

Article

Applications of Machine Learning Algorithms in Data Mining for Big Data Analytics

Jieting Lian ^{1,*}

¹ New York University, New York, USA

* Correspondence: Jieting Lian, New York University, New York, USA.

Abstract: This paper explores the integration of machine learning algorithms in data mining for big data analytics, focusing on the role of supervised, unsupervised, and deep learning techniques. It provides an overview of the foundational aspects of data mining in the context of big data and examines various machine learning algorithms that enhance data processing and analysis. Practical applications in key sectors such as healthcare, finance, marketing, and smart cities are discussed, showcasing how machine learning drives innovation and improves decision-making. The paper also addresses challenges like scalability, data privacy, and ethical considerations, and highlights future directions, including algorithm improvements, explainable AI, and edge computing. The conclusion emphasizes the transformative potential of machine learning in advancing big data analytics while ensuring ethical responsibility.

Keywords: machine learning; data mining; big data analytics; supervised learning; unsupervised learning; deep learning; data privacy

1. Introduction

1.1. Background on Big Data Analytics

Big data analytics plays a critical role in decision-making across various industries today. Big data refers to extremely large and complex datasets that traditional tools cannot efficiently handle. It is characterized by four key attributes: Volume, the massive scale of data generated; Velocity, the speed at which data is produced and must be processed; Variety, the diverse formats of data, from structured databases to unstructured content like social media posts; and Veracity, the quality and uncertainty of the data.

Big data analytics enables organizations to process these vast datasets to uncover patterns, trends, and insights that guide strategic decisions. For example, in healthcare, analytics can predict patient outcomes, while in finance, it aids in real-time fraud detection.

Traditional data analysis methods are often insufficient for handling the complexities of big data. This has led to the increasing use of machine learning algorithms, which can analyze and learn from data automatically, identifying valuable insights with minimal human intervention. The integration of machine learning into big data analytics is transforming industries by enabling more accurate predictions and smarter strategies [1].

This paper will explore the applications of machine learning algorithms in data mining for big data, focusing on key algorithms and their practical uses across various sectors.

1.2. Overview of Data Mining in Big Data

Data mining is the process of extracting valuable information from large datasets, a crucial step in big data analytics. It involves identifying patterns, trends, and relationships within the data that can be used for decision-making and predictive analysis. With the explosion of data from sources like social media, IoT devices, and e-commerce, the role of

Published: 18 January 2023



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

data mining has become more critical than ever. It helps businesses and organizations make sense of vast amounts of information that would otherwise be overwhelming.

In the context of big data, the scale, complexity, and diversity of data present unique challenges for traditional data mining methods. Handling such large volumes requires more advanced techniques to process and analyze data efficiently. This is where machine learning algorithms come into play, automating much of the data mining process by learning from the data and continuously improving over time.

Data mining is typically divided into several key tasks, such as classification, clustering, association, and anomaly detection. These techniques help in organizing data, identifying patterns, and predicting future trends. For instance, in customer segmentation, clustering algorithms can group customers based on their purchasing behavior, while classification models can be used for tasks like fraud detection or disease diagnosis.

The value of data mining in big data lies in its ability to provide actionable insights from seemingly unmanageable datasets. Whether it's improving customer experiences, optimizing operational efficiency, or making accurate predictions, data mining is fundamental to unlocking the full potential of big data [2]. In the following sections, we will explore how machine learning algorithms enhance the data mining process, offering solutions to the challenges posed by big data.

1.3. Role of Machine Learning in Data Mining

Machine learning (ML) has become a powerful tool in enhancing data mining processes, especially in the context of big data. Traditional data mining techniques rely heavily on predefined rules and human intervention, making it difficult to handle the scale and complexity of modern datasets. Machine learning, however, enables systems to automatically learn from data and improve their performance over time without explicit programming.

