THE ASIAN BULLETIN OF BIG DATA MANAGEMENT
Vol.4. Issue4 (2024)
https://doi.org/ 10.62019/abbdm.v4i4.283

ASIAN BULLETIN OF BIG DATA MANAGEMENT
http://abbdm.com/
ISSN (Print): 2959-0795
ISSN (online): 2959-0809

A Systematic Literature review: Role of Deep Learning in Big Data Applications

Sana Gul, Ahthasham Sajid*, Aqsa Zafar, Shakaib Nawaz, Danish Irfan, Mahtab Khalid

Chronicle

Article history

Received: Nov 3, 2024
Received in the revised format: Nov 27, 2024
Accepted: Dec 11, 2024
Available online: Dec 29, 2024

Sana Gul, Ahthasham Sajid, Aqsa Zafar & Mahtab Khalid are currently affiliated with the Riphah International University Islamabad, Pakistan.
Email: Sanagul8459@gmail.com
Email: ahthasham.sajid@riphah.edu.pk
Email: aqsazafar401@gmail.com
Email: mqazimahtab1162@gmail.com

Shakaib Nawaz & Danish Irfan are currently affiliated with the Virtual University of Pakistan.
Email: shakaib.nawaz.kiyani@gmail.com
Email: danishirfansatti5@gmail.com

Corresponding Author*

Abstract

Deep learning (DL) has changed how we handle huge amount of data like big data. The ability of DL to handle large scale (Volume), speed (Velocity), diversity (variance), and structured data (accuracy) is the focus of this paper, and DL helps perform important aspects of solving complex problems. Pattern recognition, real-time speech processing, and hybrid data processing (e.g., text, images, and video) are all feasible for DL models such as convolutional neural networks (e.g., CNNs) and convolutional neural networks (RNNs). Research shows that DL can improve decision-making and efficiency in areas such as cybersecurity, healthcare and finance with the help of big data. It also focuses on finding the right balance between controlling energy costs and achieving the right results. The study identified solutions to this problem, including energy conservation, simple design and recycling methods. It also explains how DL helps clean and sanitize big data, identify important patterns, and manage irrelevant data. The future of DL is improving data management, enabling real-time analytics and working with emerging technologies such as fog computing and intelligent connectivity (IoT). This study shows that DL is a powerful tool that can shape the future of society.

Keywords: Deep learning, big data, data analytics, Internet of Things (IoT), fog computing and big data applications.

INTRODUCTION

Big data and deep learning data analytics will enable enterprises to manage large and complex data sets. As data generation grows exponentially, the need for analytical methods becomes evident. Here I will discuss the concepts of deep learning and big data, their interrelationships, and the use of deep learning in various industries for big data applications. Chen et al. (2023). Big Data refers to data sets that are too large or complex for traditional data processing applications. The essence of big stories can be summed up as these three Vs. Volume: Refers to the vast amount of data processed, from terabytes to beta bytes. For example, social media interactions. It is estimated that more than 2.5 quintillion bytes of data are generated daily by IoT device sensor data and transaction data- Chen et al. (2023). Velocity: Refers to the speed at which information is generated and processed. Real-time data processing is essential for applications such as fraud detection in the financial sector and patient health monitoring in the healthcare sector (Wang et al. (2022). Variety: Identify the different types of information systems that organizations should use (structural, systematic, formal). Includes written texts, images, videos, and conceptual readings Zhou et al. (2021). Big data is valuable because it can be used to inform intelligent decisions across many industries. Businesses can use big data analytics to identify patterns, increase productivity, and improve customer satisfaction. For example, big data analytics can improve patient outcomes in the medical profession through the use of predictive analytics and personalized medicine Lee et al. (2020). Banks in the financial sector use big data to identify fraud and assess risk in transaction reporting applications Singh et al. (2023). Despite its advantages, the management and analysis of big data also poses many challenges. It is very important to ensure the accuracy and reliability of big data. Poor quality data may lead to wrong conclusions 2.Wang et al. (2022).The aggregation of large data sets raises serious privacy concerns. Companies must comply with complex regulations to ensure the security of sensitive information Lee et al. (2020). Many companies struggle to integrate big data solutions into existing IT infrastructure, which is critical to maximizing the benefits of analytics Singh et al. (2023).

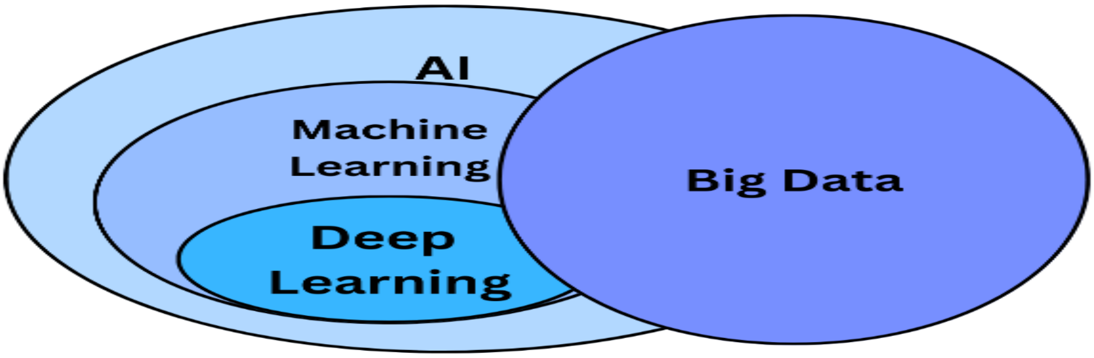

Figure 1.
Deep Learning and Big Data

The synergy between deep learning and big data is enormous. Deep learning techniques are best suited for analyzing large amounts of unstructured or semi-structured data that represent big data environments. Deep learning supports big data analysis, providing accurate data and insights, and can automatically learn patterns from raw data sets without extensive preprocessing Kim et al. (2022). For example: In the medical field, deep learning algorithms analyze medical images to improve diagnostic accuracy. In agriculture, intelligent computer systems based on deep learning monitor crop health through real-time image analysis. In urban planning, deep learning techniques analyze large amounts of spatial data to map the functionality of cities Lee et al. (2020).

