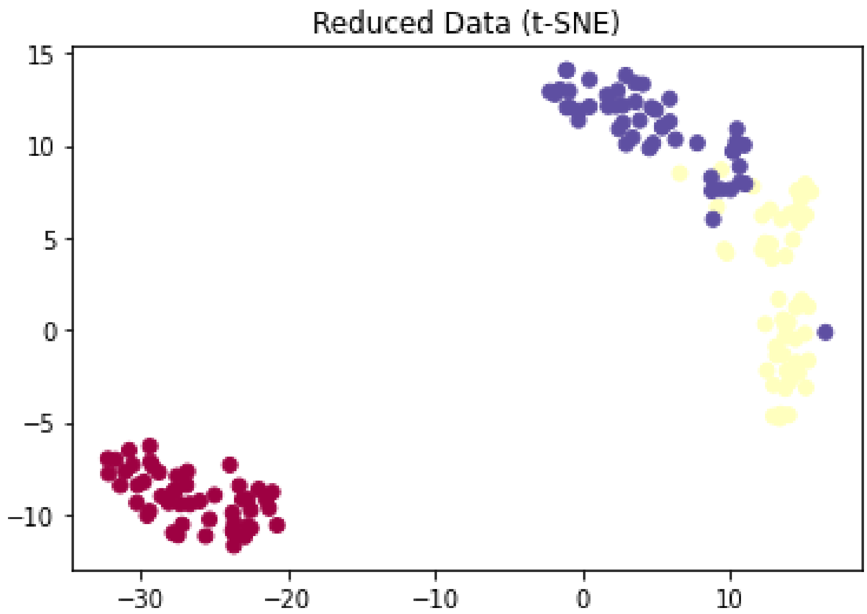

Fig 3: Reduced data (t-SNE)

Figure 4 illustrates the dataset after undergoing dimensionality reduction, likely using techniques such as topological nonlinear dimensionality reduction via manifold learning. This reduced data representation condenses the original high-dimensional feature space into a lower-dimensional format while preserving the essential topological and geometric properties of the data.

Table 4: Preprocessing accuracy comparison table

	Algorithm	Preprocessing Accuracy
Existing methods	Standard Scalar	33
	Label Encoding	55
	Principal Component Analysis	56
	Linear Discriminant Analysis	60
	Recursive Feature Elimination	88
Proposed methods	TNDRM	99

