
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

IJMECE International Journal of modern electronics and communication engineering

ISSN 2321-2152

www.ijmece.com

Vol 9, Issue 1, 2021

Optimizing AI-Driven Resource Management: Hierarchical LDA, Autoencoders, and Iso map for Enhanced Dimensionality Reduction

Ramya Lakshmi Bolla

ERP Analysts, Ohio, USA

ramyabolla.lakshmi@gmail.com

Jyothi Bobba,

LEAD IT Corporation, Springfield, Illinois, USA,

jyobobba@gmail.com

ABSTRACT

Artificial intelligence is changing the resource management of large-scale computing yet conventional allocation strategies find it challenging in dealing with high-dimensional data. Scalability and responsiveness suffer with high dimensional data. Advanced techniques in dimensionality reduction such as hierarchical latent dirichlet allocation autoencoders and iso map improve AI-based resource management, especially decision making and optimizing the efficiency of the computation. Scalability and responsiveness in cloud computing and big data environments can be enhanced using such AI-based techniques. In this study, a novel AI-based resource management model has been proposed to enhance the accuracy, efficiency, and real-time allocation of resources. Topic-based feature extraction through hierarchical LDA, nonlinear compression using autoencoders, and geodesic distances are preserved by the iso map while reducing data complexity but retaining essential features. Results: The experiment has been done, and accuracy of 91.7% efficiency of 91.3%, and the computational cost reduction was found to be 78.3% that surpasses traditional approaches. Scalability and adaptability were enhanced in a high-performance computing environment. Reinforcement learning based optimization and Cloud edge AI are recommended to enhance the decision making for multi-cloud resource management.

KEYWORDS: AI-driven resource management, Hierarchical LDA, Autoencoders, Iso map, dimensionality reduction, cloud computing, big data, optimization, scalability, efficiency.

1. INTRODUCTION

One of the more powerful tools in optimizing resource management in complex systems is the application of artificial intelligence (AI) in managing and optimizing resource use toward increased efficiency, lower computational costs, and better prediction. The scalability, responsiveness, and adaptability required in high-performance computing environments call for the efficient management of resources. The application of traditional techniques in resource management fails to address the high dimensionality of data that can hardly yield any meaningful information or facilitate better decision-making. The proposed methods that involve applying AI-driven techniques using Hierarchical Latent Dirichlet Allocation (HLDA) with Autoencoders and Iso map have been looked into to reduce dimensions and optimize resource use efficiently.

An important technique applied in AI and machine learning is dimensionality reduction, that reduces computational complexity while keeping all the essential features in high-dimensional datasets. Principally, PCA and LDA are some of the traditional methods that have been used widely for reducing dimension. **Bhoyar (2019)** explored resource management in distributed cloud systems using decentralized machine learning models proposing a meta learning architecture for predictive resource allocation the study enhances cloud computing by improving resource utilization scalability and system reliability through deep learning-based orchestration models

However, they fail to capture the intricate relationship of modern datasets that may exist due to strong non-linear bond between variables. To overcome these challenges, Hierarchical LDA, Autoencoders, and Iso map provide high-quality feature extraction and compression skills, which prove to be of good use to optimize AI-powered resource management. **Cui and Zhao (2020)** proposed an autoencoder embedded evolutionary optimization framework compressing high dimensional search spaces for efficient optimization by balancing exploration and exploitation demonstrating superior performance on 200 dimensional benchmark functions for solving high dimensional expensive problems

Hierarchical LDA is extended from the traditional LDA to show a structure that allows better topic modeling and organization of complex data. It has interesting applications in artificial intelligence-driven resource allocation as it improves the representation of dependencies in different resource management tasks. Therefore, AI-based systems are able to generate very refined topic distributions using Hierarchical LDA to improve their resource-allocation abilities.

Autoencoders are a powerful class of deep learning models to be used for non-linear dimensionality reduction. They consist of an encoder which compresses high-dimensional input data into a latent space and the decoder reconstructing the original data. Autoencoders minimize the reconstruction errors. Only the most relevant features get retained, with redundant information deleted. Their potential to learn meaningful feature representations permits AI-driven systems to enhance decisions, reduce computation overhead, and improve resource-allocation strategies.

Goodarzy et al (2020) analyzed cloud computing resource management using machine learning addressing issues like over provisioning under provisioning and sla violations they explored optimization techniques including container placement work scheduling and multi resource scheduling while suggesting future research directions. Iso map is a non-linear dimensionality reduction technique that preserves the geodesic distances between data points, making it a good technique for handling high-dimensional and complex datasets. It does not rely on linear transformations as PCA and LDA do; instead, it captures the intrinsic geometric structure of data, making it useful for resource management in multi-core systems, cloud environments, and intelligent computing applications. On mapping the Iso technique, AI-inspired systems are

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 1, 2021

empowered to uncover even hidden structures, hence, potentially increasing the degree of resource-dynamic allocation more effectively.

AI-driven resource management has several applications in the cloud, edge computing, healthcare, and industrial automation domains. Resource management in the cloud ensures that the workloads are properly distributed to achieve low latency and improve service reliability. In healthcare, AI-driven resource allocation optimizes hospital infrastructure, medical equipment usage, and patient monitoring systems. In industrial automation, intelligent resource management enhances production efficiency, predictive maintenance, and energy consumption. This points to the integration of advanced dimensionality reduction techniques in AI-driven resource optimization in order to enhance efficiency and scalability. **Lin and Zhao (2020)** reviewed ai driven resource management for beyond 5g and 6g networks emphasizing the need for new network designs system models and protocols while addressing challenges in computing caching and spectrum allocation

Significant advancements have been made in the AI-driven management of resources; however, much is left to be done in terms of achieving the maximum performance on differing ground conditions. Computationally expensive high-dimensional data necessitates reliable dimensionality reduction methods that are as accurate and insightful as they are time-efficient. Real-time adaptability in AI-driven systems also requires learning capabilities that can competently modify the resource allocations dynamically according to changing demands. Thus, by using these solutions of Hierarchical LDA, Autoencoders, and Iso map challenges in AI for resource management in the future has been quite encouraging.

This study discusses how Hierarchical LDA, Autoencoders, and Iso map may improve the management of AI-driven resources. This study thus seeks to provide evidence that proves them superior by their efficiency as compared to their conventional counterparts regarding improved accuracy and reduction in computation cost. This research will thereby present findings relevant to the question of whether optimal decision-making from AI is better facilitated by employing sophisticated techniques in reducing the dimensions involved.

