Indian Journal of Natural Sciences *www.tnsroindia.org.in ©IJONS*

Vol.15 / Issue 86 / Oct / 2024 *International Bimonthly (Print) – Open Access* *ISSN: 0976 – 0997*

REVIEW ARTICLE

A Literature Review of the Dimensionality Reduction Techniques for the Healthcare Domain

G. Chitra[1]* and Hari Ganesh S[2]

[1]Research Scholar, PG & Research Department of Computer Science, H.H. The Rajah's College (Autonomous), Pudukkottai, (Affiliated to Bharathidasan University, Tiruchirappalli), Pudukkottai, Tamil Nadu, India.

[2]Assistant Professor, PG & Research Department of Computer Science, H.H. The Rajah's College (Autonomous), Pudukkottai, (Affiliated to Bharathidasan University, Tiruchirappalli), Tamil Nadu, India.

Received: 21 Aug 2024 Revised: 03 Jul 2024 Accepted: 26 Oct 2024

***Address for Correspondence**
G. Chitra
Research Scholar, PG & Research Department of Computer Science,
H.H. The Rajah's College (Autonomous) , Pudukkottai,
(Affiliated to Bharathidasan University, Tiruchirappalli),
Tamil Nadu, India.

ABSTRACT

The rapid advancement of healthcare technology has led to an explosion of multi-modal data, including electronic health records, medical imaging, genomics, and wearable device outputs. This diverse data landscape poses significant challenges in terms of dimensionality reduction, which is essential for effective analysis and interpretation. This literature review explores various multi-modal data fusion techniques aimed at enhancing dimensionality reduction in healthcare analytics. We categorize the existing approaches into three main frameworks: feature-level fusion, decision-level fusion, and hybrid methods, each exhibiting unique strengths and limitations. The review critically evaluates recent studies that leverage machine learning algorithms, deep learning architectures, and statistical methods for integrating multi-modal data. By synthesizing findings from various domains, we highlight the impact of dimensionality reduction on predictive modeling, disease diagnosis, and personalized treatment strategies. Furthermore, we discuss the challenges and future directions in the field, emphasizing the need for robust methodologies that ensure data integrity and interpretability while maintaining patient privacy. This review aims to provide a comprehensive understanding of current trends and advancements in multi-modal data fusion techniques, offering insights for researchers and practitioners in the realm of healthcare analytics.

Keywords: Healthcare Analytics, Dimensionality Reduction, Data Fusion Technique

INTRODUCTION

In recent years, the healthcare sector has experienced a profound transformation driven by technological advancements and the proliferation of data [1] [2]. The integration of diverse data sources—ranging from electronic health records (EHRs) and medical imaging to genomic sequences and data from wearable devices—has created a multi-modal data landscape. Each of these modalities provides unique insights into patient health, but they also introduce significant complexity due to their inherent heterogeneity in format, structure, and information content. Multi-modal data fusion refers to the process of integrating data from multiple sources to achieve a more comprehensive understanding of a given phenomenon. In the context of healthcare analytics, effective data fusion is crucial for enhancing clinical decision-making, improving patient outcomes, and advancing personalized medicine. However, the diverse nature of these data sources often results in high dimensionality, making it challenging to extract meaningful patterns and insights. High-dimensional datasets can lead to issues such as the "curse of dimensionality," where the volume of the feature space increases exponentially, potentially diminishing the performance of machine learning algorithms and complicating the interpretability of models.

Dimensionality reduction techniques play a pivotal role in addressing these challenges by reducing the number of features while preserving the essential information required for analysis. By minimizing dimensionality, these techniques can improve computational efficiency, reduce storage costs, and enhance model performance. Moreover, effective dimensionality reduction can mitigate overfitting, improve generalization, and facilitate the visualization of complex data structures. Various methodologies have been proposed to achieve dimensionality reduction in the context of multi-modal data fusion. These can broadly be categorized into feature-level fusion, where data from different modalities is combined at the feature level before analysis, and decision-level fusion, where models built on individual modalities are combined to form a consensus decision. Hybrid approaches that integrate both feature-level and decision-level fusion techniques have also gained traction, promising to leverage the strengths of each method [3] [4].