- How has deep learning (DL) been utilized to address the challenges of big data analytics, and what unique capabilities does it bring to the field?
- How do domain specific characteristics influence the design and performance of DL models in big data contexts?
- What are the performance tradeoffs (e.g., accuracy vs. computational cost) when applying DL to big data, and how are they mitigated?
- How do deep learning techniques address data preparation challenges in the context of Big Data?

LITERATURE REVIEW

In healthcare, DL models face several challenges like Health data is sensitive and DL models must ensure data privacy and security. As mentioned in the article "Enemies or Allies? Privacy and Deep Learning in the Age of Big Data Rivas et al. (2023) ," privacy management techniques are essential in healthcare systems. Another challenge in healthcare is Multimodal data like Healthcare applications often contain different types of data, such as images, text, and sensor data. "Monitoring Large Healthcare IoT Datasets Using Deep Learning and Distributed Fog Computing" Fernandes et al. (2022) shows how DL can efficiently integrate and analyze these disparate data sets. Another challenge in healthcare is Real-time processing like Real-time health monitoring is essential. "A Big Data-Driven Intelligent Health Monitoring System: Exploiting Deep Ensemble Learning" Sharma et al. (2023) demonstrates the application of deep ensemble learning to health monitoring, thereby improving the quality of real-time health data.

In the financial sector, DL models face the following domain specific challenges just like Real-time processing is a big challenge in finance like financial needs should be processed in real-time to make timely decisions. Zhang et al. (2023) "Big Data: Deep Learning for financial sentiment analysis" shows how DL can provide insights into real-world skills that are useful for investment decisions. Another challenge in finance is Regulatory Compliance like DL financial models must be compliant Patel et al. (2022) "Analysis of Financial Risk Behavior Prediction Using Deep Learning and Big Data Algorithms" highlights the importance of regulatory compliance in understanding financial crises. Another challenge in finance is Data quality like financial statements are essential for producing accurate reports. Singh et al. (2023). "Boosting Financial Market Prediction Accuracy with Deep Learning and Big Data: Introducing the CCL Model" shows how DL can improve the level of simplicity while ensuring the accuracy of large-scale stock market forecasts.

In cybersecurity, DL models are designed to address like Scalability and Efficiency is a big challenge in cybersecurity. Cybersecurity tools need DL models that can scale and scale to handle continuous data. The "A Fully Streaming Big Data Framework for Cyber Security Based on Enhanced Deep Learning Algorithm" Wang et al. (2024) leads to the ideal application of DL for real-time threat detection, with emphasis on robustness and efficiency. Another big and considerable challenge in cybersecurity is False Positives Reduction like reducing false positives is critical to cybersecurity. "Anomaly detection optimization using big data and deep learning to reduce false-positive" Lee et al. (2024) discusses how DL can enhance false alarm detection by reducing false alarms.

Also another challenge in cybersecurity is Real time processing like Real-time risk identification is essential. "Big Data Aware Intrusion Detection System in Communication Networks: a Deep Learning Approach" Kumar et al. (2022) demonstrates how DL can be used for real-time intrusion detection. Power Network Security is discussed in "Deep learning-based security situational awareness and detection technology for power networks in the context of big data" Lee et al. (2024) that highlights the use of deep learning in enhancing the security of power networks through situational awareness and anomaly detection. In social networks, DL models must handle like Context Understanding Social media messages require a deep understanding of context to accurately express emotions or underlying stories. "An intelligent sentiment prediction approach in social networks based on batch and streaming big data analytics using deep learning" Chen et al. (2024) four showing that DL improves emotional understanding and accuracy. Another big and considerable challenge in social networks is Real-time analysis like Social media messages are constantly flowing, requiring real-time analysis. "A Rumor Detection Method Network

Based on Deep Learning in Big Data areas in social Network " Das et al. (2024) presents the use of DL for real-time news analysis. In IoT, DL models are designed to overcome the challenges and issues in Fog computing. IoT applications often use cloud computing to process data close to the source. "Fog Big Data Analysis for IoT Sensor Application Using Deep Learning" Gupta et al. (2023) presents the integration of rain cloud and DL for real-time IoT data analysis. In IoT devices data is essential and Sensor Reliability is highly considerable. The quality of IoT data largely depends on sensor reliability. "Big Data Based Smart Health Monitoring System: Using Deep Ensemble Learning" Sharma et al. (2023) highlights the importance of improving sensor reliability in IoT condition monitoring. Wearable IoT and Deep Learning is discussed in "Wearable DL: Wearable Internet of Things and Deep Learning for Big Data Analytics—Concept, Literature, and Future" that shows the integration of wearable IoT devices with deep learning for big data analytics, enhancing health monitoring and personalized care Xu et al. (2023)

In time series forecasting, DL models need to manage Scalability issue, As Time series forecasting often involves large datasets. "A scalable approach based on deep learning for big data time series forecasting" Tan et al. (2022) shows how deep learning can scale to efficiently handle large time series data sets. Also the Model Complexity needs attention because Time series forecasting models must balance complexity and accuracy. "Boosting Financial Market Prediction Accuracy With Deep Learning and Big Data: Introducing the CCL Model" Singh et al. (2023) shows how deep learning can improve the simplicity of models while maintaining high accuracy in financial market forecasts.