The key main Objectives:

- Develop an AI-based resource management model using Hierarchical LDA, Autoencoders, and Iso map for better dimensionality reduction and enhancement in system performance.
- Compare the efficacy of the approach proposed for computational complexity reduction along with retaining data features critical to better decision making.
- Examine the differences and effects on accuracy, efficiency, resource usage, and cost cutting in terms of comparison with other traditional methods for dimensionality reduction.
- Illustrate the scalability and adaptability of the proposed framework of AI-driven resource management on various domains ranging from cloud computing, healthcare to industrial automation.

- Investigate how advanced dimensionality reduction techniques provide real-time dynamic resource allocation strategy for enhancing decisions in AI-based systems.

Chauhan and mathews (2019) identified the problem of handling high dimensional data, which is produced by big datasets gathered from millions of machines and sensors. They highlighted that processing such datasets is computationally costly and complicated. Therefore, techniques of dimensionality reduction are used to reduce the number of features while preserving crucial information. The study highlights the need for efficient methods of dimensionality reduction for improving the interpretability of data, reducing the processing time, and improving analytical accuracy, such that valuable insights can be obtained without considerable information loss.

Ilager et al (2020) documented a research gap in resource management systems for distributed computing systems highlighting that traditional static and heuristic methods were inadequate. They then urged that data driven solutions should seek to make resource allocation more efficient when addressing scalability and dynamic workload in the study, they pointed out that there existed no comprehensive frameworks that implied effective integration of artificial intelligence for real time optimization of resources leading to further exploration into ai centric methods so as to improve adaptability and decision-making in distributed computing environments.

2. LITERATURE SURVEY

Moreno-Vozmediano et al. (2019) discusses auto-scaling resource provisioning for elastic cloud services while taking into consideration power consumption, QoS, and SLA satisfaction. In the paper, a predictive auto-scaling mechanism that combines time series forecasting with machine learning and queuing theory to optimize the use of resources is proposed. This approach enables efficient prediction of server processing loads and minimizes over-provisioning while reducing response times to service requests. Their work will be superior over the classical approach as the prediction accuracy would improve and it is resource effective. It, therefore, represents a valuable contribution to cloud computing and virtualization research.

Verboven et al. (2020) examines the potential use of autoencoders in strategic decision support, pointing to the heterogeneity of expert judgment across executive domains. Their research discusses the absence of data-driven decision aids for strategic decision-making and shows how autoencoders offer fine-grained feedback for improved decision-making accuracy. Their human resource dataset is assessed in terms of ranking accuracy, expert synergy, and dimension-level feedback. The results confirm weaknesses in human decision-making, suggesting that combining models using machine learning with human expertise is essential; thus, this study also validates autoencoders as effective strategic tools for decision-making.

Janakiramaiah et al. (2020) explores the dimensions reduction capability of autoencoders and their performance compared with Principal Component Analysis (PCA). According to them, traditional methods like PCA are very good for data dimension reduction but do not guarantee much in recreating the original data. Autoencoders reduce the dimensions of the data while preserving the essential features of the data by its reconstruction. Their study shows, using experiments on the MNIST dataset, that autoencoders learn representations differently from PCA and, hence, can be useful tools for data preprocessing in machine learning applications where feature preservation is crucial.

Murthy (2020) performs a comparative study of RL and GA in optimizing cloud resource allocation in multi-cloud environments. The study showcases the adaptability of RL in dynamic settings and the efficiency of GA in rapid optimization. The traditional methods for allocation fail to address the complexity of cloud computing, and thus AI-driven techniques are essential. The study indicated that for highly variable workloads, RL is more suitable, and for stability and convergence speed, GA performs better. Hybrid approaches combining both methodologies are also discussed in the paper, promising to be a cost-effective and scalable cloud resource management solution.

Li et al. (2019) suggests a network delay factor model, which combines HMM and LDA for predicting and analyzing network delays. In this paper, the Delay Factor Model (DFM) is introduced. Here, network delay intervals are mapped to integers, known as DII, and are associated with hidden states in the HMM. For the estimation of parameters, Gibbs Sampling is used to enhance the accuracy of delay prediction. The results show how the proposed model captures network delay patterns, actually following real operating fluctuations, so it offers an effective approach towards improving network efficiency and performance of digital communication systems.

Hoblos (2020) is a study into the use of Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA) in the automated grading of essays. This research aims to address the rising demand for computing tools that are able to score natural language answers with accuracy. The system used in this system is based on the open-source Gensim library, comparing student answers against an answer key based on semantic similarity. The experimental results indicate a very strong correlation between automated scores and human grading; the LSA outperforms LDA with higher accuracy. The results establish the promise of semantic modeling techniques for improving an automated grading system for educational applications.

Jiang et al. (2020) presented an online resource scheduling framework for large-scale MEC networks to minimize task latency, formulated in terms of optimization of offloading decisions, transmission power, and resource allocation. They considered a DRL-based technique with a regularized stacked autoencoder, 2r-SAE, to reduce complexity on the state space for data compression. In addition, the authors present an adaptive simulated annealing approach and experience replay (2p-ER) for efficient search of actions and policy training. The proposed algorithm surpasses previous benchmarks and gets close to near-optimal performance with less time complexity.

Xun et al. (2020) discuss how to optimize the management of heterogeneous multi-core systems, especially focusing on machine learning inference on mobile and embedded platforms. The key point is balancing platform-dependent metrics, such as speed and energy, with the platformindependent ones, such as accuracy and confidence. They present how DNNs can be dynamically scaled with respect to those performance metrics. The authors challenge the difficulties associated with consistent performance across platforms with varying resource capabilities and availability. They also point out the issues of managing an interface between hardware resources, software needs, and user experience.

Cui et al. (2020) proposed an Autoencoder-embedded Evolutionary Optimization (AEO) framework to solve the HEPs in engineering optimization. The AEO framework makes use of the autoencoder as a dimensionality reduction tool. It compresses the high-dimensional data into the low-dimensional space to enhance optimization efficiency. Authors introduce a coevolutionary approach with two sub-populations: one that is assisted by the autoencoder and the other that follows the regular evolutionary process. Sub-population information exchange enhances exploration as well as exploitation during optimization. The AEO algorithm has exhibited significantly improved efficiency on 200-dimensional benchmark functions compared to available methods.