Despite the significant progress in multi-modal data fusion techniques, several challenges remain. Issues such as data inconsistency, missing values, and the need for effective alignment of disparate data sources complicate the fusion process. Additionally, the selection of appropriate dimensionality reduction techniques that suit the specific characteristics of each data modality is critical for achieving optimal results [5] [6]. This literature review aims to synthesize the current state of research on multi-modal data fusion techniques for enhanced dimensionality reduction in healthcare analytics. By critically examining the existing methodologies, we will identify trends, challenges, and opportunities in this evolving field. Our review will provide insights into how these techniques can be effectively applied to improve predictive modeling, disease diagnosis, and personalized treatment strategies, ultimately contributing to the advancement of healthcare analytics. The findings of this review will serve as a valuable resource for researchers and practitioners seeking to harness the potential of multi-modal data in healthcare.

Background Study on Healthcare Analytics

Healthcare analytics is the systematic analysis of healthcare data to derive actionable insights that enhance decision-making, improve patient outcomes, and optimize operational efficiency. The integration of data science techniques and healthcare information systems has paved the way for the emergence of analytics as a vital tool in the healthcare sector. It encompasses a wide range of methodologies, including descriptive, predictive, and prescriptive analytics, each serving distinct purposes in analyzing healthcare data [7] [8].

Descriptive Analytics provides insights into historical data, allowing stakeholders to understand trends, patterns, and correlations within healthcare datasets. It typically involves the use of data visualization tools to present findings in an understandable format.

Predictive Analytics utilizes statistical algorithms and machine learning techniques to forecast future events based on historical data. In healthcare, predictive models are employed for various applications, such as predicting disease outbreaks, patient readmissions, and treatment outcomes.

Prescriptive Analytics goes a step further by recommending actions based on predictive insights. This form of analytics helps healthcare providers optimize resource allocation, enhance treatment plans, and improve patient care protocols. The effectiveness of healthcare analytics is heavily dependent on the quality and variety of data available. Data sources in healthcare are diverse, encompassing:

Electronic Health Records (EHRs): Digital versions of patients' paper charts, EHRs contain a wealth of information, including patient demographics, medical history, medications, treatment plans, and laboratory results.

Medical Imaging: Imaging modalities such as X-rays, MRIs, and CT scans generate complex data that require advanced analytics for interpretation. Techniques such as image recognition and analysis have emerged to extract valuable information from medical images.

Genomic Data: The advent of genomics has provided insights into individual genetic makeups, paving the way for personalized medicine. Analyzing genomic data enables healthcare providers to tailor treatments based on genetic predispositions.

Wearable Devices: Devices such as smartwatches and fitness trackers collect real-time health data, including heart rate, physical activity, and sleep patterns. This data is increasingly used to monitor chronic conditions and promote preventive care. The landscape of healthcare analytics has evolved significantly over the past few decades:

Early Days: Initially, healthcare analytics focused primarily on descriptive statistics to report on patient outcomes and operational efficiency. Data was often siloed, with limited integration across departments.

Emergence of Predictive Analytics: With the advent of advanced statistical methods and machine learning algorithms, predictive analytics gained traction in the late 2000s. Healthcare organizations began employing predictive models to forecast patient needs and identify high-risk patients.

Current Trends: Today, the integration of artificial intelligence (AI) and machine learning in healthcare analytics has revolutionized the field. AI-driven tools can analyze complex datasets, identify patterns, and provide real-time insights, leading to improved clinical decision-making and personalized treatment approaches.

Background Study on Dimensionality Reduction Techniques

Dimensionality reduction refers to a set of techniques aimed at reducing the number of input variables or features in a dataset while retaining its essential information. In the context of healthcare analytics, the explosion of data from diverse sources—such as electronic health records, medical imaging, genomic studies, and wearable devices—has resulted in high-dimensional datasets. While this wealth of data can provide rich insights into patient health and treatment outcomes, high dimensionality often leads to challenges such as overfitting, increased computational costs, and difficulties in visualization and interpretation. Dimensionality reduction techniques serve to simplify these complex datasets by extracting relevant features and eliminating redundant or irrelevant information. This simplification not only enhances the performance of machine learning algorithms but also aids healthcare practitioners in making more informed decisions based on the underlying patterns present in the data. Dimensionality reduction techniques can be broadly categorized into two main types: feature selection methods and feature extraction methods.