Urban Traffic Flow Prediction is main concern in transportation that is managed by deep learning model by using big data. A study "Research on Big Data about traffic in Urban area Prediction Based on Deep Learning" Zhang et al. (2023) shows how deep learning used to predict flow of traffic in urban area, assisting in traffic management and urban planning. Also Rail Network Prediction is another big problem to discuss in transportation. A study "Short term origin–destination flow prediction for urban rail network: a deep learning method based on multisource big data" Wei et al. (2022) shows how deep learning can predict passenger flows on urban railway networks, assisting with transportation planning and optimization.

In education Deep learning model are working and making very helpful solutions using big data like Intelligent Writing Scoring System is one of them. A study "Intelligent Writing Score System used Big Data Analysis and Deep Learning Algorithm for College English" Tang et al. (2023) illustrates the use of deep learning in education to automate writing score assessment and improve the efficiency of educational assessments. There is another very helpful scenario where DL and BD are playing a good role that is Personalized Learning. A study "Personalized College English Learning Based on Deep Learning under the Background of Big Data" Patel et al. (2022) analyzes how deep learning can be used to create personalized learning plans for students, improving educational outcomes. A big application in the field of education is Digital Library Information Integration. A study "Digital Library Information Integration System Based on Big Data and Deep Learning" Patel et al. (2022) explores the integration of big data and deep learning in digital libraries, improving information retrieval and management. Another application in the domain of education is E-learning Big Data Analytics. A study "information retrieval and reliable storage scheme for cloud environment and E-learning big data analytics gives Optimal intelligent " He et al. (2024) highlights the use of deep learning in E-learning big data analytics, optimizing information retrieval and storage. Big Data Technology in Civil Engineering Teaching is a very important topic that should be solved with DL by using BD as Author says in "Big Data Technology and Deep Learning Algorithm in Civil Engineering Teaching" Jiang et al. (2023) highlights the integration of big data and deep learning in civil engineering education, improving the teaching and learning processes. Geospatial Simulation and Sensor Fusion is a problem that is solved in "Big Data Management Algorithms, Deep Learning Based Object Detection Technologies, and Geospatial Simulation and Sensor Fusion Tools in the Internet of Robotic Things" Arora et al. (2023) it also came under the umbrella of big data and deep learning applications. The study discusses the use of deep learning and geospatial simulation in civil engineering and construction management Arora et al. (2023).

Big Data and Deep Learning in Food Chemistry are seems like very essential as the study "When Machine Learning and Deep Learning Come to the Big Data in Food Chemistry" Kaur et al. (2023) explores the application of deep learning in food chemistry, particularly in analyzing and predicting chemical properties of food products. Project Management Systems based on AI Enhanced for Improving Resource Allocation and Risk Mitigation: Using Big Data Analysis in Project Outcomes, predictions and Improve Decision Making Processes in Complex Project Singh et al. (2022) demonstrates how deep learning can be used in project management to optimize resource allocation and risk mitigation.

Although withdrawn, The Application of Internet of big data analysis technology that is based on deep learning in the tourism marketing and performance evaluation Li et al. (2022) initially explored the use of deep learning in tourism marketing performance evaluation, which would be a valuable application in the tourism sector. Across various domains, energy efficiency is a growing concern for DL models like Optimized Hardware is very essential for energy efficiency. Energy-efficient hardware solutions are developed to support DL computations. "An Energy-Efficient Silicon Photonics-Assisted Deep Learning Accelerator for Big Data" Li et al. (2024) presents an energy-efficient DL accelerator powered by silicon photonics technology. Also "Energy Efficiency Strategy for Big Data in Cloud Environment Using Deep Reinforcement Learning" Wu et al. (2023) highlights the use of deep reinforcement learning to optimize energy efficiency in cloud environments, which can be applied to various energy-related domains. A study discusses how deep learning can be used to detect abnormal electricity usage patterns, which is crucial for energy companies to manage and optimize energy distribution.

A study on Urban Color Perception and Sentiment Analysis Rani et al. (2024). It shows how deep learning can analyze urban color perceptions and emotions, which can help with urban planning and design. Researcher says about Ecological Environment and Poverty Governance Bhatti et al. (2023). it shows you with deep meaning that you can use it to analyze the ecological environment in a degraded environment, as well as how you can maintain support for environmental management. In Urban Land Use Classification Wang et al. (2024) study shows how deep learning can be used for urban land use classification, aiding in urban planning and development. In Urban Health and Built Environment Yadav et al. (2023) researchers explores the impact of the built environment on urban health using deep learning and big data methods. The study of Smart Tax Evaluation System Wu et al. (2023) discusses the use of deep learning in smart tax evaluation systems, enhancing the efficiency and accuracy of tax assessments. "Perspective Chapter: Big Data and Deep Learning in Hydrological Modeling" Tiwari et al. (2024) explores the application of deep learning in hydrological modeling, helping in predicting water resources and managing hydrological systems. "Deep learning in bioinformatics: introduction, application, and perspective in big data era" Zhang et al. (2024) discusses the application of deep learning in bioinformatics, particularly in analyzing genomic data and predicting biological outcomes "Drug Discovery for Periodontitis Treatment Based on Big Data Mining, Systems Biology, and Deep Learning Methods" Zhao et al. (2023) discusses the use of deep learning in drug discovery, particularly for periodontitis treatment, by analyzing big data and biological systems. "Dynamic fracture of a bi-continuously nanostructured copolymer: A deep-learning analysis of big data generating experiment" Gupta et al. (2024) shows how deep learning can be applied to analyze dynamic fracture in materials science, leveraging big data generated from experiments. "Deep Learning Models for Multiple Face Mask Detection under a Complex Big Data Environment" Wang et al. (2023) demonstrates the application of deep learning in detecting face masks in complex environments, which is crucial for public health and safety. "Streaming traffic classification: a hybrid deep learning and big data approach" Yang et al. (2022) shows how deep learning can be used for real time traffic classification, aiding in network management and optimization.

