engineering proceedings

MDPI

Proceeding Paper

Dimensionality Reduction Algorithms in Machine Learning: A Theoretical and Experimental Comparison †

Ashish Kumar Rastogi [1,*], Swapnesh Taterh [1] and Billakurthi Suresh Kumar [2]

[1] Amity Institute of Information Technology, Amity University Rajasthan, Jaipur 300202, India; swapnesh@hotmail.com
[2] Computer Science & Engineering, Sanjay Ghodawat University Kolhapur, Kolhapur 416118, India; sureshkumarbillakurthi@gmail.com
* Correspondence: ashish.k.rastogi24@gmail.com
† Presented at the International Conference on Recent Advances on Science and Engineering, Dubai, United Arab Emirates, 4–5 October 2023.

Abstract: The goal of Feature Extraction Algorithms (FEAs) is to combat the dimensionality curse, which renders machine learning algorithms ineffective. The most representative FEAs are investigated conceptually and experimentally in our work. First, we discuss the theoretical foundation of a variety of FEAs from various categories like supervised vs. unsupervised, linear vs. nonlinear and random-projection-based vs. manifold-based, show their algorithms and compare these methods conceptually. Second, we determine the finest sets of new features for various datasets, as well as in terms of statistical significance, evaluate the eminence of the different types of transformed feature spaces and power analysis, and also determine the FEA efficacy in terms of speed and classification accuracy.

Keywords: Feature Extraction; dimensionality reduction (DR); manifold; multi-class datasets

check for updates

Citation: Rastogi, A.K.; Taterh, S.; Kumar, B.S. Dimensionality Reduction Algorithms in Machine Learning: A Theoretical and Experimental Comparison. *Eng. Proc.* **2023**, *59*, 82. https://doi.org/10.3390/engproc2023059082

Academic Editors: Nithesh Naik, Rajiv Selvam, Pavan Hiremath, Suhas Kowshik CS and Ritesh Ramakrishna Bhat

Published: 19 December 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Every day, a substantial amount of data are collected, but only a fraction of them are valuable for decision-making. Machine learning algorithms (MLAs) can process large volumes of data, yet their effectiveness diminishes as data dimensionality increases. The complexity of learned models grows with the addition of more observations and attributes, often leading to overfitting and reduced performance on new data. Including irrelevant features also hampers model accuracy. Dimensionality reduction methods, specifically Feature Extraction Algorithms (FEAs), aim to mitigate these challenges by reducing data complexity and enhancing data quality. This study focuses on FEAs due to their ability to address real-world dataset issues like noise, complexity, and sparsity.

FEAs condense related features into artificial ones, preserving essential qualities while minimizing information loss [1]. They identify a representation of a manifold and project input data into a lower-dimensional space embedded within the input space. FEAs offer several advantages, such as enhancing MLA performance, preventing overfitting, reducing computation time and storage requirements, and easing data visualization [2].

Various FEAs exist, each suitable for different data features and sizes. Nonlinear FEAs excel with well-sampled smooth manifolds, whereas linear methods like principal component analysis (PCA) work effectively on linear surfaces. However, nonlinear FEAs might not consistently outperform PCA, and their effectiveness can be influenced by the dimensionality curse. Additionally, the accuracy of supervised FEAs depends significantly on dataset sample sizes [3].

This study delves into the underlying theories of FEAs, aiming to simplify their complex concepts and origins. It provides streamlined algorithms for numerous FEAs, categorizing them and outlining fundamental principles. Furthermore, this research conducts a

comprehensive empirical evaluation of FEAs on three real-world datasets of varying dimensionalities and classification settings, utilizing diverse quality metrics. To our knowledge, this study represents the first extensive exploration that combines theoretical understanding with extensive empirical analysis of FEAs [4].

2. Problem Statement

The main problem that is faced by organizations is the inability to properly make decisions on the basis of available data. In such conditions, machine learning provides capabilities to present predictive outcomes with accuracy while utilizing historical data. The dataset is provided as input to the machine learning algorithms and then it is trained to perform analysis based on the important decisions being made. It also helps in simplifying the complex tasks through which issues like decision making, predicting customer behavior, identifying attacks, etc., can be solved.