Feature Selection: This approach involves selecting a subset of the most relevant features from the original dataset [9][10]. It is particularly useful when the original features contain irrelevant or redundant information that can obscure meaningful patterns. Common feature selection techniques include:

Filter Methods: These methods evaluate the relevance of features based on their statistical properties, independent of the machine learning algorithm used. Examples include chi-square tests, correlation coefficients, and mutual information.

Wrapper Methods: These techniques use a predictive model to evaluate the performance of feature subsets. They iteratively select features based on model accuracy, often employing techniques like recursive feature elimination.

Embedded Methods: These approaches perform feature selection during the model training process, integrating the selection with the learning algorithm. Examples include Lasso regression and decision tree-based methods like Random Forest.

Feature Extraction: This approach transforms the original features into a new feature space with reduced dimensionality. Feature extraction techniques create composite features that capture the underlying structure of the data. Common feature extraction methods include:

Principal Component Analysis (PCA): A linear transformation technique that reduces dimensionality by projecting the data onto a lower-dimensional space while maximizing variance.

t-distributed Stochastic Neighbor Embedding (t-SNE): A nonlinear technique particularly suited for visualizing high-dimensional data in a lower-dimensional space. It emphasizes preserving local structures in the data.

Linear Discriminant Analysis (LDA): A supervised method that seeks to find the linear combinations of features that best separate different classes in the dataset.

Autoencoders: A type of neural network used for unsupervised feature extraction, where the model learns to encode the input data into a lower-dimensional representation and then reconstruct it. \
The application of dimensionality reduction techniques in healthcare is vast and varied, with implications across several areas:

Disease Diagnosis and Prediction: Dimensionality reduction techniques are often employed to enhance the performance of predictive models in disease diagnosis. By reducing the number of features, these techniques help mitigate overfitting and improve model interpretability, which is crucial in clinical settings. For example, PCA has been used to analyze genomic data for predicting cancer outcomes, while LDA has been applied to differentiate between various types of diseases based on clinical parameters.

Medical Imaging: In medical imaging, dimensionality reduction techniques are used to extract relevant features from high-dimensional image data, facilitating the identification of patterns indicative of specific conditions. For instance, PCA and autoencoders have been utilized to enhance image quality and reduce noise in MRI and CT scans.

Personalized Medicine: The integration of multi-modal data, including genomic, clinical, and lifestyle data, is essential for advancing personalized medicine. Dimensionality reduction techniques enable the identification of significant biomarkers that can inform tailored treatment plans, enhancing the efficacy of interventions.

Patient Monitoring and Wearable Devices: Data collected from wearable devices, which often include numerous features related to physical activity, heart rate, and sleep patterns, can be analyzed using dimensionality reduction techniques. This analysis helps in identifying trends and anomalies in patient health, facilitating timely interventions.

LITERATURE REVIEW

Rani, Ridhima, *et al* [11] In the era of big data, diverse data types characterised by extensive samples and high dimensionality are proving essential across various domains, including data mining, pattern recognition, machine learning, and the Internet of Things (IoT), among others. The intricacy of data processing escalates with the augmentation of the dataset's size. The term "complexity" denotes the challenge of identifying and utilising correlations among many elements of a dataset. Consequently, employing a dimensionality reduction (DR) method can eliminate the complexity among various features. This article examines the literature on data recovery approaches in the context of enhancing storage and processing of large data across various IoT applications, highlighting its advantages, characteristics, classification, and evaluation criteria. Moreover, the essay delineates prospective research issues and provides insights into the applications of data reduction (DR) across several areas, thereby informing readers about the relevance of a specific data reduction technique. Rashid, Lubaba, *et al* [12] intended to ascertain the impact of dimensionality reduction on IoT data on storage and communication expenses. It also examines the impact of dimensionality reduction of IoT data on the efficacy of various classification algorithms applied to it. Dimension reduction has been shown to decrease the storage and communication expenses of IoT data, albeit at the expense of the performance of classification algorithms applied to the reduced-dimensional data. Nonetheless, this decline in performance is insignificant relative to the optimisation of storage and transmission costs.