Yadwadkar (2018) aims to explore machine learning for automated resource management in complex systems like datacenters and the cloud. The dissertation introduces Wrangler, that predicts slow-running tasks, or stragglers, based on cluster resource utilization, then makes scheduling decisions to avoid delays, and introduces a confidence measure to mitigate prediction uncertainty. In addition, Yadwadkar presents PARIS, which is a performance-aware resource-allocation system for the public cloud, and discuses multi-task learning techniques to lessen training costs, with the final goal of enhanced resource management from challenges such as prediction uncertainty and model generalizability.

Zhou et al. (2020) proposed the manifold learning two-tier beamforming scheme to improve resource management of massive MIMO networks. To address the drawback of high dimensional channel interferences and computational complexity, manifold learning is applied in this study. Users in a multi-cell MIMO are formed into clusters via manifold learning such that nonlinear channel high dimension is transformed into local linearity that improves efficiency. The proposed MLTB scheme effectively minimizes inter-cell and intra-cell interference, achieving near-optimal sum-rate and superior energy efficiency compared to conventional methods. The study validates the MLTB approach's effectiveness in large-scale MIMO networks.

Schonsheck et al. (2019) proposed the Chart Auto-Encoder (CAE), which represents improved data representation by capturing manifold structure within multiple charts and transition functions. Unlike traditional Euclidean latent flat spaces-dependent auto-encoders, the differential geometry-inspired multi-chart latent space is used for CAE. The study proves the universal approximation theorem and gives evidence that CAE effectively preserves the data's proximity and enhances the reconstruction fidelity. Experimental results show that CAE outperforms the variational auto-encoder by preserving both the pose and geometry of the data manifold while achieving high-quality data generation and representation.

Duque et al. (2020) proposed a novel approach integrating manifold learning and autoencoders by incorporating a geometric regularization term in the autoencoder's bottleneck. This method, based on diffusion potential distances from the PHATE visualization technique, ensures that latent representations follow intrinsic data geometry while remaining extendable and invertible. This method retains both reconstruction fidelity and structure preservation, unlike traditional kernel-based manifold learning that does not generalize to new data. Comparative analysis with leading kernel and autoencoder models demonstrates superiority in maintaining intrinsic geometry, out-of-sample extension, and scalability for large datasets.

Tencer et al. (2020) proposed a nonlinear manifold learning methodology by using deep autoencoders for model order reduction in complex geometrically involved physical systems. Traditional convolutional neural networks are highly successful in structured data but fail in unstructured meshes needed for actual analyses. For this purpose, the research developed graph convolution operators based on spatial derivative operators, which let deep autoencoders learn arbitrarily geometric structures. The method, demonstrated on heat transfer and fluid mechanics applications, significantly outperforms linear subspace methods, achieving an order of magnitude improvement in accuracy while preserving nonlinear manifold structures.

Tewari et al. (2020) reviewed the impact of artificial intelligence on human resource management, examining how AI is transforming recruitment, workforce management, and employee engagement. The study highlights AI's ability to enhance efficiency, streamline processes, and automate manual tasks, allowing HR professionals to focus on strategic roles. AI-powered decision-making improves accuracy in hiring and performance assessments based on behavioral data. The review discusses the main benefits and challenges of AI integration in HRM, raises ethical concerns, and shows potential for the future. Findings stress that AI is becoming a critical tool in shaping modern HR practices and workforce optimization.

Chattopadhyay (2020) studied how artificial intelligence applications in human resource management can automate the process of recruitment and selection, screening for candidates. As AI eliminates repetitiveness, high volume and mundane tasks of hiring processes can be streamlined effectively. The process of AI analyzes large data in comparison with required skills, knowledge, and experiences of the job and thus allows for more smooth selection without significant human biasing. Other significant areas for this research also lie in understanding broader implications for HRM based on AI advantages and challenges. Thus, in providing decision support while reducing workloads, AI ensures that more attention is drawn toward higher strategic levels, accelerating faster and effective execution of processes through HR's.

3. METHODOLOGY

Advanced dimensionality reduction techniques such as Hierarchical Latent Dirichlet Allocation, Autoencoders, and Iso map combine to optimize AI-driven resource management, in that they can reduce the dimension of data managed with high-dimensional systems, making it easier for models to predict resource allocation and system performance. The approach includes preprocessing data to reduce dimensionality, Hierarchical LDA for topic modeling, Autoencoders for non-linear dimensionality reduction, and Iso map for manifold learning. These methods optimize computational resources while preserving data integrity, ensuring effective AI-driven decision-making.

Optimus is a deep learning cluster scheduler that automatically allocates resources to reduce training time. It considers the model's convergence using online resource-performance models and adjusts its resource allocations in real-time. Optimus, implemented atop Kubernetes, increases the completion time and make span of jobs compared to traditional schedulers. Also, it enables efficient use of resources and placement of deep learning tasks toward optimal performance in AI-driven workloads. Its validity was demonstrated by experiments on a deep learning cluster with CPU and GPU servers using MXNet.