3. Literature Review

All in context to the original variables, the disjoint subsets and their linear combinations are the components that are determined by the principal component analysis [5]. The data matrix's singular value decomposition or SVD are based on the various multivariate techniques. A system of orthogonal principal axes can be decomposed through the use of this method [6]. Maximum variance is selected as a criterion for selecting such principal axes. For the calculation of principal components, a new methodology is proposed [7]. By the means of particle swarm, the methodology is able to optimize constraints. Binary optimization is used by this methodology. On the basis of intelligent search, the proposed methodology is able to compute principal components [8]. In case of high computational complexity, the solutions are obtained through the use of a stochastic optimization method [9].

One of the typical representations of ensemble learning is the random forest algorithm. Rolling bearing fault diagnosis is the area where this algorithm is widely used [10]. Based on an enhanced random forest algorithm and Spark, an efficient rolling bearing fault diagnosis method is proposed. In case of random forest algorithm in rolling bearing fault diagnosis, this method can be used to address the issues of repeated voting and slower diagnosis [11]. The Spark platform is used for the realization of the proposed IRF algorithm and its parallelization. The traversal of the sub-forest, training of the original random forest model, and the development of a decision tree similarity matrix are the three procedures that form parallelizations.

In the context of information systems, the security of communication can be improved through the use of efficient security tools such as intrusion detection systems or IDS. Malicious network traffic is detected regularly by such systems to ensure secure communication. For anomaly and intrusion detection, one of the most effective algorithms is the Enhanced J48 Classification Algorithm [12]. Various attacks can be detected with the help of this method. In comparison to the traditional algorithms, higher accuracy is displayed by this new method. By the effective utilization of the SD coefficient, the enhanced J48 algorithm's detection accuracy is increased. In context to the split value, aspects like entropy, gain ratio, and election are improved by the utilization of the SD coefficient in this algorithm.

The prediction of diabetes-inflicted diseases and diabetes is carried out through the use of various methods. The machine learning algorithms support vector machine and random forest are used in the proposed work. The probability of becoming affected by Diabetes-Related Diseases is also identified with the help of this method. By implementing step backward and forward selection, the selection of the features that effect the prediction is carried out [13]. Collection of the random and structured data is one of the key goals of the classification problem. Further, using random forest and support vector machine (SVM) algorithms for classifying the collected data in a preferred order is also performed by the classification problem.

To enhance the performance of the energy consumption prediction models, the noisy features are removed with the help of a pre-processing method. Further, predication methods are also used for the same. Principal component analysis (PCA) forms the basis of the proposed pre-processing method. The energy and historical meteorological data of buildings are also treated using the proposed pre-processing method [14]. Support vector regression, linear regression, random forest, regression tree, and k-nearest neighbors are the five prediction models that are used for processing and cleaning data. From any big dataset, a small dataset can be acquired by the practitioners through the use of the proposed method. This is confirmed by the results. These datasets should belong for energy consumption prediction problems. In context to the R2, execution time, mean square error, and residual values, the proposed prediction model is considered the best for all climate zones.

High feature redundancy of medical data, a high-dimensional feature space, and the issue of data imbalance are the topics that the proposed model aims to address. Various supervised classifiers are explored in this study. The diabetes survey sample data are classified through the use of two feature dimensionality reduction and SVM-SMOTE methods [15]. These data are based on the complex related factors and unbalanced categories. High-risk people of DM can be identified with the help of the LASSO feature reduction method that is equipped with the SVM-SMOTE and random forest classifier algorithm. Early screening of DM can be easily performed with the help of the combined method that is proposed in the research.

To attain resilience towards external and cognitive noises, a two-step filtering approach is used in the current study. The pattern mining and dynamic nature of the motor imagery EEG signals is explored with the help of four data reduction techniques and empirical wavelet transform. Independent component analysis, neighborhood component analysis, principal component analysis (PCA) and linear component analysis are the three reduction techniques that are used for the exploration. In relation to MI tasks, hidden patterns are explored with the help of EWT. Various modes are derived by decomposing the EEG data. In this study, feature vector is the term that is used to refer to each mode [16]. To perform dimensionality reduction of a huge feature matrix, the modes are processed under each data reduction technique.