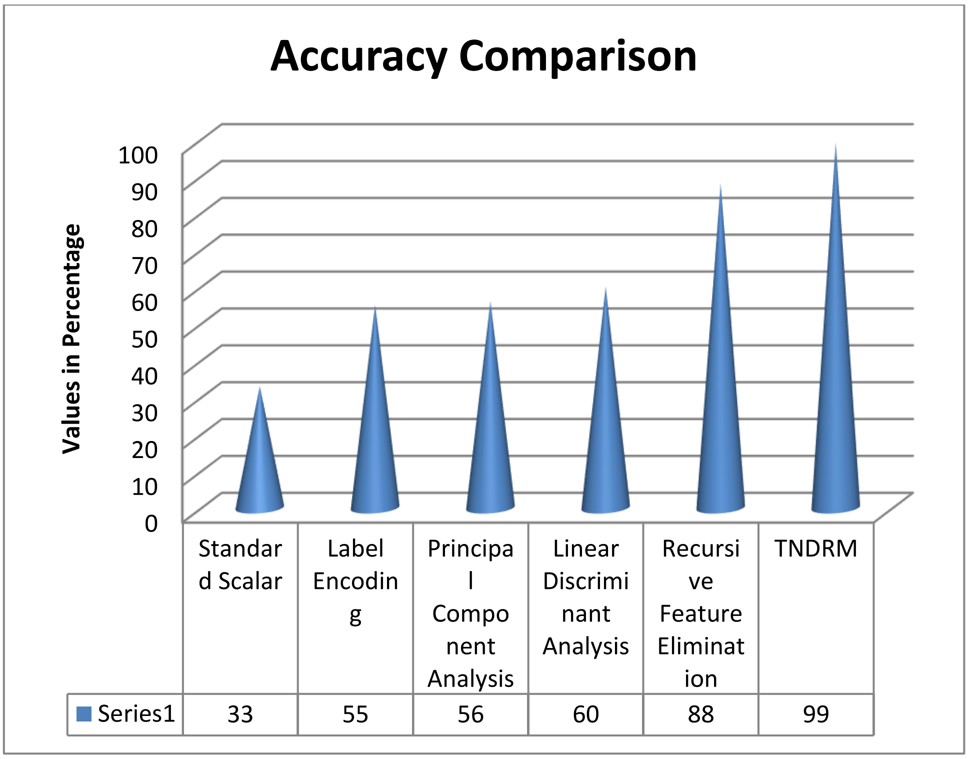

Fig 4: Preprocessing accuracy comparison chart

The table 4 and figure 4 shows, various algorithms were employed for data preprocessing, and their effectiveness was evaluated based on preprocessing accuracy scores. Among the existing methods, Principal Component Analysis (PCA) achieved an accuracy of 56%, Linear Discriminant Analysis (LDA) reached 60%, and Recursive Feature Elimination (RFE) resulted in 88% accuracy. In comparison, a proposed method referred to as TNDRM (presumably a new technique) demonstrated the highest preprocessing accuracy of 99%. These accuracy scores indicate the efficacy of each algorithm in transforming and preparing the data for subsequent analysis or machine learning tasks. A higher accuracy score suggests that the corresponding preprocessing method was more successful in capturing essential features or reducing the dimensionality of the data while preserving relevant information. Therefore, TNDRM outperforms the existing methods, showcasing its potential as a promising technique for data preprocessing.

5. Conclusion

Finally, TNDRM has presented a robust preprocessing approach for improving the accuracy of DDOS attack detection in Wireless Sensor Networks (WSNs). We tackled the delicate issues provided by the high-dimensional and complex nature of WSN data by mixing topological nonlinear dimensionality reduction via manifold learning with painstaking data standardization techniques. We effectively translated the complicated high-dimensional feature space of WSN data into a lower-dimensional representation while keeping the key topology and geometry of the original data using a variety of learning approaches. This change allowed for a deeper comprehension of the various data structures, establishing the groundwork for more effective analysis and detection. The use of z-score normalization and min-max scaling to normalize data within this reduced dimensional space was critical. Normalization maintained uniformity and consistency by minimizing the influence of various scales and outliers, giving a dependable input for machine learning algorithms. This process removed aberrations that may jeopardize the accuracy of the algorithms, improving the dependability of our DDOS attack detection system. For Further to improve the classification accuracy using the feature selection methods.

Reference

[1] Abidoye, A. P., & Obagbuwa, I. C. (2018). DDoS attacks in WSNs: detection and countermeasures . IET Wireless Sensor Systems, 8(2), 52–59. doi:10.1049/iet-wss.2017.0029

[2] Al-issa Abdulaziz I., Al-Akhras, M., ALsahli, M. S., & Alawairdhi, M. (2019). Using Machine Learning to Detect DoS Attacks in Wireless Sensor Networks. 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). doi:10.1109/jeeit.2019.8717400

[3] Anand, C., & Vasuki, N. (2021). Trust Based DoS Attack Detection in Wireless Sensor Networks for Reliable Data Transmission. Wireless Personal Communications. doi:10.1007/s11277-021-08855-9

[4] Chen, H., Meng, C., Shan, Z., Fu, Z., & Bhargava, B. K. (2019). A novel Low-rate Denial of Service attack detection approach in ZigBee wireless sensor network by combining Hilbert-Huang Transformation and Trust Evaluation. IEEE Access, 1–1. doi:10.1109/access.2019.2903816

[5] Edlund, A., Yang, Y., Yooseph, S., He, X., Shi, W., & McLean, J. S. (2018). Uncovering complex microbiome activities via metatranscriptomics during 24 hours of oral biofilm assembly and maturation. Microbiome, 6(1). doi:10.1186/s40168-018-0591-4

[6] Fong, Y., & Xu, J. (2021). Forward Stepwise Deep Autoencoder-Based Monotone Nonlinear Dimensionality Reduction Methods. Journal of Computational and Graphical Statistics, 30(3), 519–529. doi:10.1080/10618600.2020.1856119

[7] Gadd, C., Xing, W., Nezhad, M. M., & Shah, A. A. (2018). A Surrogate Modelling Approach Based on Nonlinear Dimension Reduction for Uncertainty Quantification in Groundwater Flow Models. Transport in Porous Media. doi:10.1007/s11242-018-1065-7