In data mining, machine learning algorithms are used to discover hidden patterns, relationships, and trends within large datasets. These algorithms can handle tasks such as classification, regression, clustering, and anomaly detection, all of which are essential for extracting meaningful insights. For example, supervised learning algorithms like decision trees and support vector machines (SVMs) are commonly used to classify data, while unsupervised learning algorithms like K-means clustering help in grouping similar data points without prior knowledge.

The key advantage of machine learning in data mining lies in its ability to automate the analysis process. With big data, manually analyzing data becomes impractical due to the sheer volume and variety of information. Machine learning models can continuously learn from new data and adapt to changing patterns, providing more accurate and scalable solutions. This makes it particularly useful in industries such as finance, healthcare, and marketing, where timely insights are critical for decision-making.

By integrating machine learning into data mining, organizations can handle the challenges of big data more effectively, making better predictions and discovering insights that were previously inaccessible. This paper will further explore the most widely used machine learning algorithms in data mining and their practical applications in big data analytics [3].

1.4. Purpose and Structure of the Paper

The purpose of this paper is to explore the applications of machine learning algorithms in data mining for big data analytics. As the volume and complexity of data continue to grow, traditional data analysis methods struggle to extract meaningful insights. Machine learning provides advanced techniques that enhance the ability to analyze, predict, and interpret data in real-time, making it invaluable in sectors such as healthcare, finance, marketing, and smart cities.

This paper is structured into four key chapters:

- 1) provides an introduction to big data analytics, an overview of data mining, and the role of machine learning in enhancing data analysis capabilities.
- 2) explores the most commonly used machine learning algorithms in data mining, including supervised, unsupervised, and deep learning methods.
- 3) discusses practical applications of these algorithms across different industries, highlighting real-world examples and case studies.
- 4) addresses the challenges and future trends in the use of machine learning for big data mining, focusing on scalability, privacy concerns, and emerging technologies.
- 5) By the end of this paper, readers will have a clearer understanding of how machine learning algorithms are transforming data mining processes, particularly in managing and analyzing large, complex datasets [4].

2. Machine Learning Algorithms for Data Mining

2.1. Supervised Learning Algorithms

Supervised learning is one of the most widely used machine learning approaches in data mining, particularly when labeled data is available. In this method, the algorithm is trained on a dataset containing input-output pairs (X, y) , where the goal is to learn a mapping function $f: X \rightarrow y$, which predicts the output y based on the input features X . The model continuously improves its predictions by minimizing the loss function $L(y, \hat{y})$, where \hat{y} is the predicted output [5].

Decision Trees: Decision trees split the dataset recursively based on feature values. At each node, a feature x_i is selected to minimize the impurity measure, such as Gini index or entropy H . For binary classification, the entropy at a node is given by:

$$H(p) = -P \log(p) - (1 - P) \log(1 - p)$$

where p is the probability of the class at the node. The goal is to minimize the overall entropy across the tree's splits.

Random Forests: A random forest is an ensemble of N decision trees, where each tree T_i is trained on a random subset of the data. The final prediction \hat{y} for a new instance is the majority vote (classification) or average (regression) of individual tree predictions:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(X)$$

This reduces variance and improves generalization.

Support Vector Machines (SVMs): SVMs aim to find the hyperplane $w^T x + b = 0$ that maximizes the margin between two classes. The margin is maximized by solving the following optimization problem:

$$\begin{aligned} & \min_{w,b} \frac{1}{2} \|w\|^2 \\ & \text{subject to } y_i(w^T x_i + b) \geq 1, \forall i \end{aligned}$$

where w is the weight vector, b is the bias, and x_i, y_i are the input-output pairs.

Supervised learning algorithms play a crucial role in big data mining, enabling tasks like classification, regression, and predictive modeling. For instance, in healthcare, supervised models predict patient outcomes based on historical data, while in finance, they help detect fraud by classifying transactions as legitimate or fraudulent.

2.2. Unsupervised Learning Algorithms

Unsupervised learning algorithms are used when the dataset lacks labeled output, meaning the model must identify patterns and structures on its own. Unlike supervised learning, the goal is to discover hidden relationships or groupings in the data without known labels.