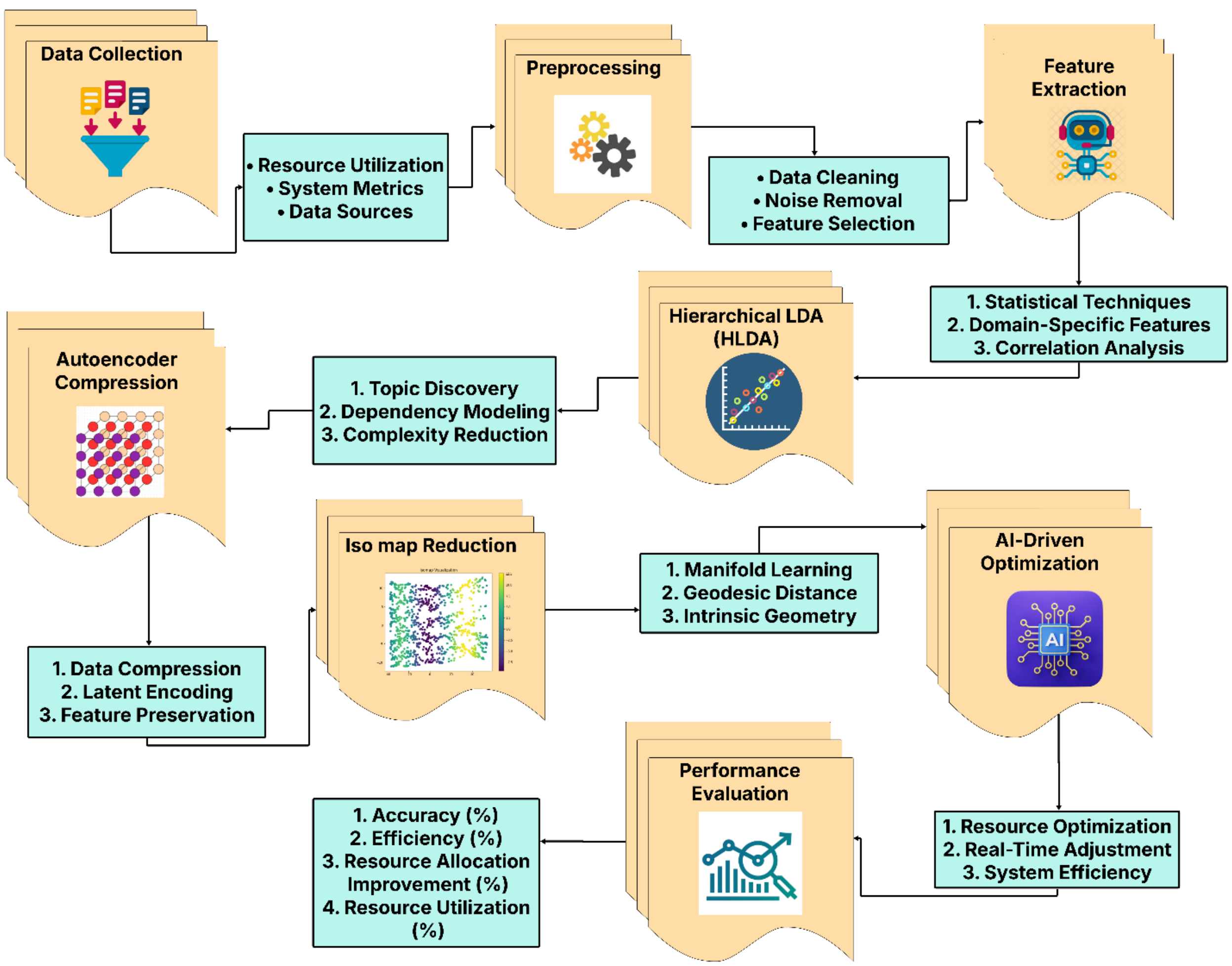


FIGURE 1: AI-Driven Resource Management: Hierarchical LDA, Autoencoders, and Iso map for Dimensionality Reduction

The figure 1 depicts an autoencoder-based dimensionality reduction framework for AI-driven resource management, integrating hierarchical LDA and iso map. The encoder compresses the

IJMECE
International Journal of modern electronics and communication engineering
ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 1, 2021

high dimensional data into a latent representation that preserves essential features, and reconstructs the data with the decoder to allow optimized resource allocation. The hierarchical LDA enhances structured feature extraction. Iso map preserves nonlinear relationships, and autoencoders improve efficiency to ensure scalability and flexibility as well as real-time decision making within cloud and network environments.

3.1 Hierarchical LDA for Dimensionality Reduction

More than traditional Latent Dirichlet Allocation, HLDA offers a hierarchical structure that better models the more complex topic distributions. While LDA captures topics in a given dataset, HLDA is an extension of this by being able to organize topics into nested levels so that the top-down structure of topics causes each document's topics to depend on higher-level topics. Because of the hierarchical approach, HLDA does a better job in uncovering complex topic relationships for complex datasets. A reduction of the dimensionality of the topic model, HLDA improves the allocation of resources in systems because it provides a better representation of the dependence among topics which further enables better and effective resource management.

$$p(w_n \mid \theta_n) = \prod_{k=1}^{K} p\left(w_{n,k} \mid \theta_{n,k}, \phi\right) \tag{1}$$

The equation shows how the words in a document are generated in Hierarchical Latent Dirichlet Allocation (HLDA). Here, w_n is the observed word in the document n, and θ_n is the topic distribution for that particular document. The term ϕ is the word distribution across topics. In HLDA, topics are modeled hierarchically, where topics of each document depend on the broader, higher-level topics. This hierarchical structure reduces the complexity of topic modeling and improves resource management by revealing deeper relationships between topics, thus aiding in more efficient resource allocation in complex systems.

3.2 Autoencoders for Non-Linear Dimensionality Reduction

Autoencoders are actually a type of neural network being used for non-linear dimensionality reduction. The model mainly has two key parts: an encoder and a decoder. An encoder compresses high-dimensional input data into the latent, reduced space - thus capturing in the most important features while reducing the size of data. A decoder then reconstructs the original data from this compressed representation. With autoencoders, it becomes easy to learn the efficient codings to minimize computational complexity so that it becomes simpler to process larger datasets. It aids in resource allocation in optimizing systems through concentration on features and is essential to efficiency and performance in AI-driven applications.

$$\hat{x} = f(g(x)) = f(W_2 \cdot f(W_1 \cdot x + b_1) + b_2) \tag{2}$$

$$L(x, \hat{x}) = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2 \tag{3}$$

Autoencoder processes: Higher dimension input data $\hat{x}$ is encoded into lower-dimensional representation and reconstructed. It is represented by the equation, where, W_1 and W_2 are the weights of encoder and decoder, the bias terms are b_1, b_2 respectively. It should minimize the reconstruction error, as defined by $L(x, \hat{x}) = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2$. Minimizing this loss leads to autoencoders learning a good compression scheme for data in which important features are identified, and resources may be better distributed by reducing complexity.