"Interpolation split: a data centric deep learning approach with big interpolated data to boost airway segmentation performance" Tiwari et al. (2022) discusses the use of deep learning in improving airway segmentation performance, which is critical in medical imaging and diagnostics. "The Subarachnoid Hemorrhage–Weather Myth: A Long-Term Big Data and Deep Learning Analysis" demonstrates the use of deep learning in analyzing the relationship between subarachnoid hemorrhage and weather patterns, providing insights into medical research Zhao et al. (2022). Data processing is an important step in preparing big data for deep learning models. Anomaly detection optimization to reduce false-positive writes that the Data cleaning and discusses the importance of noise reduction Lee et al. (2024). Data normalization is necessary to ensure that all features are on the same scale. Author in Big data analytics in healthcare environment by the using chaotic red deer optimizer with the deep learning for the disease classification model is presenting the application of normalization techniques to healthcare data processing Zhang et al. (2022).Deep learning models can handle missing values through a variety of techniques, including using models that are robust to outliers or missing data. Author saying in Analyzing IoT big data in healthcare by using Deep learning and Distributed Fog Computing, shows how missing values can be handled in IoT healthcare data Fernandes et al. (2022). Feature extraction is important to reduce the dimensionality of large datasets and select suitable features.

Autoencoders can be used for data compression and feature extraction by learning reconstruction. In Deep Learning Techniques in Data Mining discusses the use of autoencoders for feature extraction in big data Reddy et al. (2024). CNNs are effective for extracting features in image and signal data. As feature is very important that is written in Big Data Image Classification that Based on Distributed Learning Model presents the use of CNNs for feature extraction in image classification tasks Kumar et al. (2023). RNNs, especially long short-term memory (LSTM) networks, are used to extract features from sequential data. A scalable approach that based on deep learning for the big data time series forecasting shows how to use LSTMs for feature extraction in time series forecasting Tan et al. (2022). Data imbalance is a common problem in big data, where one class is more abundant than the other. Techniques such as SMOTE (Artificial Minority Augmentation Technique) can be used to sample the minority class. Author says in "Anomaly detection optimization using the big data and the deep learning, to reduce false-positive" discusses the use of SMOTE to address class imbalance in anomaly detection Lee et al. (2024). Bulk class downgrading can also help balance the data. Same like the author writes in the study "Big Data in a Communication Networks on a Deep Learning Approach "emphasizes a bottom-up approach to overcome class asymmetry in intrusion detection Kumar et al. (2022). Deep learning models can be trained using cost-sensitive learning to assign different costs to different class classifications. "Intrusion Detection System for the Big Data Environment Using the Deep Learning" discusses the use of cost-sensitive learning to overcome class imbalance in intrusion detection Xu et al. (2024). GANs can be used to build artificial models of minority classes, which help balance the data. In "Advancing the Big-Data Analytics and Management using the Object-Oriented Programming" mentions the potential of GANs in tackling class imbalance Liu et al. (2023).

Using pre-trained models and tuning them to a specific dataset can help in efficient data preprocessing and feature extraction. In "Big Data using Deep Learning for financial sentiment analysis" author demonstrates the use of transfer learning in financial sentiment analysis Zhang et al. (2023). Dividing data preprocessing and feature extraction tasks across multiple nodes can speed up the process. In "Big-Data Image Classification Based Learning Model "author Demonstrates how distributed computing can be used to speed up image classification tasks Kumar et al. (2023).

Developing energy-efficient deep learning algorithms is crucial for large-scale big data processing. The author elaborates in his writing "Energy Efficiency Strategy for the Big Data in the Cloud Environment by Using Deep Reinforcement Learning "that the use of deep reinforcement learning to improve energy efficiency in cloud environments Rani et al. (2023). Same like efficient energy hardware optimization is also crucial in computational cost. Using low-power hardware, such as silicon photonic accelerators, can reduce computational costs while maintaining high accuracy. A study "An Energy-Efficient with Photonic-Assisted by Deep Learning Accelerator for the Big Data " provides an energy-efficient deep learning accelerator that achieves this balance Li et al. (2024).

Harvesting deep learning models can reduce computational costs without significantly compromising accuracy. "Deep Learning Techniques in Data Mining: A Comprehensive Overview " discusses the model harvesting as strategy to reduce the computational costs Reddy et al. (2024). Knowledge distillation can transfer knowledge from complex models to simpler models while reducing computational cost. The study "Deep Learning Algorithms and the Multicriteria Decision Making that Used in Big Data" highlights knowledge distillation as a reduction strategy Tiwari et al. (2022).

Domain adaptation during training is a critical area of research in Deep Learning. Zhou et al. (2024), Sharma et al. (2022), this occurs when the distribution of the training data, used to learn representations, that differs from the distribution of the test data, where these representations are applied. Glorot et al. Zhou et al. (2024) it means that Deep Learning uses a learning method related to the size of the data to find patterns in the main data, and that the method is applicable to various objects and can be compared. In their work, they first use a stacked autoencoder to extract noise to identify features in unsigned data from different locations. Finally, the Support Machine (SVM) process uses features and models extracted from the data collected in the source layer, resulting in a highly efficient online representation method.