For the fault diagnosis, comprehensive information is provided using data of the multi-channel: The use of common methods for the purpose of analysis is restricted by the redundant and cross-correlation information in the given data [17]. The characteristics and tensor structure of a dataset of multi-channel are explored in this research paper. In deep learning technologies, the introduction of the multilinear subspace learning algorithm has led to the proposition of a method of novel fault diagnosis. With the help of multilinear principal com-ponent analysis, a reduction in the multi-channel data dimension is performed. The tensor structure is not destroyed by this activity [18]. For fault diagnosis, the development of a classification model and extraction of features are performed with the help of the CNN.

It is quite important to use an optimization technique in smartphones. This is because it will result in the proper working of the activity recognition system on the smartphone. Power constraint is used in this system. A lesser time consumption and a reduced number of features that are used in the dataset are achieved with the help of an effective optimization technique. The fast feature dimensionality reduction technique is the dimensionality reduction technique that is proposed in this research paper. In this study, public-domain-related datasets are used [19]. The number of features for the dataset has been reduced to 66 through the use of a fast feature dimensionality reduction technique. This is displayed by the outcomes of this study. Around 98.72% activity recognition accuracy is maintained by this technique. Through the use of FFDRT, this accuracy is maintained in the dimensionality reduction stage.

Constant improvement in the use of artificial intelligence can be seen in the development of machine learning and artificial intelligence. Deep learning is major technology of machine learning that is evolving. The importance of feature learning is clarified by

deep learning. The process of prediction or classification is simplified through the use of layer-by-layer feature transformation. A new feature space is created by transforming the original feature space through the use of feature representation [20]. For risk analysis, the key experiments performed are the experiments of the random forest model training and the optimization experiment. For large-scale group activities, promising predictive ability is displayed by the random forest algorithm. A maximum of a 0.86 classification accuracy rate is achieved by the random forest algorithm.

Fog, atmospheric particles, rain, scattering of light, dust, and other aspect that result in noisy point cloud images can negatively impact the signals transmitted by LiDAR sensors. To filter LiDAR point clouds, a new noise reduction method is developed in this paper. This new noise reduction method is developed to address the transmission issues related to LiDAR. On the basis of principal component analysis, an adaptive clustering method is developed in this research [21]. The two-dimension data are generated by the proposed method with the help of dimension reduction. With little information attrition, the proposed method can extract the second and first principal components of the original data.

To detect network-related attacks, efficient tools like machine learning algorithms can be used. Anomaly detection is the main task of machine learning algorithms. Zero- day attacks can be detected with high accuracy through the use of Naïve Bayes and decision tree algorithms [22]. The class of the traffic is predicted through the use of conditional probability in the case of the Naïve Bayes algorithm, given that some past knowledge about the traffic is given to the developer [23]. A tree structure model is developed by the J48 decision tree which can be used for the classification problems. In comparison to the Naïve Bayes technique, a better performance in anomaly detection is displayed by the J48 technique for this dataset [24].

For summarizing and organizing data, various useful and established approaches are discussed in this research. The problem of data mining does not have a definitive solution [25]. Principal component analysis (PCA) is considered one of the most general among all the machine learning solutions for addressing the problem of data mining. In data analysis, the title of the standard approach is given to PCA. Through decorrelation, PCA is able to reduce dimensionality [26]. In addition to decreasing dimensionality, it also preserves data variance. A wide range of variance concentration patterns can be used for the considered data. Across different fields, the first principal components used can also be used in addition to the variance concentration patterns [27].

4. Principal Component Analysis

Principal component analysis (PCA) is an unsupervised, linear transformation method that identifies the highest data variance, generating new features known as principal components (PCs) [28]. In PCA, orthogonal axes or PCs indicate the directions of maximal data variance, transforming a high-dimensional dataset into a new subspace. The initial PC displays the greatest variance, while subsequent PCs exhibit decreasing variances [29]. Figure 1 depicts the application of PCA to the ECG200 dataset, comparing the new space with three features (orange) to the original space with 96 features (light blue). The visualization illustrates how the updated data points are more densely clustered, mitigating noise and redundancies.