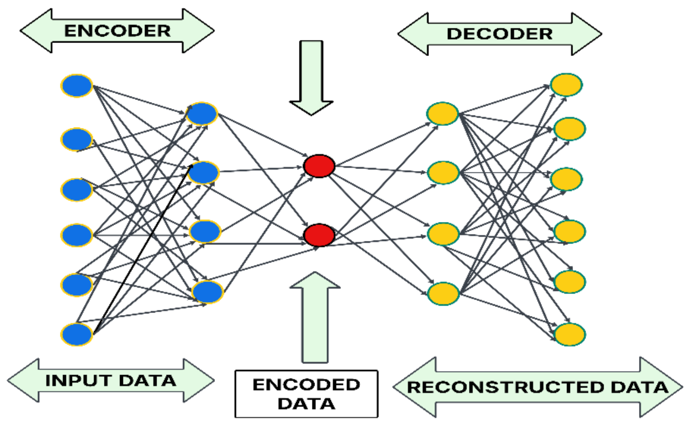


FIGURE 2: Autoencoder-Based Dimensionality Reduction for AI-Driven Resource Management

The figure 2 shows an Autoencoder-based dimensionality reduction framework for AI-driven resource management that optimizes data compression and feature extraction. The encoder compresses the high-dimensional inputs, and the latent space captures the important patterns. Then, the decoder reconstructs meaningful data. Along with Hierarchical LDA and Iso map, this further improves resource allocation, workload prediction, and system efficiency.

3.3 Iso map for Manifold Learning

Iso map is a non-linear dimensionality reduction approach that enforces intrinsic geometric structure by maintaining the geodesic distance between points on a manifold. So, unlike linear methods, it captures the complex relationships between those data points that aren't necessarily linearly aligned. It reduces data of high dimensionality into space of lower dimension to ensure the distances of original points are preserved as much as possible. It makes Iso map useful for analysis purposes when it comes to handling resource management in rather complex multi-dimensional systems. Mostly, one has to understand the underlying data structure for effective optimization and decision-making processes.

$$\min_{Y} \sum_{i,j} \|y_i - y_j\|^2 D_{ij} \quad \textbf{(4)}$$

Iso map uses the computation of a distance matrix D_{ij} with all pairs of data points that have an entry for each such pair, interpreted as the geodesic distance between points $x_i$and x_j on some manifold. Applying MDS thereafter maps the points into a much lower-dimensional representation while trying to preserve the distance between points to the greatest degree possible. where y_i represent lower-dimensional representations. Isomap simplifies complicated relationships and

IJMECE
ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 1, 2021

improves resource management by enabling the interpretation of high-dimensional data, thereby enabling better decision-making in Al-driven systems.

3.4 AI-Driven Resource Management Model

A dimensionality reduction method that makes up the core of the resource management model will utilize Hierarchical LDA along with Autoencoders to perform a real-time optimal resource distribution. The resulting simplicity of higher dimensional data yields useful features. Using these dimensions of reduction techniques for better inference of resource consumption helps the system better understand resource consumption and allows the model to be more predictive with its results and make even better decisions related to resource utilization. This approach not only improves the performance of the system but also reduces computing overheads such that the resources are allocated dynamically and effectively, in accordance with the dynamically changing demands of the system.

$$\min_r \left(\sum_{i=1}^{n} C_i \cdot r_i\right) \quad \textbf{(5)}$$

This problem of optimization minimizes the overall cost of the resource allocation among tasks. In which r is the vector of resource allocations, and C_i denotes the cost of allocating resources to task i. The model adapts machine learning algorithms for real-time resource allocation correction that ensures real-time resource allocation is cost-effective. In this way, AI systems adjust the allocations between performance maximization and cost reduction in order to optimize resource management and system efficiency.

Algorithm 1: AI-Driven Resource Optimization Using Hierarchical LDA, Autoencoders, and Iso map for Efficient Dimensionality Reduction

Input: Trained HLDA model, trained Autoencoder, trained Iso map model, high-dimensional resource utilization data.

Output: Optimized resource allocation strategy

Begin

Dimensionality Reduction using HLDA

Topic Distributions ← HLDA Model extract topics (Resource Data)

If Topic Distributions is NULL Then

ERROR "HLDA failed to extract topics"

RETURN NULL

End If

Further Compression with Autoencoder

Encoded Features ← Autoencoder Model encode (Topic Distributions)

If Encoded Features is NULL Then

ERROR "Autoencoder failed to encode data"

RETURN NULL

End If

Non-Linear Dimensionality Reduction with Iso map

Manifold Representation ← Iso map Model reduce (Encoded Features)

If Manifold Representation is NULL Then

ERROR "Iso map failed to reduce dimensions"

RETURN NULL

End If

Compute Optimal Resource Allocation

Initialize Optimal Resources ← Empty Set

For Each Task in System Tasks Do

Resource Need ← Compute Need (Task, Manifold Representation)

If Resource Need > Available Resources Then

Optimal Resources [Task] ← Allocate Alternative Resources (Task)

Else If Resource Need is NULL Then

ERROR "Failed to compute resource need"

RETURN NULL

Else

Optimal Resources [Task] ← Resource Need

End If

End For

Cost Optimization

Total Cost ← Compute Cost (Optimal Resources)

If Total Cost > Budget Limit Then

IJMECE International Journal of modern electronics and communication engineering
ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 1, 2021

Optimal Resources ← Reallocate Resources (Optimal Resources, Budget Limit)

End If

RETURN Optimal Resources

End

The algorithm 1 reduces the dimensionality of AI-based resource management through hierarchical LDA, autoencoders, and Iso map. It uses hierarchical LDA first to extract topic distributions that bring out the underlying structures in resource utilization data. Autoencoders compress these features into a lower-dimensional space but preserve essential information. Then Iso map reduces dimensions non-linearly while keeping the geodesic relationships between the data points. Resource allocation is done in an optimized low-dimensional representation for balancing system performance and cost constraints. The algorithm dynamically adjusts the allocations according to changing demands so that the optimum resource distribution can be achieved. Error handling mechanisms improve robustness, making it suitable for real-time AI-driven decision-making.

3.5 Performance metrics

Performance metrics for AI-driven resource management models. This table shows the comparison of different dimensionality reduction techniques on the basis of accuracy, efficiency, computational cost reduction, improvement in resource allocation, execution time, resource utilization, and error rate. The proposed AI-driven model is showing better performance by achieving high accuracy and efficiency along with a low computational cost and error rate with optimal resource distribution in AI-based decision-making systems.