Ashraf, Mohsena, *et al* [13] Contemporary data analysis entails managing extensive datasets, including time-series data. This data is distinguished by its high dimensionality, substantial volume, and the existence of noise and redundant features. Nevertheless, the "curse of dimensionality" frequently presents challenges for learning methodologies, which may struggle to recognise the temporal correlations inherent in time-series data. To resolve this issue, it is imperative to diminish dimensionality while maintaining the inherent characteristics of temporal dependencies. This will mitigate diminished learning and prediction performance. This paper introduces twelve distinct dimensionality reduction techniques tailored for time-series data, categorised by supervision, linearity, time and memory complexity, hyper-parameters, and limitations. Vinutha, M. R., *et al* [14] An Enhanced Principal Component Analysis (EPCA) is proposed, which minimises the dimensions of the medical dataset while meticulously preserving critical information, thereby attaining superior outcomes. The notable dimensionality reduction approaches, including Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Partial Least Squares (PLS), Random Forest, Logistic Regression, Decision Tree, and the proposed EPCA, are examined in relation to the following Machine Learning (ML) algorithms: Support Vector Machine (SVM), Artificial Neural Networks (ANN), Naïve Bayes (NB), and Ensemble ANN (EANN) evaluated by statistical metrics including F1 score, precision, accuracy, and recall. EPCA directly transferred the data to a lower-dimensional space to enhance the distribution of the data in that form.

Henouda, Salah Eddine, *et al* [15] This study aims to examine the impact of dimensionality reduction methods (DRTs) on the classification of breast cancer (BC). We concentrated on the following five dimensionality reduction techniques (DRTs): Auto-Encoders (AE), T-Distributed Stochastic Neighbour Embedding (T-SNE), Recursive Feature Elimination (RFE), Isometric Feature Mapping (Isomap), and Principal Component Analysis (PCA). These methods are integrated with two renowned classifiers: Support Vector Machine (SVM) and Multilayer Perceptron (MLP). They are utilised for BC categorisation. The Breast Cancer Wisconsin Diagnostic (WDBC) dataset was utilised to validate the experiments conducted in this study. The former was supplied by the machine learning repository of the University of California, Irvine (UCI). Ahmad, Noor, and Ali Bou Nassif [16] Dimensionality reduction strategies are essential for the analysis and interpretation of high-dimensional data. These strategies collect various data attributes of significance, including dynamic structure, input-output linkages, inter-data set correlation, covariance, and others. Dimensionality reduction involves transforming a collection of high-dimensional data features into a lower-dimensional representation. This study addresses the inadequate performance of learning models caused by high-dimensional data by examining five distinct dimensionality reduction techniques. A comprehensive comparison is made between reduced dimensionality data and the original dataset utilising statistical and machine learning algorithms.

Rafieian, Bardia, Pedro Hermosilla, and Pere-Pau Vázquez [17] presented a straightforward yet potent transformation for vector datasets that alters their values according to weight frequencies. This strategy greatly enhances the efficacy of dimensionality reduction algorithms in many contexts. We analyse a selection of renowned labelled datasets to illustrate the effectiveness of our methods. The results indicate enhanced clustering efficacy in classifying data within the limited space. The idea offers a thorough and flexible strategy to improve the results of dimensionality reduction for visual data analysis. Ali, Mehak, *et al* [18] This paper combines Principal Component Analysis (PCA) with eigenvector integration techniques to present a new approach for dimensionality reduction in time-domain optimisation. Effective dimensionality reduction is increasingly obstructed by data complexity, which is crucial for enhancing computational efficiency and improving model performance. Principal Component Analysis (PCA) is a crucial instrument in machine learning and data processing, particularly advantageous for high-resolution data. This study examines the influence of Principal Component Analysis (PCA) on the efficacy and precision of three classification algorithms: Support Vector Machine (SVM), Random Forest (RF), and Convolutional Neural Network (CNN) in the context of medical image categorisation. Data photos of melanoma and eczema were utilised, with Visual Geometry Group 16 (VGG16) employed for feature extraction, followed by Principal Component Analysis (PCA) for dimensionality reduction. The findings indicate that Principal Component Analysis (PCA) enhances processing speed without significantly impacting accuracy or other performance metrics.