K-means Clustering: The K-means algorithm divides data into K clusters by minimizing the variance within each cluster. The objective function is:

$$\min_C \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

where C_i represents the i -th cluster and μ_i is the centroid of that cluster. The algorithm iteratively updates the centroids μ_i and reassigns points to clusters to minimize the sum of squared distances.

Hierarchical Clustering: Hierarchical clustering builds a tree-like structure of nested clusters. In the agglomerative approach, it starts with each data point as its own cluster and then merges clusters based on a distance metric $d(x_i, x_j)$. Common distance metrics include Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$$

Principal Component Analysis (PCA): PCA is a dimensionality reduction technique that seeks to project data into a lower-dimensional space while retaining most of the variance. Given a dataset $X \in R^{m \times n}$, PCA computes the eigenvectors and eigenvalues of the covariance matrix Σ of X :

$$\Sigma = \frac{1}{m} X^T X$$

The principal components are the eigenvectors corresponding to the largest eigenvalues, representing directions of maximum variance.

2.3. Deep Learning in Big Data Mining

Deep learning, a subset of machine learning, has emerged as a powerful tool in big data mining due to its ability to process and analyze vast amounts of complex data. Unlike traditional machine learning algorithms, deep learning models, particularly neural networks, consist of multiple layers that allow them to learn intricate patterns and representations from raw data. This makes them highly effective in tasks like image recognition, natural language processing, and real-time data analysis.

Key deep learning techniques used in big data mining include:

Convolutional Neural Networks (CNNs): CNNs are widely used for image and video data. They automatically detect spatial hierarchies of patterns through convolutional layers, making them effective in fields such as computer vision, where they are applied in tasks like object detection, facial recognition, and medical imaging analysis [6].

Recurrent Neural Networks (RNNs): RNNs are designed for sequential data and are commonly used in applications like time series forecasting, speech recognition, and natural language processing (NLP). Their ability to retain information from previous inputs allows them to model data where order and context are critical, such as in financial market analysis or predictive text generation.

Autoencoders: Autoencoders are neural networks that aim to compress data into a lower-dimensional form and then reconstruct it, often used for anomaly detection, data denoising, and dimensionality reduction. In big data mining, they are useful for identifying outliers or rare events within large datasets, such as fraud detection in financial transactions or fault detection in industrial systems.

Deep learning excels in big data environments because of its ability to automatically extract features from unstructured data, such as images, text, or audio, without requiring manual feature engineering. This makes deep learning particularly valuable when dealing with the vast, varied, and complex datasets typical of big data.

Furthermore, the scalability of deep learning models allows them to handle the immense volume of data in real-time applications, providing faster and more accurate predictions. As big data continues to grow, deep learning will play an increasingly vital role

in enabling advanced data mining techniques, uncovering deeper insights, and driving innovation across industries.

3. Practical Applications in Big Data Analytics

3.1. Application in Healthcare

Data mining and machine learning have transformed healthcare by enabling data-driven decisions that improve patient care, diagnosis, and efficiency. With vast amounts of data from electronic health records, medical imaging, and wearable devices, machine learning algorithms help analyze this information for real-time insights.

Key applications include:

Disease Prediction and Diagnosis: Machine learning models predict the onset of diseases like cancer or diabetes by analyzing patient data. Algorithms such as SVMs and random forests help detect early signs, while medical image analysis assists in accurate diagnoses [7-9].

Personalized Medicine: Data mining supports personalized treatment plans by evaluating genetic, lifestyle, and medical data to identify the most effective therapies for individual patients, particularly in cancer treatment.

Hospital Management: Predictive analytics optimize hospital resources by forecasting patient admissions and managing emergency room traffic, leading to more efficient operations.

Fraud Detection: Machine learning helps detect anomalies in billing and claims data, reducing fraud and ensuring financial accuracy in healthcare.