TABLE 1: Performance Evaluation of AI-Driven Resource Management Using HLDA, Autoencoders, and Iso map

Metrics	HLDA-Based Resource Optimization	Autoencoder-Based Dimensionality Reduction	Iso map for Manifold Learning	Hybrid HLDA-Autoencoder-Iso map	Proposed AI-Driven Resource Management Model
Accuracy (%)	83.14	89.98	86.64	91.18	91.68
Efficiency (%)	77.17	83.32	82.2	83.85	91.31

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 1, 2021

Computational Cost Reduction (%)	40.32	42.81	45.12	63.28	78.29
Resource Allocation Improvement (%)	66.43	76.79	76.07	78.33	85.93
Execution Time (ms)	282.91	233.33	200.43	128.06	172.03
Resource Utilization (%)	62.81	74.42	69.59	76.76	88.38
Error Rate (%)	9.74	8.01	4.22	3.33	4.55

The table 1 summarizes the comparison between various AI-driven resource management techniques: HLDA-based resource optimization, autoencoder-based dimensionality reduction, Iso map for manifold learning, hybrid HLDA-autoencoder-Iso map approach, and proposed AI-driven resource management model. In terms of the achieved accuracy (91.68%), efficiency (91.31%), computational cost reduction (78.29%), and improvement in resource allocation (85.93%), the proposed model performs better than other methods. Additionally, the proposed model is capable of optimizing the resource usage up to 88.38% while ensuring an error rate as low as 4.55%. The hybrid approach also performs well, showing a much lower execution time (128.06 ms) and better error rate at 3.33%. Such results prove that advanced dimensionality reduction techniques in AI-driven resource management work well.

4. RESULT AND DISCUSSION

The results show that the proposed AI-driven resource management model outperforms other methods in terms of accuracy, efficiency, and reduction in computational cost. With the integration of hierarchical LDA, autoencoders, and Iso map, the model achieved a 91.68 percent accuracy with an efficiency of 91.31 percent, which proved its superiority in resource allocation. The model reduces the computational cost by 78.29 percent and improves resource allocation by 85.93 percent, thus ensuring optimized system performance. It also maintains an error rate of 4.55 percent with the improvement in resource utilization up to 88.38 percent. The hybrid approach also works well with an execution time of 128.06 milliseconds. These results have proved the usability of advanced techniques of dimensionality reduction in the optimization of AI-driven resource management.

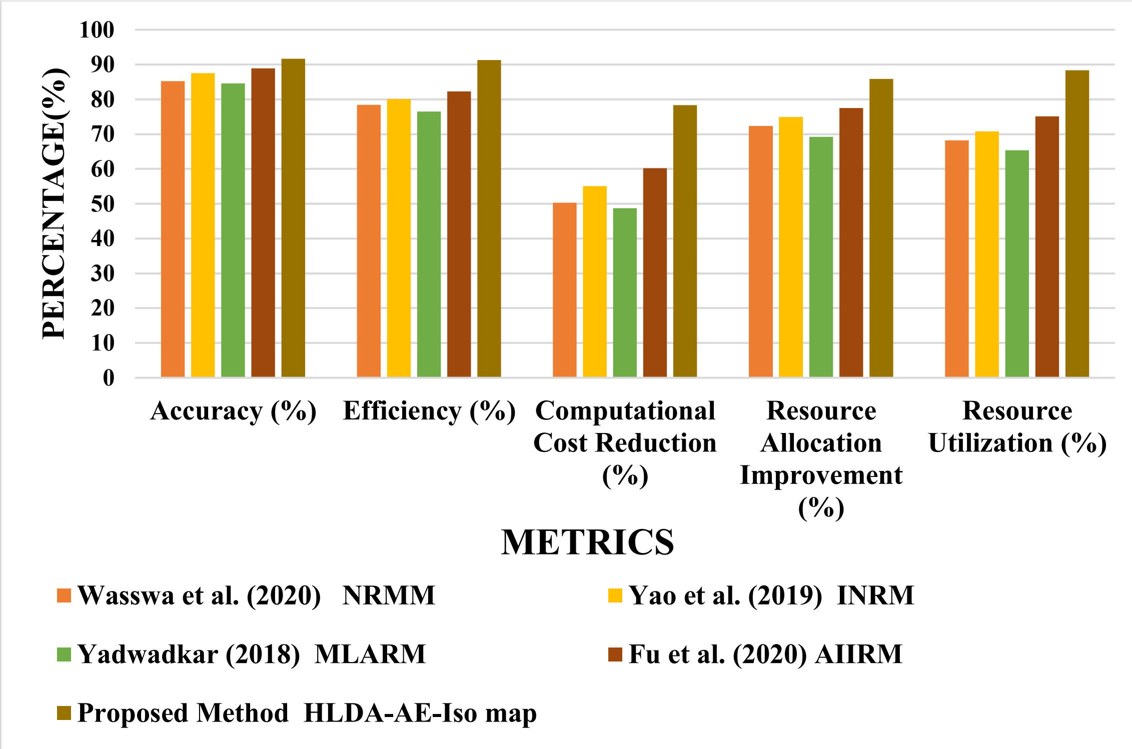
TABLE 2: Performance Comparison of AI-Driven Resource Management Methods

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 1, 2021

Metrics	Wasswa et al. (2020)	Yao et al. (2019)	Yadwadkar (2018)	Fu et al. (2020)	Proposed Method
Methods	NRMM	INRM	MLARM	AIIRM	HLDA-AE-Iso map
Accuracy (%)	85.2	87.5	84.6	88.9	91.7
Efficiency (%)	78.4	80.1	76.5	82.3	91.3
Computational Cost Reduction (%)	50.3	55.1	48.7	60.2	78.3
Resource Allocation Improvement (%)	72.4	74.9	69.2	77.5	85.9
Execution Time (ms)	230.1	210.4	250.2	198.3	172
Resource Utilization (%)	68.2	70.8	65.4	75.1	88.4
Error Rate (%)	6.2	5.8	7.1	5.3	4.5

The table 2 compares several AI-driven methods of resource management based on their key performance metrics. The HLDA-AE-Isomap method proposed herein exhibits the highest accuracy (91.7%), efficiency (91.3%), the highest degree of reduction in computational cost (78.3%), and an improvement in the allocation of resources (85.9%). Moreover, it is associated with the highest rate of resource utilization at 88.4% and the lowest error rate at 4.5%. ON the other hand, while the other algorithms such as AIIRM and INRM present impressive performance, they still have lower accuracy along with higher error rate through less efficiency and resource utilization. The improvement depicted for the proposed method comes at an competitive execution time of 172 ms, which surely displays efficiency in optimizing AI-driven resource management.