[8] Gajamannage, K., Paffenroth, R., & Bollt, E. M. (2018). A Nonlinear Dimensionality Reduction Framework Using Smooth Geodesics. Pattern Recognition. doi:10.1016/j.patcog.2018.10.020

[9] Guo, W., & Banerjee, A. G. (2017). Identification of key features using topological data analysis for accurate prediction of manufacturing system outputs. Journal of Manufacturing Systems, 43, 225–234. doi:10.1016/j.jmsy.2017.02.015

[10] He, X., Sun, L., Lyu, C., & Wang, X. (2020). Quantum locally linear embedding for nonlinear dimensionality reduction. Quantum Information Processing, 19(9). doi:10.1007/s11128-020-02818-y

[11] Islam, M. T., & Xing, L. (2021). Geometry and statistics-preserving manifold embedding for nonlinear dimensionality reduction. Pattern Recognition Letters, 151, 155–162. https://doi.org/10.1016/j.patrec.2021.07.012

[12] Jinhui, X., Yang, T., Feiyue, Y., Leina, P., Juan, X., & Yao, H. (2018). Intrusion Detection System for Hybrid DoS Attacks using Energy Trust in Wireless Sensor Networks*. Procedia Computer Science, 131, 1188–1195. doi:10.1016/j.procs.2018.04.297

[13] Kaya, S., & Vural, E. (2019). Multi-Modal Learning With Generalizable Nonlinear Dimensionality Reduction. 2019 IEEE International Conference on Image Processing (ICIP). doi:10.1109/icip.2019.8803196

[14] Liu, Z., Wu, C. T., & Koishi, M. (2018). A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. Computer Methods in Applied Mechanics and Engineering. doi:10.1016/j.cma.2018.09.020

[15] Ma, M., Deng, T., Wang, N., & Chen, Y. (2018). Semi-supervised rough fuzzy Laplacian Eigenmaps for dimensionality reduction. International Journal of Machine Learning and Cybernetics. doi:10.1007/s13042-018-0784-y

[16] Paul, R., & Chalup, S. K. (2017). A study on validating non-linear dimensionality reduction using persistent homology. Pattern Recognition Letters, 100, 160–166. doi:10.1016/j.patrec.2017.09.032

[17] Puschmann, D., Barnaghi, P., & Tafazolli, R. (2018). Using LDA to Uncover the Underlying Structures and Relations in Smart City Data Streams. IEEE Systems Journal, 12(2), 1755–1766. doi:10.1109/jsyst.2017.2723818

[18] Rafique, O., & Mir, A. H. (2019). A Topological Approach for Cancer Subtyping from Gene Expression Data. Journal of Biomedical Informatics, 103357. doi:10.1016/j.jbi.2019.103357

[19] Ramesh, S., Yaashuwanth, C., Prathibanandhi, K., Basha, A. R., & Jayasankar, T. (2021). An optimized deep neural network based DoS attack detection in wireless video sensor network. Journal of Ambient Intelligence and Humanized Computing. doi:10.1007/s12652-020-02763-9

[20] Tasoulis, S., Pavlidis, N. G., & Roos, T. (2020). Nonlinear Dimensionality Reduction for Clustering. Pattern Recognition, 107508. doi:10.1016/j.patcog.2020.107508

[21] Tran, Q. H., Chen, M., & Hasegawa, Y. (2021). Topological persistence machine of phase transitions. Physical Review E, 103(5). doi:10.1103/physreve.103.052127

[22] Trozzi, F., Wang, X., & Tao, P. (2021). UMAP as a Dimensionality Reduction Tool for Molecular Dynamics Simulations of Biomacromolecules: A Comparison Study. The Journal of Physical Chemistry B, 125(19), 5022–5034. doi:10.1021/acs.jpcb.1c02081

[23] Yang, Y., Sun, Z. Z., Ran, S. J., & Su, G. (2021). Visualizing quantum phases and identifying quantum phase transitions by nonlinear dimensional reduction. Physical Review B, 103(7), 075106.

[24] Yousaf, M., Rehman, T. U., & Jing, L. (2020). An Extended Isomap Approach for Nonlinear Dimension Reduction. SN Computer Science, 1(3). doi:10.1007/s42979-020-00179-y