Mwanga, Emmanuel P., *et al* [19] Fourier transform infrared spectrometers, categorised into two distinct age classes. The dimensionality of the spectral data was diminished by unsupervised principal component analysis or t-distributed stochastic neighbour embedding, subsequently employed to train deep learning and conventional machine learning classifiers. The efficacy of transfer learning was assessed to enhance the transferability of models in predicting mosquito age classes from novel populations. Kabir, Md Faisal, Tianjie Chen, and Simone A. Ludwig [20] Examined the effects of various dimensionality reduction strategies on machine learning models employed for cancer prediction. Dimensionality reduction methods, including principal component analysis (PCA), kernel PCA, and autoencoders, were employed to diminish the dimensionality of the RNA sequencing data. Two machine learning classifiers, specifically a neural network and a support vector machine, were trained and evaluated utilising the original, dimensionally reduced, and cancer-relevant datasets. Multiple metrics, including accuracy, precision, recall, F-measure, receiver operating characteristic curve, and area under the curve, were employed to evaluate classifier performance.

Bharadiya, Jasmin Praful [21]Anomaly detection has emerged as an essential technology across various application domains, particularly in network security. This document outlines the categorisation difficulty of anomaly detection utilising machine learning algorithms on network data. The KDD99 dataset is utilised to explore and evaluate dimensionality reduction and classification algorithms for network intrusion detection systems (IDS). Principal Component Analysis for dimensionality reduction and Support Vector Machine for classification have been utilised in the application of network data, and the outcomes have been analysed. The results indicate a reduction in execution time for classification when the dimensionality of the input data is diminished. Additionally, the precision and recall metrics of the classification algorithm demonstrate that the SVM with PCA technique exhibits greater accuracy, evidenced by a decrease in misclassifications. The vast data in health research is highly intriguing, as data-driven studies can progress more rapidly than hypothesis-driven research, despite the increasing prevalence of large databases, which complicates interpretation. Principal Component Analysis (PCA) can be employed to reduce the dimensionality of certain datasets. improves interpretability while preserving the majority of the information. It accomplishes this by introducing novel variables that are independent of each other. Saheed, Yakub Kayode [22] Intended to offer machine learning-based methodologies for the classification of acute myeloid leukaemia and acute lymphoblastic leukaemia utilising microarray gene expression patterns. We utilised logistic regression, very randomised trees classifier, ridge classifier, AdaBoost classifier, linear discriminant analysis, random forest, gradient boosting, and k-nearest neighbours classifier. Principal component analysis was employed for dimensionality reduction. We utilise two separate cross-validation methods in this work as they yield more precise skill evaluations than prior approaches. Six unique performance metrics for categorisation were employed to assess these methodologies.

Pandey, Rajiv, *et al* [23] offered an investigation employing Principal Component investigation (PCA), a prevalent dimensionality reduction technique, to address the dimensionality issue of data. We evaluate the importance of employing PCA to diminish the dimensions of the dataset utilised in an IoMT-enabled system, integrating our research with a previously established framework titled "Prenatal Healthcare System of Remote Mother and Foetal Surveillance via IoMT." The prenatal device enhances the probability of a safe and healthy delivery while mitigating pregnancy hazards. The survival of a foetus relies on regular health evaluations, which are both beneficial and crucial. The information utilised in the experiments comprises essential prenatal device parameters for a foetus, potentially assisting medical professionals with real-time health updates. We employ PCA to emphasise variance and reveal major patterns in the dataset to reliably predict outcomes. Saidulu, D., and R. Sasikala [24] developed a computationally efficient approach for dimensionality reduction and categorisation of healthcare-related data. The developed framework is capable of handling data with both discrete and continuous properties. The experimental assessment is conducted on the Parkinson's disease categorisation database (Sakar *et al*., 2018). The statistical performance metrics employed include validation and test accuracy, precision, recall, F1-score, among others. The decreased dimensional data will confer computational complexity advantages when processed for modelling and constructing prediction systems. To demonstrate the optimality of the proposed framework, a comparative analysis is conducted with notable existing techniques.