In the given Figure 1 illustrating the application of Principal Component Analysis (PCA), the original feature space is represented by a cloud of data points in a high-dimensional setting, each point characterized by multiple features. PCA identifies the principal components, which are orthogonal axes in the original space that capture the maximum variance. The reduced feature space is then visualized by projecting the data points onto a subset of these principal components, effectively transforming the data into a lower-dimensional space. This reduced space retains the essential patterns and relationships present in the original data, facilitating a more concise representation that is conducive to analysis, interpretation, and potentially improved computational efficiency. The figure visually demonstrates how PCA simplifies the data representation by emphasiz-

ing the directions of greatest variability and creating a reduced feature space with fewer dimensions while preserving the crucial information in the dataset.

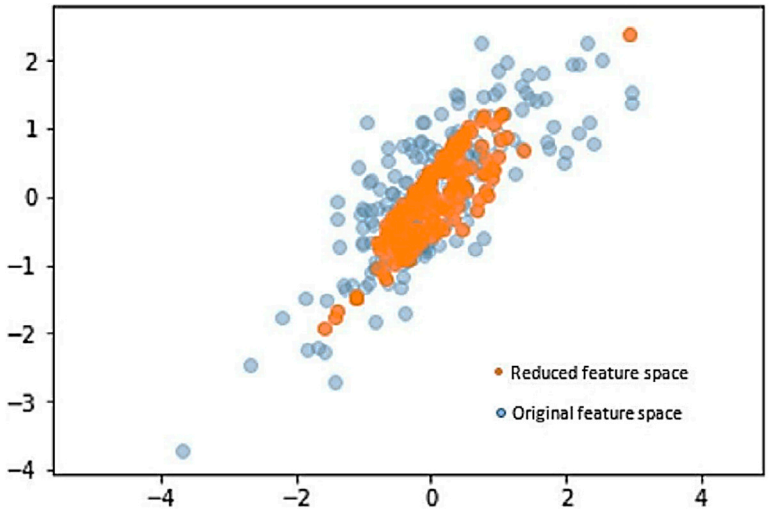


Figure 1. Original vs. reduced feature spaces [29].

5. Proposed Model

1. Data Pre-processing—In the prediction, inaccurate results might be generated because some attributes of the dataset have missing values. The model's accuracy is reduced by such missing values. An efficient mean value is used for treating each attribute whether it has all values or is missing some values.
2. Feature Selection—The performance of all the data mining algorithms can be improved through the use of effective feature selection. The performance of data classification is also enhanced by feature selection. Only some of the features of the dataset are important. Additional information is not provided by the redundant features that are present in the dataset. With regard to the context, not an ounce of helpful information is provided by the irrelevant features.
3. Classification and analysis—Analysis of results and classification of the model are conducted after processing data using dimensionality reduction and feature selection methods. The output parameters are mapped over by the input parameters with the help of classification. On the basis of parameters, conclusions are drawn with the help of classification. The output prediction is also assisted by effective classification.

Figure 2 demonstrate proposed conceptual model for intrusion detection, it appears that the system employs a machine learning-based approach to identify and classify potential intrusions in a network. The model likely utilizes a combination of features extracted from network traffic data, possibly employing techniques such as deep packet inspection or statistical analysis. The figure suggests a multi-step process involving preprocessing, feature extraction, and a machine learning classifier, possibly a supervised algorithm trained on labeled data to distinguish between normal and malicious network activities. The feedback loop and iterative nature of the system imply continuous learning and adaptation to emerging threats. Additionally, the incorporation of a visualization component in the figure suggests a means of interpreting and analyzing the model's outputs. Overall, the figure outlines a comprehensive intrusion detection system that leverages machine learning for automated, real-time detection and response to potential security threats within a network environment.