By leveraging big data analytics, healthcare providers can enhance patient outcomes, streamline operations, and improve decision-making.

3.2. Application in Finance

The integration of data mining and machine learning in finance has significantly enhanced decision-making processes, risk management, and customer service. With the proliferation of big data, financial institutions are harnessing advanced analytical techniques to gain insights from vast amounts of data, enabling them to remain competitive in a rapidly evolving market.

Key applications include:

Credit Scoring and Risk Assessment: Machine learning algorithms analyze customer data, including credit history, transaction patterns, and social factors, to evaluate creditworthiness. Models such as logistic regression and decision trees improve the accuracy of credit scoring, enabling lenders to make informed decisions and reduce the risk of default.

Fraud Detection: Financial institutions use data mining techniques to identify fraudulent transactions in real time. Machine learning models can detect unusual patterns in transaction data, flagging potentially fraudulent activities before they escalate. This proactive approach helps minimize financial losses and enhance security.

Algorithmic Trading: Machine learning algorithms are increasingly used in algorithmic trading to analyze market trends and make trading decisions at high speeds. By processing large datasets and identifying patterns, these algorithms can execute trades based on market movements, optimizing investment strategies and maximizing returns.

Customer Segmentation and Personalization: Data mining enables financial institutions to segment their customers based on behavior, preferences, and financial needs. Machine learning models can provide personalized product recommendations, improving customer engagement and satisfaction.

Market Risk Management: Advanced analytics allow financial institutions to assess and manage market risks more effectively. Machine learning techniques can analyze historical data to predict market fluctuations, enabling firms to adjust their investment strategies and mitigate potential losses [10].

3.3. Application in Marketing

Data mining and machine learning have revolutionized the marketing landscape by providing valuable insights into consumer behavior, enabling targeted campaigns, and optimizing marketing strategies. By leveraging big data, marketers can analyze vast amounts of information to make data-driven decisions that enhance customer engagement and boost sales.

Key applications include:

Customer Segmentation: Machine learning algorithms, such as k-means clustering, are commonly used for customer segmentation. K-means works by grouping customers into clusters based on their features (e.g., demographics, purchasing behavior), and the goal is to minimize the variance within each cluster. The objective function for k-means clustering is:

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2$$

Where x_j represents a data point, μ_i is the centroid of cluster C_i , and k is the number of clusters. This process helps marketers design targeted campaigns for distinct customer groups.

Predictive Analytics: Predictive analytics in marketing involves forecasting customer behavior, such as purchase likelihood or churn rates. One common approach is using logistic regression for binary outcomes like predicting churn. The probability $P(y = 1)$ (e.g., a customer will churn) is given by:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}}$$

Where X_1, X_2, \dots, X_n are customer attributes (e.g., purchase frequency, demographics), and $\beta_0, \beta_1, \dots, \beta_n$ are model coefficients. This helps marketers forecast customer behavior and improve retention efforts.

A/B Testing and Optimization: A/B testing is widely used in marketing to evaluate the effectiveness of different campaign variations. The conversion rate for each variation is calculated as:

$$CR = \frac{\text{Number of Conversions}}{\text{Total Number of Visitors}}$$

The performance of two variations (A and B) is compared statistically using a z-test:

$$z = \frac{CR_A - CR_B}{\sqrt{SE_A^2 + SE_B^2}}$$

Where SE_A and SE_B are the standard errors of the conversion rates for versions A and B. A significant z-score can help marketers choose the better-performing version.

3.4. Application in Smart Cities and IoT

The integration of data mining and machine learning within smart cities and the Internet of Things (IoT) has led to significant advancements in urban management, resource allocation, and overall quality of life. By harnessing the vast amounts of data generated by connected devices, city planners and administrators can make informed decisions that enhance efficiency, sustainability, and citizen engagement.

Key applications include:

Traffic Management: Data mining algorithms analyze real-time traffic data from sensors and cameras to optimize traffic flow, reduce congestion, and improve public transportation systems. Machine learning models can predict traffic patterns, enabling the implementation of adaptive traffic signal control and efficient routing.