Hussein, Safa Saad, *et al* [25] This work examined the efficacy of data dimensionality reduction approaches and machine learning algorithms in enhancing the detection accuracy of cardiac anomalies in WBAN sensors. Dimensionality reduction was executed utilising principal component analysis (PCA), independent component analysis (ICA), and spatial correlation techniques. Decision Tree and Multilayer Perceptron algorithms were employed for arrhythmia prediction, and their performances were compared. Numerical simulations and Python code analysis shown that the implementation of data reduction strategies markedly enhanced the reliability and efficacy of WBAN sensors in managing extensive datasets. Moreover, the implementation of PCA, ICA, and spatial correlation techniques significantly diminished the battery energy consumption of WBAN sensors, as well as the requirements for data storage, computational complexity, and processing duration. These realistic methods may enable healthcare practitioners to react proactively before patients face life-threatening diseases. Karthikeyani, S., S. Sasipriya, and M. Ramkumar [26] This study examined the amalgamation of dimensionality reduction techniques with diverse deep learning classifiers to enhance the precision and efficacy of cardiac illness classification. Uniform Manifold Approximation and Projection, in conjunction with Principal Component Analysis, is employed for dimensionality reduction, effectively capturing both global and local data structures. Classification is performed using deep learning classifiers, including convolutional neural networks, capsule networks, recurrent neural networks, graph neural networks, deep long short-term memory networks, and attention-based convolutional neural networks. The Adaptive Spiral Flying Sparrow Search algorithm optimises classifier parameters to boost accuracy. Performance is assessed using multiple criteria, including the area under the receiver operating characteristic curve, accuracy, F1-Score, precision, and recall.

Kherwa, Pooja, *et al* [27] Conducted an extensive literature review to furnish a comprehensive application-oriented understanding of diverse dimensionality reduction strategies, serving as a reference for selecting the appropriate dimensionality reduction approach to enhance performance across distinct applications. The authors conducted comprehensive tests on two distinct datasets to compare the efficacy of several linear and non-linear dimensionality reduction strategies. PCA, a linear dimensionality reduction method, surpassed all other strategies included in the studies. Indeed, nearly all linear dimensionality reduction methods far surpassed the non-linear strategies on both datasets, exhibiting a substantial margin of error. Dessureault, Jean-Sébastien, and Daniel Massicotte [28] provided a novel approach for selecting the optimal dimensionality reduction technique inside a supervised learning framework. It is also beneficial to eliminate or reconstruct features until the desired resolution is attained. The target resolution may be user-defined or automatically determined by the algorithm. The method employs regression or classification, assesses the outcomes, and provides a diagnosis for the optimal dimensionality reduction technique inside this particular supervised learning framework. The primary algorithms employed are the random forest method, the principal component analysis technique, and the multilayer perceptron neural network algorithm. Six use cases are delineated, each grounded in a recognised approach for generating synthetic data. This research examines each option available in the process, seeking to elucidate the complexities surrounding the overall decision-making process of feature selection or extraction.

Mehrpooya, Adel, *et al* [29] employed matrix factorisation (MF) as a method for high-dimensional reduction in systems pharmacology. We have introduced three innovative feature selection methods based on the mathematical concept of a basis for features. We employed these strategies together with three additional MF methods to analyse eight distinct gene expression datasets in order to examine and compare their efficacy for feature selection. Our findings indicate that these techniques can effectively diminish feature spaces and identify predictive features related to phenotypic determination. The three proposed methodologies surpass the alternative methods employed and can isolate a 2-gene signature indicative of a response to tyrosine kinase inhibitor treatment in the Cancer Cell Line Encyclopaedia. Hernàndez-Carnerero, Àlvar, *et al* [30]Concentrated on forecasting antibiotic resistance in Pseudomonas aeruginosa nosocomial infections within the ICU, employing Long Short-Term Memory (LSTM) artificial neural networks as the predictive approach. The data analysed were sourced from the Electronic Health Records (EHR) of patients admitted to the University Hospital of Fuenlabrada between 2004 and 2019 and were structured as Multivariate Time Series. A data-driven dimensionality reduction method is developed by modifying three feature importance methodologies from the literature to the specific data and presenting an algorithm for determining the optimal number of features. This is accomplished through the sequential capabilities of LSTM, allowing for the consideration of the temporal dimension of features. Additionally, a collection of LSTMs is employed to mitigate performance volatility.