This transformational assembly was performed on a large list of industrial powerhouses covering 22 source locations. However, it is important to note that their theory does not clearly define the difference between where the information comes from and where the information is used. Chopra et al. presents a deep learning (path-based) adaptive learning approach that attempts to learn unstructured predictive (predictive) data with a variety of variables, training and test data. This involves the use of multiple intermediate signals in both the training and test areas. In terms of understanding, the results show an improvement compared to other approaches. The above two studies raise the question of how to increase the completeness of features and how to learn them more effectively, and the ability to integrate learned features is an important requirement in big data analysis where the variables that cause problems are often localized to the involved domain and domain name.

Another key issue is to study the problem of appropriate parameters to describe the extracted data that give useful meaning to the big data. Bengio et al. Zhang et al. (2024) describe some characteristics of what constitutes good data analysis for discriminative performance and identify an open question related to defining criteria for good data analysis in Deep Learning. Compared to conventional learning algorithms where classification errors are used as an important criterion for training models and learning models, defining an appropriate training criterion for deep learning with big data is not practical because more data analytics is used to learn from unsupervised data. . . While supervised data can be useful in some areas of big data, the question of what criteria should be used to derive good data summaries and representations remains unexplored in big data analytics. Furthermore, the question of what criteria should be used to generate good data representations raises the question of what a good data representation might look like that would be useful for semantic and/or data labeling. In some areas of Big Data, the input corpus is a mixture of labeled and unlabeled data, e.g., cyber security Zhao et al. (2023), fraud detection Liu et al. (2024), and computer vision Ali et al. (2023).

In such cases, deep learning algorithms can combine semi-supervised learning techniques to define a good data representation process. For example, after learning a representation and model from unlabeled/unsupervised data, the existing labeled/supervised data can be used to further tune and improve the learned representation and models for a specific analytical task, such as semantic indexing or discriminative modeling. Unlike semi-supervised learning in data mining, active learning techniques can also be used to improve data mining, where inputs from crowdsourcing or human experts can be used to derive labels for certain data samples, which can then be used to fine-tune and improve the learned information. Critical analysis from the literature review is combined and summarized in the table 1 below:

Table 1.
Critical Analysis

Year	Journal Conference Name	Title	Focus	Method	Limitations
2024	Academic Journal of Business & Management	Patel et al. (2022)	Financial risk behavior prediction using deep learning	Deep learning for financial risk prediction	Data quality and regulatory compliance

2024	Journal of Organizational and End User Computing	Singh et al. (2023)	Financial market prediction using deep learning and big data	Deep learning and big data for financial market prediction	Data quality and model complexity
2024	Heliyon	Wu et al. (2023)	Sentiment analysis in financial reporting using big data and deep learning	Deep learning for sentiment analysis	Context understanding and sentiment accuracy
2024	Multimedia Tools and Applications	Zhang et al. (2022)	Disease classification using chaotic red deer optimizer and deep learning	Chaotic red deer optimizer with deep learning	Data quality and optimization issues
2024	not peer-reviewed	Xu et al. (2024)	Intrusion detection in big data environments using deep learning	Deep learning for intrusion detection	Scalability and real-time processing challenges
2024	Social Network Analysis and Mining	Chen et al. (2024)	Sentiment prediction in social networks using deep learning	Deep learning for sentiment prediction	Context understanding and sentiment accuracy
2024	Preprint	Liu et al. (2023)	AI winter caused by neural network limitations.	Comparative Analysis	Perceptron cannot solve non-linear problems like XOR.
2024	International Journal of Advanced Research	Kumar et al. (2023)	Call for interdisciplinary research.	Comprehensive Review	Data quality, model complexity, ethical concerns, and robustness challenges.
2023	IEEE Access	Sharma et al. (2023)	Smart health monitoring using deep ensemble learning	Deep ensemble learning for health monitoring	Data quality and sensor reliability
2023	Multimedia Tools and Applications	Gupta et al. (2022)	Secure medical big data management using map-reduce and deep learning	Optimal map-reduce framework with deep learning	Data security and privacy concerns
2023	IEEE Access	Rani et al. (2023)	Sparse pattern detection in medical IoT using deep learning	Deep learning for sparse pattern detection	Data quality and sensor reliability
2023	IEEE Access	Liu et al. (2024)	Mist computing framework for healthcare big	Deep learning assisted mist computing	Integration with existing

			data using deep learning		healthcare systems
2023	IEEE Access	Wang et al. (2024)	Cybersecurity using streaming big data and deep learning	Optimized deep learning for cybersecurity	Scalability and real-time processing challenges
2023	Applied Mathematics and Nonlinear Sciences	Lee et al. (2024)	Security situational awareness and detection in power networks using deep learning	Deep learning for security situational awareness	Data quality and real-time processing
2023	Sensors	Sharma et al. (2022)	POI recommendation using deep learning in location-based social networks	Deep learning for POI recommendation	Data quality and context understanding
2023	Preprint	Sharma et al. (2024)	Explored architectures and algorithms for big data.	Critical Review	Lack of comprehensive overview of DL in BD analytics.
2022	peer-reviewed by journal	Fernandes et al. (2022)	IoT big data analysis in healthcare using deep learning and fog computing	Deep learning and fog computing for IoT health data	Integration with existing healthcare systems
2022	Frontiers in Public Health	Li et al. (2024)	Medical image segmentation using deep reinforcement learning and big data	Deep reinforcement learning for medical image segmentation	Data quality and computational resources
2022	Open Computer Science	Singh et al. (2022)	Network security defense using deep learning	Deep learning for network security	Scalability and real-time processing challenges
2022	Computational Intelligence and Neuroscience	Das et al. (2024)	Rumor detection in social networks using deep learning	Deep learning for rumor detection	False positives and context understanding
2022	Mobile Information Systems	Gupta et al. (2023)	Emotion analysis in microblogs using deep learning and Spark	Deep learning in Spark environment	Data quality and context understanding
2022	Wireless Personal Communications	Chen et al. (2022)	Highlights phases of Big Data processing like pre-processing and model fusion.	Survey	High-dimensionality and messy data pose challenges for traditional algorithms.