Public Safety: In smart cities, machine learning is employed to enhance public safety through predictive policing and emergency response systems. By analyzing historical crime data, algorithms can identify patterns and predict potential hotspots, allowing law enforcement to allocate resources more effectively.

Energy Management: IoT devices monitor energy consumption patterns in real time, enabling smart grids to optimize energy distribution. Data mining techniques analyze this data to forecast demand, identify inefficiencies, and promote energy conservation, ultimately reducing costs and environmental impact.

Waste Management: Smart waste management systems use data mining to analyze waste generation patterns and optimize collection routes. IoT sensors in waste bins provide real-time data on fill levels, allowing for more efficient scheduling of collection services and reducing operational costs [11].

Citizen Engagement: Machine learning algorithms analyze feedback from citizens through social media, surveys, and mobile applications to gauge public sentiment and identify areas for improvement. This data-driven approach enhances communication between city officials and residents, fostering a collaborative environment for urban development.

4. Challenges and Future Directions

4.1. Scalability and Performance Issues

As data mining and machine learning applications continue to expand in various domains, scalability and performance issues have emerged as significant challenges. With the ever-increasing volume, variety, and velocity of big data, ensuring that algorithms can effectively handle large datasets is crucial for maintaining efficiency and accuracy.

Key challenges include:

Data Volume: The sheer amount of data generated from sources such as IoT devices, social media, and transaction records can overwhelm traditional data processing systems. Ensuring that machine learning models can scale effectively to handle large datasets without compromising performance is a critical concern.

Algorithm Complexity: Many advanced machine learning algorithms, particularly deep learning models, require substantial computational resources and time to train. As the complexity of these models increases, so does the demand for processing power, memory, and storage. This can lead to longer training times and challenges in deploying models in real-time applications.

Distributed Computing: To address scalability, many organizations are adopting distributed computing frameworks such as Apache Hadoop and Apache Spark. However, implementing these systems requires expertise in distributed algorithms and may introduce additional complexity in managing data consistency and synchronization across multiple nodes [12].

Latency Issues: Real-time data analytics is often hindered by latency, particularly when processing large volumes of data. Ensuring that data mining algorithms can provide timely insights without delays is essential for applications in fields such as finance, healthcare, and smart cities, where decisions must be made rapidly.

Data Quality and Integration: As data comes from diverse sources, maintaining data quality and ensuring seamless integration pose significant challenges. Inconsistent, incomplete, or noisy data can adversely affect the performance of machine learning models, leading to inaccurate predictions and insights.

To overcome these challenges, future research and development must focus on enhancing the scalability of algorithms, optimizing computational efficiency, and improving data processing techniques. Innovations in hardware, such as graphics processing units (GPUs) and specialized chips for machine learning, as well as advancements in distributed computing and cloud-based solutions, can help address scalability and performance issues in big data mining. As these technologies evolve, they will enable organizations to harness the full potential of big data analytics and machine learning, paving the way for more effective and efficient applications across various sectors.

4.2. Data Privacy and Ethical Considerations

As the application of data mining and machine learning expands, significant concerns regarding data privacy and ethical implications have emerged. The collection and analysis of vast amounts of personal data raise critical questions about how this information is used, who has access to it, and the potential consequences of its misuse.

Key considerations include:

Data Privacy Regulations: With increasing public awareness of data privacy issues, regulations such as the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the United States have been enacted to protect individuals' personal information. Organizations must ensure compliance with these regulations, which often impose strict guidelines on data collection, storage, and usage. Failure to adhere to these regulations can result in significant legal and financial penalties [13].

Informed Consent: Obtaining informed consent from individuals whose data is being collected is a fundamental ethical consideration. Users should be aware of how their data will be used, stored, and shared. Transparent data practices foster trust between organizations and consumers, but many users remain unaware of the extent of data collection and its implications.