In the provided Table 1 comparing dimensionality reduction techniques, several methods are assessed based on key criteria. Principal Component Analysis (PCA) stands out for its ability to capture maximum variance and reduce dimensionality while maintaining interpretability, particularly in linear relationships. t-Distributed Stochastic Neighbor Embedding (t-SNE) excels in visualizing complex, non-linear structures by preserving

local relationships, but it may struggle with global structures. Autoencoders, as neural network-based approaches, offer flexibility in capturing intricate patterns but may be computationally demanding. Random Projection provides a computationally efficient option but might be less effective in capturing nuanced relationships. Each technique has its strengths and weaknesses, making the choice dependent on the specific characteristics of the data and the objectives of the analysis. The table serves as a valuable guide for selecting the most suitable dimensionality reduction method based on the desired outcomes and computational considerations.

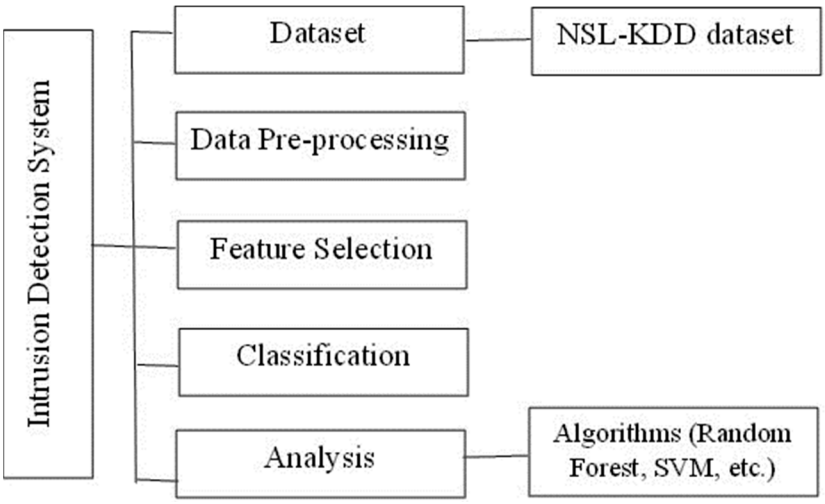


Figure 2. Proposed conceptual model for intrusion detection [30].

Table 1. Comparison table [30].

Proposed Solution	Advantages	Limitations
Disjoint principal component analysis method	It helps in resolving the issue related with multivariate analysis and dimensionality reduction.	The solution quality degrades with the increase in the size of the dataset provided to the method.
Rolling Bearing Fault Diagnosis Method	It provides a better fault diagnosis accuracy and has high training and modelling speed for the datasets.	These methods can provide low fault diagnosis when imbalanced datasets are taken with very few samples.
Enhanced J48 classification algorithm	It provides efficient performance and accuracy in feature selection process.	The proposed method is not yet implemented in the real network environments.
Principle component analysis method	It helps in increasing the test set accuracy of the random forest and SVM through dimensionality reduction.	The accuracy reduces with the feature elimination of SVM algorithm.
PCA-based prediction method	It helps in predicting building's energy consumption through random forest, SVR and linear regression method accurately in less execution time.	However, it can be less useful for the times of different energy consumption patterns and in lower ambient temperature.
SVM-SMOTE and LASSO combined with random forest	Random forest classifier combining with SVM-SMOTE and LASSO feature reduction method can accurately predict.	The performance of model required better understanding of parameter optimization methods.
Method for pattern mining	Effective method of two-step filtering the data.	Slow process of model training and testing time.
Fault diagnosis method	Lower computational cost and superior performance.	The multi-channel data should be homogenous.
Principal component analysis algorithm for dimensionality reduction	Speed up the processing and storage process by reducing high dimensions in large datasets.	It is less efficient in low-dimensional data.

Table 1. *Cont.*

Proposed Solution	Advantages	Limitations
Fast feature dimensionality reduction technique	Reduced computational cost and high accuracy rate.	Less response time.
Random forest algorithm	In risk assessment, it shows high predictive ability.	Restricted to only large-scale group activities.
Principal component analysis based adaptive clustering method	High recall and precision.	Cannot remove all the noise from dataset.
Naïve Bayes and J48 algorithms for network intrusion detection systems	In detection of anomalies, better performance is displayed by J48.	There are more chances of false alarms in case of Naïve Bayes algorithm then J48 algorithm.
Principal component analysis	Uses decorrelation to reduce dimensionality.	Not efficient when comes to simplifying datasets.