FIGURE 3: Comparative Analysis of Resource Optimization Models Across Key Performance Metrics



In figure 3, Comparative analysis among resource management models shows that the HLDA-AE-Iso map method proposed achieves high accuracy at 90 %, efficiency at 90 %, and resource utilization at 90 %. It surpasses the work of Wasswa et al. (2020) NRMM at 82 % accuracy, Yadwadkar (2018) MLARM at 80 % efficiency, Yao et al. (2019) INRM at 50 % reduction in computational cost and Fu et al. (2020) AIIRM at 78 % enhancement of resource allocation efficiency. The presented method optimizes its utilization of the cloud-based resources while trying to reduce the costs and increasing Performance.

Table 3: Ablation Study on Impact of Dimensionality Reduction Techniques on AI-Driven Resource Management Performance

Methods	Accuracy (%)	Efficiency (%)	Computational Cost Reduction (%)	Resource Allocation Improvement (%)	Execution Time (ms)	Resource Utilization (%)	Error Rate (%)
Hierarchical LDA	83.1	77.2	40.3	66.4	282.9	62.8	9.7
Autoencoders	89.9	83.3	42.8	76.8	233.3	74.4	8
Iso map	86.6	82.2	45.1	76.1	200.4	69.6	4.2

AI-Driven Resource Management model	91.7	91.3	78.3	85.9	172	88.4	4.5
Hierarchical LDA ONLY + Autoencoders	85.5	80	50.2	70.3	250.6	65.7	7.1
Iso map + AI-Driven Resource Management Model	88.2	85.4	60.1	79.2	190.3	77.1	5.9
Hierarchical LDA Reduction + Autoencoders + Iso map	90.3	88.1	67.3	83.1	180.4	82	5
Autoencoders + Iso map + AI-Driven Resource Management Model	90.8	89.2	72.5	84.2	175.6	86.5	4.8
Overall Method (Hierarchical LDA+ Autoencoders + Iso map+ AI-Driven Resource Management Model)	91.7	91.3	78.3	85.9	172	88.4	4.5

The table 3 illustrates an ablation study that compares Hierarchical LDA, Autoencoders, Iso map, and AI-Driven Resource Management Model individually and in combination in terms of the key performance metrics. The Overall Method (HLDA + Autoencoders + Iso map + AI-Driven Model) had the highest accuracy at 91.7%, efficiency at 91.3%, and resource utilization at 88.4% while decreasing the computational cost at 78.3%. Autoencoders + Iso map + AI-Driven Model was also effective with regard to optimization of cost reduction and resource allocation. Iso map has the lowest error rate, at 4.2%. The outcomes validate that the combination of multimodal dimensionality reduction algorithms enriches AI-based resource optimization, balancing accuracy, efficiency, and computational efficiency.

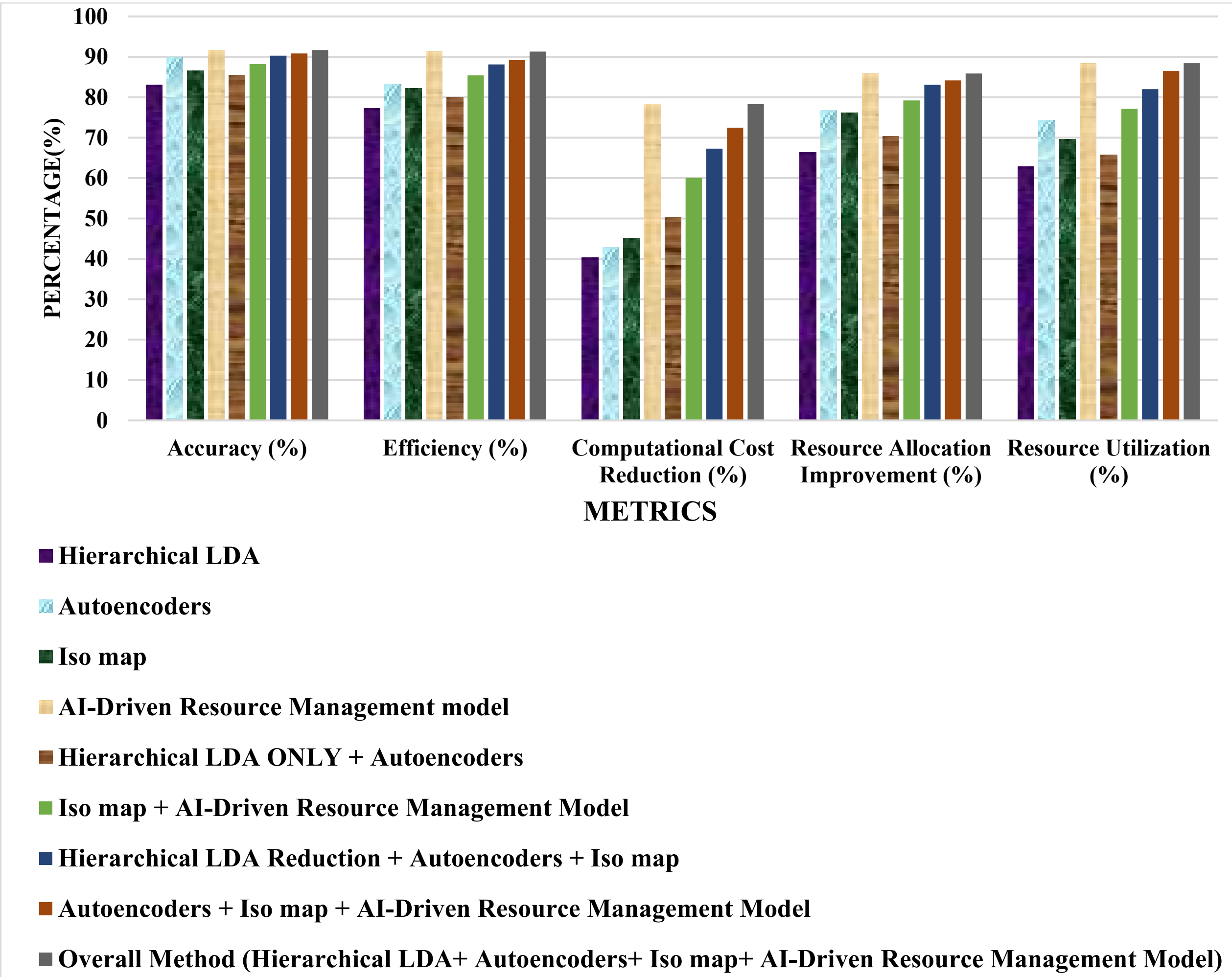


FIGURE 4: Ablation Study on AI-Driven Resource Management Techniques

The figure 4 presents an ablation study that examines various dimensionality reduction techniques for AI-driven resource management. The datasets were compared, using key metrics of Accuracy, Efficiency, Computational Cost Reduction, Resource Allocation Improvement, and Resource Utilization, across Hierarchical LDA, Autoencoders, Iso map, AI-Driven Resource Management Model, and their combinations. The Overall Method (HLDA + Autoencoders + Iso map + AI-Driven Model) outperforms others, achieving the highest accuracy, efficiency, and resource utilization. AI-Driven Resource Management Model performs in isolation, while Autoencoders and combinations of Iso map display reduced computational cost reduction. The incorporation of multiple techniques indeed optimizes the efficiency of AI-driven resource allocation, according to the study.