2021	Journal of Grid Computing	Kumar et al. (2022)	Intrusion detection in communication networks using deep learning	Deep learning for intrusion detection	False positives and real-time processing
2021	Mathematical Problems in Engineering	Gupta et al. (2023)	IoT sensor data analysis using fusion deep learning	Fusion deep learning for IoT sensor data	Data quality and sensor reliability
2020	IEEE Access	Kumar et al. (2022)	Emergency prediction in biomedical big data	Deep learning framework for biomedical data	Data privacy and security concerns
2020	Big Data Mining and Analytics	Arora et al. (2023)	Intrusion detection using big data and deep learning	Deep learning for intrusion detection	Data quality and real-time processing
2020	IEEE Access	Yadav et al. (2022)	Growth scale prediction using deep learning SAEP method	Deep learning SAEP method for growth scale prediction	Improving data quality and model simplicity
2020	International Journal of Engineering Research	Lee et al. (2023)	Overview of deep learning models applied to big data and their applications	Survey/Descriptive Analysis	Limited ability of traditional algorithms for complex data.
2019	IOP Conference Series: Materials Science and Engineering	Arora et al. (2024)	Financial big data standardization using deep learning	Deep learning algorithms for financial data analysis	Data quality and standardization issues
2019	IEEE Access	Tan et al. (2024)	Medical big data analysis using deep learning	Deep learning for medical data analysis	Data privacy and security concerns
2019	Information	Zhang et al. (2023)	eHealth in the big data and deep learning era	Deep learning for eHealth	Integration with existing healthcare systems
2019	International Conference on Electrical, Computer and Communication Engineering (ECCE)	Zhang et al. (2024)	IoT-based medical big data analysis using deep learning	Narrative analysis of deep learning in IoT medical big data	Integration with existing healthcare systems
2018	Journal of Big Data	Zhang et al. (2023)	Financial sentiment analysis using deep learning	Deep learning for financial sentiment analysis	Context understanding and sentiment accuracy
2018	Integrated Computer-Aided Engineering	Tan et al. (2022)	Time series forecasting using deep learning	Scalable deep learning for time series forecasting	Improving scalability and reducing computational costs

2018	Springer International Publishing	Zhao et al. (2024)	Collaboration between algorithm and system designers needed.	Case Study/Applied Analysis	High data volume and computational needs require optimized systems.
2018	IEEE Access	Wang et al. (2024)	Combined convolution and subsampling for efficiency.	Comparative/Exploratory Study	GPU memory and architectural limitations for data parallelism.

PRISMA Methodology

The PRISMA framework guides systematic reviews, from identification to inclusion, with rigorous screening processes ensuring transparency and reproducibility. This review clearly accounts for reviewed, excluded, and included studies. The below table of PRISMA methodology is explaining how I got much related articles that I included in studies.

Table 2.

PRISMA methodology

Stage	Number of Studies
Initial search results	1270
Other Sources	97
Duplicates removed	203
Articles screened	1164
Articles excluded	852
Full-text articles assessed	312
Studies included	95

The following figure 2 PRISMA diagram is showing the number of records identified, screened, excluded, and ultimately included. The initial search yielded total articles screened at each stage using PRISMA diagram summarizing selection process ensuring only relevant high-quality studies included within review.

To ensure the reliability and relevance of the study, specific inclusion and exclusion criteria were established. These criteria were designed to define the study population clearly, ensure the validity of the results, and minimize potential biases. Participants/data/samples that met the inclusion criteria were considered eligible for the study, while those falling under the exclusion criteria were excluded to maintain consistency and accuracy. The detailed criteria are outlined in the table below.