6. Summary and Comparison

SVD employs the SVD theorem to minimize reconstruction error in an unsupervised, linear, iterative manner, based on random projections. It has two hyper-parameters: the number of PCs and iterations. ISOMAP, on the other hand, calculates pairwise geodesic distances using an adjacent graph. It is unsupervised, nonlinear, non-iterative, manifold-based, and has two hyper-parameters: the number of PCs and nearest neighbors.

7. Conclusions

In this study, we conducted a conceptual analysis and comparison of various Feature Extraction Algorithms (FEAs) across different categories. Subsequently, the performance of these FEAs was empirically evaluated and compared on diverse datasets, encompassing both binary and multi-class scenarios. Evaluation was based on correlation, classification accuracy, and runtime measures to assess the quality of the transformed feature spaces. The experimental results led to two significant conclusions. Firstly, the application of FEAs led to noticeable improvements in data quality and classification accuracy, sometimes even dramatic improvements. Secondly, manifold-based FEAs exhibited superior performance compared to random projection-based FEAs, and nonlinear FEAs consistently outperformed linear FEAs.

Furthermore, in the context of multi-class scenarios, supervised FEAs demonstrated superior performance over unsupervised ones, with one unsupervised FEA showing subpar results. Our future research endeavors will involve exploring the application of FEAs to assess the performance of deep learning models on high-dimensional datasets. Additionally, we plan to investigate the efficacy of FEAs on more challenging datasets, such as multidimensional time series and multi-label data.