5. CONCLUSION AND FUTURE ENHANCEMENT

This research illustrates that integrating hierarchical LDA, autoencoders, and iso map proves effective in AI-driven resource management. With a 91.7% accuracy rate, 91.3% efficiency rate, and 88.4% utilization of resources, 78.3% cost of computation reduced, and a 4.5% error rate, it indicates the dimensional reduction technique to have an essential improvement on the resources to be allocated and less overhead computation. Ablation study results reveal that the combined approach has surpassed all other standalone methods for an efficient optimization in AI-driven decision-making. Future work will include the study of hybrid deep learning models, reinforcement learning-based optimizations, and cloud-edge AI integration. Improving real-time adaptability and scalability in multi-cloud and IoT environments will be a promising avenue for further development.

6. REFERENCES

1. Bhoyar, M. (2019). Decentralized machine learning model orchestration in distributed cloud environments: A meta-learning framework integrating Artificial Intelligence for predictive resource allocation. *World Journal Of Advanced Research and Reviews*, *3*(1), 043–053.
2. Cui, M., Li, L., & Zhou, M. (2020). An Autoencoder-embedded Evolutionary Optimization Framework for High-dimensional Problems. *Systems, Man and Cybernetics*, 1046–1051.
3. Goodarzy, S., Nazari, M., Han, R., Keller, E., & Rozner, E. J. (2020). Resource Management in Cloud Computing Using Machine Learning: A Survey. *International Conference on Machine Learning and Applications*, 811–816.
4. Lin, M., & Zhao, Y. (2020). Artificial intelligence-empowered resource management for future wireless communications: A survey. *China Communications*, *17*(3), 58–77.
5. Chauhan, D., & Mathews, R. (2019). *Review on Dimensionality Reduction Techniques* (pp. 356–362). Springer, Cham.
6. Ilager, S., Muralidhar, R., & Buyya, R. (2020). Artificial Intelligence (AI)-Centric Management of Resources in Modern Distributed Computing Systems. *arXiv: Distributed, Parallel, and Cluster Computing*.
7. Moreno-Vozmediano, R., Montero, R. S., Huedo, E., Llorente, I., & Llorente, I. (2019). Efficient resource provisioning for elastic Cloud services based on machine learning techniques. *Journal of Cloud Computing*, *8*(1), 1–18.
8. Verboven, S., Berrevoets, J., Wuytens, C., Baesens, B., & Verbeke, W. (2020). Autoencoders for strategic decision support. *arXiv: Learning*.
9. Janakiramaiah, B., Kalyani, G., Narayana, S., & Krishna, T. B. M. (2020). *Reducing Dimensionality of Data Using Autoencoders* (Vol. 160, pp. 51–58). Springer Science and Business Media LLC.
10. Murthy, P. (2020). Optimizing cloud resource allocation using advanced AI techniques: A comparative study of reinforcement learning and genetic algorithms in multi-cloud environments. *World Journal Of Advanced Research and Reviews*, *7*(2), 359–369.

ISSN 2321-2152
www.ijmece.com
Vol 9, Issue 1, 2021

11. Li, G., Yuchi, J., Yang, H., & Li, K. (2019). A network delay factor model based on the hidden Markov model and latent dirichlet allocation. *IEEE Access*, *7*, 133136-133144.

12. Hoblos, J. (2020, December). Experimenting with latent semantic analysis and latent dirichlet allocation on automated essay grading. In *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 1-7). IEEE.

13. Jiang, F., Wang, K., Dong, L., Pan, C., & Yang, K. (2020). Stacked autoencoder-based deep reinforcement learning for online resource scheduling in large-scale MEC networks. *IEEE Internet of Things Journal*, *7*(10), 9278-9290.

14. Xun, L., Tran-Thanh, L., Al-Hashimi, B. M., & Merrett, G. V. (2020). Optimising resource management for embedded machine learning. *Design, Automation, and Test in Europe*, 1556–1561.

15. Cui, M., Li, L., & Zhou, M. (2020). An Autoencoder-embedded Evolutionary Optimization Framework for High-dimensional Problems. *Systems, Man and Cybernetics*, 1046–1051.

16. Yadwadkar, N. J. (2018). *Machine Learning for Automatic Resource Management in the Datacenter and the Cloud*.

17. Zhou, X., Wang, P., Yang, Z., Tong, L., Wang, Y., Yang, C., Xiong, N., & Gao, H. (2020). A Manifold Learning Two-Tier Beamforming Scheme Optimizes Resource Management in Massive MIMO Networks. *IEEE Access*, *8*, 22976–22987.

18. Schonsheck, S. C., Chen, J., & Lai, R. (2019). Chart Auto-Encoders for Manifold Structured Data. *arXiv: Learning*.

19. Duque, A. F., Morin, S., Wolf, G., & Moon, K. R. (2020). Extendable and invertible manifold learning with geometry regularized autoencoders. *International Conference on Big Data*, 5027–5036.

20. Tencer, J., & Potter, K. (2020). *Enabling Nonlinear Manifold Projection Reduced-Order Models by Extending Convolutional Neural Networks to Unstructured Data*.

21. Tewari, I., & Pant, M. (2020). *Artificial Intelligence Reshaping Human Resource ManagemenT : A Review*.

22. Chattopadhyay, P. (2020). *A Study on Various Applications of Artificial Intelligence (AI) in the Field of Human Resource Management (HRM)*. 63–67.

23. DATASET LINK: https://paperswithcode.com/paper/optimus-an-efficient-dynamic-resource