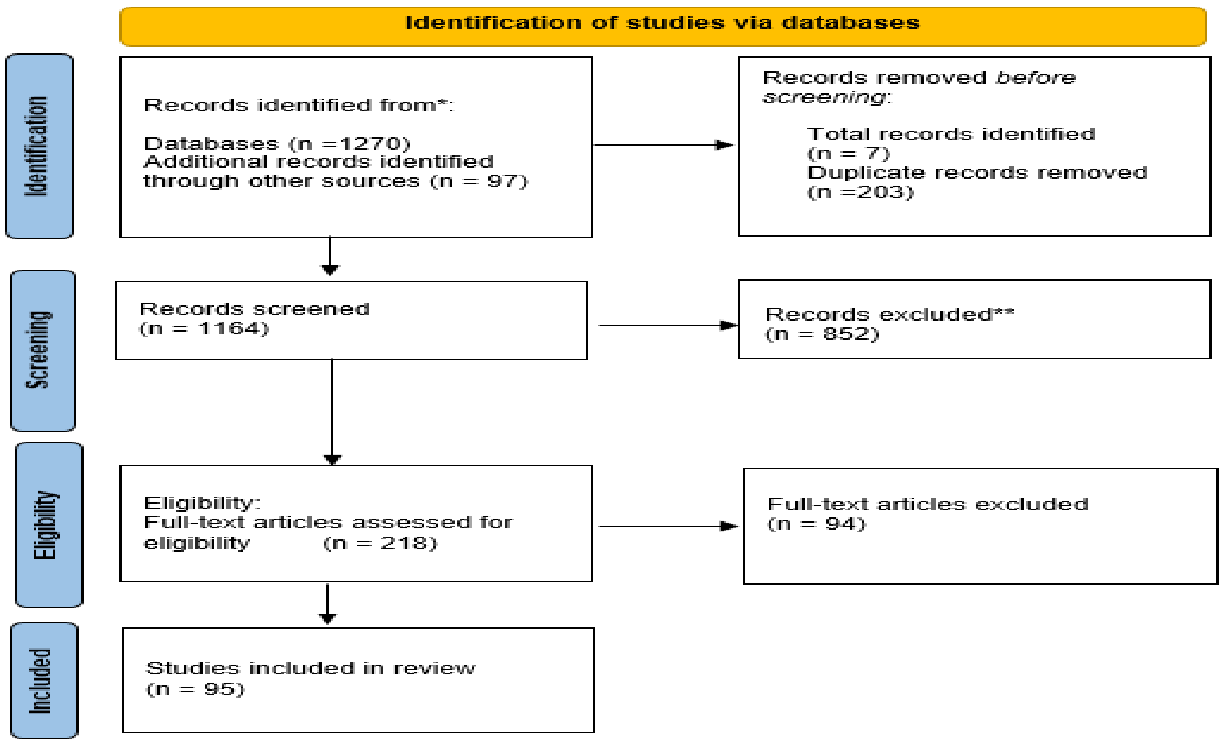

Figure 2.
PRISMA methodology

Table 2.
Inclusion and Exclusion criteria

Criteria	Inclusion	Exclusion
Relevance to Topic	Studies focused on deep learning applications in big data.	Studies unrelated to deep learning or big data applications.
Publication Type	Peer-reviewed journal articles, conference papers, and book chapters.	Opinion pieces, blogs, editorials, or non-peer-reviewed sources.
Publication Date	Published within the last 5-7 years (or foundational papers).	Studies older than 7 years unless highly cited or foundational.
Focus on Big Data	Emphasis on large-scale datasets, big data analytics, or data processing.	Papers focusing solely on small-scale data or traditional analysis techniques.
Application Context	Domain-specific (e.g., healthcare, finance) or general big data applications involving deep learning.	Studies with no practical application context or relevance to big data.
Language	Articles published in English.	Articles in languages other than English.
Technical Depth	Includes details about deep learning models, architectures, or implementations in big data contexts.	Purely theoretical studies or those lacking technical insights on deep learning in big data.
Duplicate Studies	Unique studies offering new insights or findings.	Duplicate studies or papers replicating findings without adding value.

A formal search strategy was developed to ensure adequate coverage of the relevant literature. The purpose of this search strategy was to examine multiple Deep Learning studies with big data applications from multiple reliable sources. To enhance the retrieval efficiency, keywords were selected by combining synonyms and new terms with commonly used terms from the Deep learning and big data applications. Table of Database Search strategy with results is given Below

Table 3.
Database

IEEE Xplore	**214**	**14**	**6.54%**
ACM Digital Library	186	13	7%
SpringerLink	143	11	7.69%
Elsevier ScienceDirect	299	8	2.67%
Taylor and Francis	27	2	7.4%
arXiv	128	12	9.37%
PubMed	108	8	7.4%
Frontier	165	8	4.84%

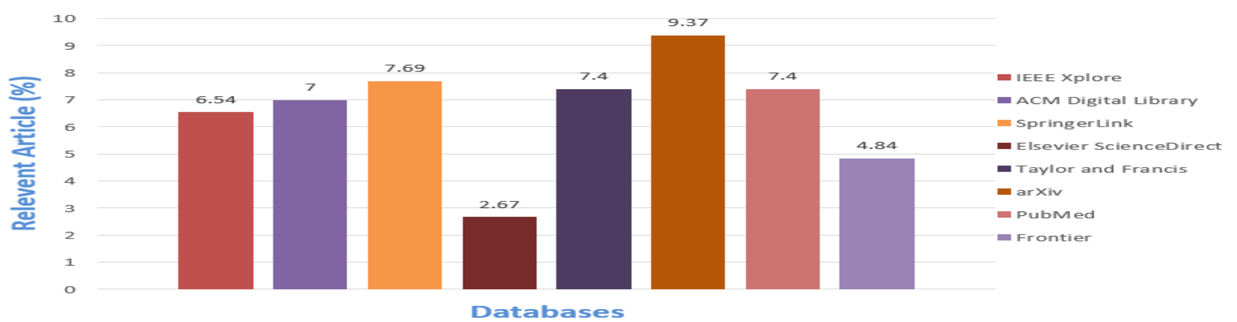

Figure 3.
Summary of Search Query for Methodology

SURVEY DISCUSSION

Deep learning (DL) has emerged as a breakthrough in solving data processing problems, providing scalability, pattern recognition, and pre-processing capabilities. Its ability to efficiently process multiple datasets, identify patterns, and support real-time decision support has already been demonstrated in fields such as organizational security, healthcare, and IoT. For instance, models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) specialize in spatial and temporal features, whereas lightweight methods with low labor costs and increasing data volume as well as solving complex problems are better for example. The effectiveness of DL is strongly influenced by the characteristics of the network. In healthcare, the importance of knowledge requires sophisticated systems with multiple applications that incorporate images, text and data. The audit function prioritizes periodic analysis and compliance, working securely to use constructive and proactive methods to effectively improve risk detection and reduce false positives. IoT and urban systems use DL for real-time data processing and predictive analysis, often relying on fog computing to address time constraints and energy consumption. Across the board, a wide range of silicon photonics and advanced technologies help me meet the growing demand for computing power. However, applying DL to data involves heterogeneous tasks, calculations within the framework and initial costs. Despite the high-quality models, they require a large number of computational resources, which increases energy consumption and acquisition time.