Bias and Discrimination: Machine learning algorithms can inadvertently perpetuate biases present in the training data. If the data reflects historical inequalities or prejudices, the resulting models may produce discriminatory outcomes, affecting marginalized groups. Addressing bias in data collection and algorithm design is essential to ensure fairness and equity in decision-making processes.

Data Security: Protecting sensitive data from breaches and unauthorized access is paramount. Organizations must implement robust security measures to safeguard personal information, as data breaches can lead to severe consequences for individuals and damage the organization's reputation.

Transparency and Accountability: The complexity of machine learning algorithms can make it challenging to understand how decisions are made, leading to calls for greater transparency in AI systems. Organizations should strive to provide explanations for their algorithms' decisions, particularly in high-stakes areas such as healthcare, finance, and criminal justice, where biased or incorrect outcomes can have profound implications.

Moving forward, organizations must prioritize ethical considerations and data privacy in their data mining and machine learning initiatives. Establishing clear ethical guidelines, conducting regular audits, and promoting a culture of responsibility can help ensure that these technologies are used for the benefit of society while minimizing potential risks. By addressing these challenges proactively, organizations can enhance public trust and create a more responsible framework for leveraging big data analytics.

4.3. Future Directions in Machine Learning for Big Data

The evolving landscape of big data presents exciting opportunities for advancements in machine learning. Key future directions include:

Algorithm Design Improvements: Developing more efficient algorithms to process vast datasets quickly and accurately is essential. Research will focus on enhancing existing techniques and creating novel algorithms that improve scalability and performance.

Explainable AI (XAI): As transparency in machine learning becomes crucial, future research will prioritize methods for interpreting model predictions. This will help users understand decision-making processes, fostering trust in AI systems.

Federated Learning: This emerging approach allows models to be trained across decentralized devices while keeping data local, addressing privacy concerns. Future developments will enhance collaboration without compromising individual privacy.

Multi-Modal Data Integration: Future applications will increasingly leverage multi-modal data, combining structured and unstructured sources like text, images, and sensor data for richer insights and improved model robustness.

Real-Time Analytics: The demand for real-time insights necessitates advancements in machine learning techniques that can efficiently process streaming data, enabling swift responses to changing conditions.

Ethics and Governance Frameworks: Establishing ethical guidelines and governance structures for responsible AI development will be crucial in addressing issues like bias and accountability [14].

Edge Computing: As IoT and smart devices proliferate, future machine learning applications will leverage edge computing to enable faster decision-making by processing data closer to its source.

In summary, the future of machine learning in big data is set for transformative advancements that enhance efficiency, transparency, and ethical considerations, unlocking the full potential of machine learning across various sectors.

5. Conclusion

In conclusion, the integration of machine learning algorithms in data mining has revolutionized big data analytics, enabling organizations to extract valuable insights from vast amounts of information. This paper has explored the foundational aspects of data mining, the crucial role of machine learning, and the various algorithms employed, including supervised, unsupervised, and deep learning techniques.

We have also examined practical applications across multiple sectors, such as healthcare, finance, marketing, and smart cities, highlighting how machine learning can drive efficiency, innovation, and improved decision-making. However, the challenges of scalability, data privacy, and ethical considerations must be addressed to ensure responsible and effective deployment.

Looking ahead, the future of machine learning in big data holds significant promise, with ongoing advancements in algorithm design, explainability, and integration of diverse data sources. By embracing these developments and prioritizing ethical governance, organizations can harness the full potential of big data analytics, leading to enhanced operational performance and better outcomes for society as a whole.