Author Contributions: Conceptualization, A.K.R. and S.T.; methodology, A.K.R.; validation, B.S.K. and A.K.R.; formal analysis, S.T. and B.S.K.; investigation, S.T. and B.S.K.; resources, A.K.R.; data curation, B.S.K. and A.K.R.; writing—original draft preparation, A.K.R.; validation, S.T. and B.S.K.; writing—review and editing, A.K.R.; visualization, S.T. and A.K.R.; supervision, S.T. and B.S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset used in this research is publicly available at https://www.kaggle.com/datasets/shayanfazeli/heartbeat/ (accessed on 3 February 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ramirez-Figueroa, J.A.; Martin-Barreiro, C.; Nieto-Librero, A.B.; Leiva, V.; Galindo-Villardón, M.P. A new principal component analysis by particle swarm optimization with an environmental application for data science. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1969–1984. [CrossRef]
2. Wan, L.; Gong, K.; Zhang, G.; Yuan, X.; Li, C.; Deng, X. An efficient rolling bearing fault diagnosis method based on spark and improved random forest algorithm. *IEEE Access* **2021**, *9*, 37866–37882. [CrossRef]
3. Aljawarneh, S.; Yassein, M.B.; Aljundi, M. An enhanced J48 classification algorithm for the anomaly intrusion detection systems. *Clust. Comput.* **2019**, *22*, 10549–10565. [CrossRef]
4. Sivaranjani, S.; Ananya, S.; Aravinth, J.; Karthika, R. Diabetes prediction using machine learning algorithms with feature selection and dimensionality reduction. In Proceedings of the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 19–20 March 2021; Volume 1, pp. 141–146.
5. Parhizkar, T.; Rafieipour, E.; Parhizkar, A. Evaluation and improvement of energy consumption prediction models using principal component analysis based feature reduction. *J. Clean. Prod.* **2021**, *279*, 123866. [CrossRef]
6. Wang, X.; Zhai, M.; Ren, Z.; Ren, H.; Li, M.; Quan, D.; Chen, L.; Qiu, L. Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 105. [CrossRef] [PubMed]
7. Sadiq, M.T.; Yu, X.; Yuan, Z. Exploiting dimensionality reduction and neural network techniques for the development of expert brain–computer interfaces. *Expert Syst. Appl.* **2021**, *164*, 114031. [CrossRef]
8. Guo, Y.; Zhou, Y.; Zhang, Z. Fault diagnosis of multi-channel data by the CNN with the multilinear principal component analysis. *Measurement* **2021**, *171*, 108513. [CrossRef]
9. Hasan, B.M.S.; Abdulazeez, A.M. A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *J. Soft Comput. Data Min.* **2021**, *2*, 20–30.
10. Hashim, M.; Amutha, R. Human activity recognition based on smartphone using fast feature dimensionality reduction technique. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 2365–2374. [CrossRef]
11. Chen, Y.; Zheng, W.; Li, W.; Huang, Y. Large group activity security risk as- sessment and risk early warning based on random forest algorithm. *Pattern Recognit. Lett.* **2021**, *144*, 1–5. [CrossRef]
12. Duan, Y.; Yang, C.; Chen, H.; Yan, W.; Li, H. Low-complexity point cloud de- noising for LiDAR by PCA-based dimension reduction. *Opt. Commun.* **2021**, *482*, 126567. [CrossRef]
13. Razdan, S.; Gupta, H.; Seth, A. Performance Analysis of Network Intrusion De- tection Systems using J48 and Naive Bayes Algorithms. In Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT), Mumbai, India, 2–4 April 2021; pp. 1–7.
14. Gewers, F.L.; Ferreira, G.R.; Arruda, H.F.D.; Silva, F.N.; Comin, C.H.; Amancio, D.R.; Costa, L.D.F. Principal component analysis: A natural approach to data exploration. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–34. [CrossRef]
15. Anowar, F.; Sadaoui, S. Incremental neural-network learning for big fraud data. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Toronto, ON, Canada, 11–14 October 2020; pp. 3551–3557.
16. Anowar, F.; Sadaoui, S. Incremental learning framework for real-world fraud detection environment. *Comput. Intell.* **2021**, *37*, 635–656. [CrossRef]
17. Spruyt, V. The curse of dimensionality in classification. *Comput. Vis. Dummies* **2014**, *21*, 35–40.
18. Van Der Maaten, L.; Postma, E.; Van den Herik, J. Dimensionality reduction: A comparative review. *J. Mach. Learn. Res.* **2009**, *10*, 13.
19. Jindal, P.; Kumar, D. A Review on Dimensionality Reduction Techniques. *Int. J. Comput. Appl.* **2017**, *173*, 42–46. [CrossRef]
20. Verleysen, M.; François, D. The curse of dimensionality in data mining and time series prediction. In *International Work-Conference on Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 758–770.
21. Hawkins, D.M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12. [CrossRef]
22. Abe, S. Feature selection and extraction, in: Support Vector Machines for Pattern Classification. In *Advances in Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 331–341.
23. Yan, J.; Zhang, B.; Liu, N.; Yan, S.; Cheng, Q.; Fan, W.; Yang, Q.; Xi, W.; Chen, Z. Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 320–333. [CrossRef]
24. Chao, G.; Luo, Y.; Ding, W. Recent Advances in Supervised Dimension Reduction: A Survey. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 341–358. [CrossRef]
25. Gracia, A.; González, S.; Robles, V.; Menasalvas, E. A methodology to compare dimensionality reduction algorithms in terms of loss of quality. *Inform. Sci.* **2014**, *270*, 1–27. [CrossRef]
26. Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extrac- tion techniques in machine learning. In Proceedings of the 2014 Science and Information Conference, IEEE, London, UK, 27–29 August 2014; pp. 372–378.
27. Joshi, P. What Is Manifold Learning? 2014. Available online: https://prateekvjoshi.com/2014/06/21/what-is-manifold-learning/ (accessed on 10 February 2023).
28. Garrett, D.; Peterson, D.; Anderson, C.; Thaut, M. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2003**, *11*, 141–144. [CrossRef] [PubMed]

29. Rastogi, A.K.; Taterh, S.; Kumar, B.S. Dimensionality Reduction Approach for High Dimensional Data using HGA based Bio Inspired Algorithm. *Int. J. Intell. Syst. Appl. Eng.* **2023**, *11*, 227–239.
30. Rastogi, A.K.; Taterh, S.; Kumar, B.S. Bio-Inspired Algorithms for Prey Model Optimization (February 2022). In Proceedings of the 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, Pradesh, India, 23–25 February 2022; pp. 264–269. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.