Challenges in the Healthcare Analytics

Despite their advantages, dimensionality reduction techniques in healthcare face several challenges:

Data Quality and Integrity: The effectiveness of dimensionality reduction methods is contingent on the quality of the input data. Incomplete, inconsistent, or noisy data can lead to misleading results, necessitating robust preprocessing steps to ensure data integrity.

Interpretability: While dimensionality reduction can simplify data, it may also complicate interpretability. For instance, the new features generated by PCA may not have clear clinical relevance, making it difficult for healthcare practitioners to derive actionable insights.

Choice of Technique: The choice of dimensionality reduction technique can significantly impact the outcomes of analyses. The effectiveness of different methods may vary based on the specific characteristics of the dataset and the goals of the analysis, necessitating careful evaluation and validation.

Computational Complexity: Some dimensionality reduction techniques, particularly those involving complex algorithms like deep learning-based autoencoders, can be computationally intensive. This complexity can pose challenges in real-time applications or when processing large datasets.

Future Research Direction

As the healthcare landscape continues to evolve, several future directions can enhance the application of dimensionality reduction techniques:

Integration of AI and Machine Learning: The incorporation of advanced machine learning and artificial intelligence techniques into dimensionality reduction processes can lead to improved performance and efficiency. Hybrid models that combine dimensionality reduction with predictive analytics are likely to yield significant insights.

Developing Robust Algorithms: Continued research is needed to develop robust algorithms that can handle diverse data types and maintain interpretability. Techniques that account for missing values and outliers will be particularly beneficial in clinical settings.

Focus on Interpretability and Explainability: As healthcare decisions increasingly rely on data-driven insights, ensuring that dimensionality reduction techniques produce interpretable and explainable results will be crucial. This focus will help bridge the gap between data science and clinical practice.

Interdisciplinary Collaboration: Collaboration between data scientists, healthcare professionals, and domain experts will be essential to effectively apply dimensionality reduction techniques in real-world scenarios. Such interdisciplinary efforts can enhance the development and validation of methods tailored to specific healthcare challenges.

REFERENCES

1. Ayesha, Shaeela, Muhammad Kashif Hanif, and Ramzan Talib. "Performance enhancement of predictive analytics for health informatics using dimensionality reduction techniques and fusion frameworks." *IEEE Access* 10 (2021): 753-769.
2. Kasthuri, S., and A. Nisha Jebaseeli. "An artificial bee colony and pigeon inspired optimization hybrid feature selection algorithm for twitter sentiment analysis." *Journal of Computational and Theoretical Nanoscience* 17.12 (2020): 5378-5385.
3. Kasthuri, S., and A. Nisha Jebaseeli. "Study on social network analysis in data mining." *International Journal of Analytical and Experimental Modal Analysis (IJAEMA),(UGC CARE-A Journal), Impact Factor 6.3* 11.VIII (2019): 111-116.
4. Kasthuri, S., and A. Nisha Jebaseeli. "Review analysis of Twitter sentimental data." *Bioscience Biotechnology Research Communications (BBRC),(UGC CARE Journal-Web of Science), Special Issue* 13.6 (2020): 209-214.
5. Shanti, M. A., and K. Saravanan. "Knowledge data map—A framework for the field of data mining and knowledge discovery." *International Journal of Computer Engineering & Technology* 8.5 (2017): 67-77.
6. Shanti, M. A., and K. Saravanan. "An Effect of Data Mining Techniques in Public Healthcare-A Case Study." *International Journal of Civil Engineering and Technology* 9.9 (2018): 115-122.
7. Shanti, M. A. "A Study to Analyse the Quality of Work Life with Special Reference to Private Sector Bank Employees in Kumbakonam Town of Thanjavur District." *Our Heritage*, vol. 68, 2020.
8. Shanti, M. A. "Earthquake Prediction Using SVM Based Time Predictable Technique." *International Journal of Computer Sciences and Engineering (IJCSE)*, vol. 7, no. 4, 2019, pp. 2347–2693..
9. Poornappriya, T. S., and M. Durairaj. "High relevancy low redundancy vague set based feature selection method for telecom dataset." *Journal of Intelligent & Fuzzy Systems* 37.5 (2019): 6743-6760.
10. Durairaj, M., and T. S. Poornappriya. "Why feature selection in data mining is prominent? A survey." *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications: AISGSC 2019*. Springer International Publishing, 2020.Related works
11. Shanti, M. A., and K. Saravanan. "Wartortle—A Data Mining and Knowledge Discovery Suite for Data Analysis and Reporting." *International Journal of Applied Engineering Research*, vol. 13, no. 10, 2018, pp. 7835–7841.
12. Ambika, G., and P. Srivaramangai. "REVIEW ON SECURITY IN THE INTERNET OF THINGS." *International Journal of Advanced Research in Computer Science* 9.1 (2018).
13. Ambika, G., and D. P. Srivaramangai. "A study on security in the Internet of Things." *Int. J. Sci. Res. Comput. Sci. Eng. Inform. Technol* 5.2 (2017): 12-21.