References

1. Sara Verde, Federica Dell'Acqua & Mario Losasso. (2024). Environmental Data, Modeling and Digital Simulation for the Evaluation of Climate Adaptation and Mitigation Strategies in the Urban Environment. *Sustainability*(5). doi:10.3390/SU16052179.
2. Wei Jiang, Zhuoxiang Li, Xi Kang, Lei Luo, Yinjie Zhou, Qisong Liu... & Ge He. (2024). Hybrid modeling approach for natural gas desulfurization process: Coupling mechanism and data modeling via compact variable identification. *Gas Science and Engineering*. doi:10.1016/J.JGSC.2024.205243.
3. A. Varadhan, P. Goyal, E. Hohlastos, K. Jenkins, R. Salem & S. Mouli. (2024). Abstract No. 99 Effect of Salvage Prostate Artery Embolization for Recurrent or Refractory Urinary Symptoms After Minimally Invasive Surgical Techniques. *Journal of Vascular and Interventional Radiology*(35). doi:10.1016/J.JVIR.2023.12.128.
4. S. Secco, A. Olivero, F. Sessa, L. Cindolo, P. Chiu, M. Baboudjian... & G. Siena. (2024). A0692 - Multicenter study evaluating the safety and feasibility of Transurethral Intraprostatic Injections of Anesthetics (TUIA) delivered via an innovative catheter to perform minimally invasive surgical techniques for benign prostatic obstructions. *European Urology*(S1). doi:10.1016/S0302-2838(24)01264-8.
5. Sara Verde, Federica Dell'Acqua & Mario Losasso. (2024). Environmental Data, Modeling and Digital Simulation for the Evaluation of Climate Adaptation and Mitigation Strategies in the Urban Environment. *Sustainability*(5). doi:10.3390/SU16052179.
6. Wei Jiang, Zhuoxiang Li, Xi Kang, Lei Luo, Yinjie Zhou, Qisong Liu... & Ge He. (2024). Hybrid modeling approach for natural gas desulfurization process: Coupling mechanism and data modeling via compact variable identification. *Gas Science and Engineering*. doi:10.1016/J.JGSC.2024.205243.
7. A. Varadhan, P. Goyal, E. Hohlastos, K. Jenkins, R. Salem & S. Mouli. (2024). Abstract No. 99 Effect of Salvage Prostate Artery Embolization for Recurrent or Refractory Urinary Symptoms After Minimally Invasive Surgical Techniques. *Journal of Vascular and Interventional Radiology*(35). doi:10.1016/J.JVIR.2023.12.128.

8. Poltavtseva M. A..(2024).Data Modeling in Big Data Systems Including Polystore and Heterogeneous Information Processing Components. *Automatic Control and Computer Sciences*(8). doi:10.3103/S0146411623080266.
9. Dianchao Han,Yongxiang Cao,Fan Yang,Xin Zhang & Min Yang.(2024).Water Quality Estimation Using Gaofen-2 Images Based on UAV Multispectral Data Modeling in Qinba Rugged Terrain Area. *Water*(5). doi:10.3390/W16050732.
10. DeCuir Gunby Jessica T. & Schutz Paul A..(2024).The Evolution of Race-Focused and Race-Reimagined Approaches in Educational Psychology: Future Directions for the Field. *Educational Psychology Review*(1). doi:10.1007/S10648-024-09873-2.
11. Atanu Bhattacharjee,Gajendra K Vishwakarma,Abhipsa Tripathy & Bhritu Kumar Rajbongshi.(2024).Competing risk multi-state censored data modeling by propensity score matching method.. *Scientific reports*(1). doi:10.1038/S41598-024-54149-Y.
12. Yee Lee Cheah & Caroline J Simon.(2024).Surgical techniques for robotic right donor hepatectomy, part 1: robotic hilar dissection and right lobe mobilization.. *Korean journal of transplantation*. doi:10.4285/KJT.23.0059.
13. Yee Lee Cheah & Caroline J Simon.(2024).Surgical techniques for robotic right donor hepatectomy, part 2: robotic parenchymal transection and bile duct division.. *Korean journal of transplantation*. doi:10.4285/KJT.23.0060.
14. Yang Frank Gong & Chun Lai.(2024).Editorial: Teaching and learning Chinese as a foreign or second language: the

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 SOAP and/or the editor(s). SOAP 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.