14. Vinutha, M. R., *et al* "EPCA—enhanced principal component analysis for medical data dimensionality reduction." *SN Computer Science* 4.3 (2023): 243.
15. Ambika, G., and P. Srivaramangai. "Encrypted Query Data Processing in Internet Of Things (IoTs): CryptDB and Trusted DB." (2018).
16. Ambika, G. "Advanced Human Activity Recognition: Leveraging Adaptive Neural Networks and Diverse Machine Learning Algorithms on IoT Data." *Fuzzy Systems and Soft Computing*, vol. 19, no. 02(V), 2024, pp. 11–17.
17. G. Ambika, "IoTCryptDB: Encrypted query data processing in Internet of Things," *International Journal of Scientific Research in Computer Science Applications and Management Studies*, vol. 7, no. 4, p. 18, Jul. 2018.
18. Ali, Mehak, *et al* "Enhanced Dimensionality Reduction in Time-Domain Optimization through PCA and Eigenvector Integration." *Journal of Computing & Biomedical Informatics* 7.02 (2024).
19. Dheepak, T. "Enhancing the Cloud Security with ECC based Key Generation Technique." *Annals of the Romanian Society for Cell Biology* 25.2 (2021): 3874-3891.
20. Malathi, T., and T. Dheepak. "Enhanced Regression Method for Weather Forecasting." *The Scientific Temper*, vol. 15, special issue, 16 Oct. 2024, pp. 146–149. *The Scientific Temper*.
21. Suresh, T., T. Dheepak, and R. Kayalvizhi. "Optimizing QOS in Mobile Ad Hoc Networks Through Advanced Routing Protocols Under Wormhole Attack Scenarios." *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 11, 2023, pp. 584–594.
22. Suresh, T., T. Dheepak, and K. Saraswathi. "Enhancing Sentiment Analysis for Autistic Children: A Hybrid Approach Using SBERT and Ensemble Learning." *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 9, 2023, pp. 4649–4656..
23. Dheepak, T. "Optimizing Routing Protocols in Mobile Adhoc Networks Using Firefly Optimization Algorithm." *Webology*, vol. 18, no. 5, 2021.
24. Dheepak, T. "An Enhanced Access Control Mechanism for Mobile Cloud Computing." *Design Engineering* (2021): 10805-10814.
25. Dheepak, T. "Trust Based Cluster Selection for Intrusion Detection in Mobile Ad Hoc Networks." *Technology*, vol. 11, no. 10, 2020, pp. 421–430.
26. Dheepak, T. "Detection of Attacks in Wireless Networks Using Data Mining Techniques." *International Journal of Management (IJM)*, vol. 10, no. 5, Oct. 2019, pp. 280–288.
27. Kherwa, Pooja, *et al* "Dimension Reduction Techniques in Distributional Semantics: An Application Specific Review." *Data Wrangling: Concepts, Applications and Tools* (2023): 147-185.
28. Dessureault, Jean-Sébastien, and Daniel Massicotte. "DPDR: A novel machine learning method for the Decision Process for Dimensionality Reduction." *SN Computer Science* 5.1 (2023): 124.
29. Mehrpooya, Adel, *et al* "High dimensionality reduction by matrix factorization for systems pharmacology." *Briefings in Bioinformatics* 23.1 (2022): bbab410.
30. Hernàndez-Carnerero, Àlvar, *et al* "Dimensionality reduction and ensemble of LSTMs for antimicrobial resistance prediction." *Artificial intelligence in medicine* 138 (2023): 102508.