Strategies such as shared resource management mitigate this problem by introducing existing tasks and using lifecycle strategies that have not yet been identified. In the meantime, sophisticated infrastructure and fog computing allow for slower latencies affecting speed. To find the right balance between efficiency and effectiveness, technologies such as robotics and artificial intelligence become easier and more efficient. The Hybrid approach, which combines both supervised and unsupervised processing, improves efficiency and saves on computational costs.

The biggest challenge with data is the organization and organization of the data and to ensure its accuracy and retrieval. DL addresses this issue using noise reduction techniques, reduction, and missing value techniques. Auto encoders compress high-dimensional data, CNNs are features of the environment, and RNNs process data sequentially, transforming raw data into well-labeled data. Unbalanced datasets, a common problem in many applications, are addressed by techniques such as artificial minority oversampling techniques (SMOTE), generative adversarial networks (GANs), and cost-aware learning, where Fair model performance in between groups is determined. Distributed systems, low-power architectures, and flexible approaches such as workflows make data organization and resource management easier. To do so, the ability of DL to balance efficiency, effectiveness, and resource requirements in each category makes it a valuable tool for data analysis. By integrating pre-trained hardware, hardware, and human approaches to mitigation, DL is continuously evolving, providing scalability, real-time processing, and low-cost approaches. Its integration with many technologies such as IoT and fog computing is the future where DL will not only appear but is also the next vision for an increasingly data-driven world. In short, deep learning techniques work by preprocessing data with denoising, normalization, and handling missing values. Feature extraction using auto encoders, CNNs, and RNNs; Data imbalance due to oversampling, under sampling, cost-related learning, and GANs. Other strategies include transfer learning, distributed computing, fog computing, energy-efficient algorithms, model harvesting, knowledge distillation, and hybrid methods. This is how deep learning and big data are working combine in various domains.

FUTURE DIRECTIONS

Deep learning (DL) has been extensively utilized to address the challenges of big data analytics, bringing several unique capabilities to the field. Here are some key aspects highlighted in the literature: The first section focuses on presenting the applications and advantages of deep learning algorithms for big data analysis. However, the properties associated with big data has experiments for designing and optimizing deep learning to solve these problems. This section highlights some of the big data areas that deep learning should explore, especially learning from streaming data, working with high-dimensional data, model scalability, and distributed computing. In summary, future directions in DL will focus on data quality, accuracy, and precision. Reduction of complexity; application scalability; Deep learning (DL) in big data analytics faces major limitations related to data requirements, although these may change. Developing effective DL strategies often requires access to large, high-quality datasets, which can be difficult in areas where datasets are scarce or expensive. Instead of relying on industry-specific datasets based on domain-specific standards. These datasets often need to be pre-processed with a large information field to ensure accuracy and functionality, increasing the complexity of the application. To overcome this limitation, new methods should be developed, such as transfer learning from pre-trained models, generating artificial data to augment the dataset.

CONCLUSION

The future direction of deep learning (DL) in big data analytics will focus on improving data quality. High levels of data protection will be prioritized, with projects focusing on pre-processing technologies. Data performance efforts and IT infrastructure should be spent on managing the volume of data and the real-time analytics required. This includes integrating DL into distributed and hybrid computing platforms, infrastructure optimization, and digital application development. The widespread use of DL has become increasingly popular in various sectors such as healthcare, telecommunications, and telecommunications. In healthcare, DL can be combined with real-time monitoring and clinical decision support systems to improve patient outcomes. Similarly, funds can benefit from a better understanding of freedom and independence. The use of IoT is highly anticipated, especially in healthcare cities and smart cities, to enhance decision-making data and increase efficiency. Another important goal is to strengthen data security and regulatory compliance to address ethical and legal issues in the industry. Improving context awareness to improve the accuracy and quality of DL models, as well as model evaluation or decision support, is also important. Reducing computational costs while maintaining or improving model performance will enable DL to survive in challenging environments.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the supervisors and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Abidi, M. H., Umer, U., Mian, S. H., & Al-Ahmari, A. (2023). Big data-based smart health monitoring system: Using deep ensemble learning. *IEEE Access*, *11*, 114880–114903. https://doi.org/10.1109/ACCESS.2023.3325323

Academic Journal of Business Management. (2024). Analysis of financial risk behavior prediction using deep learning and big data algorithms. *Academic Journal of Business Management*, 6(10). https://doi.org/10.25236/AJBM.2024.061039

Alkhudhayr, A. (2024). Deep learning in the era of big data: Foundations, advances, applications, challenges, and future directions. *International Journal of Advanced Research, 12*(4), 549–552. https://doi.org/10.21474/IJAR01/18581

Amato, F., Marrone, S., Moscato, V., Piantadosi, G., Picariello, A., & Sansone, C. (2019). HOLMeS: eHealth in the big data and deep learning era. *Information, 10*(2), 34. https://doi.org/10.3390/info10020034

Calandra, R., Raiko, T., Deisenroth, M. P., & Pouzols, F. M. (2012). Learning deep belief networks from non-stationary streams. In A. E. P. Villa, W. Duch, P. Érdi, F. Masulli, & G. Palm (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2012* (Vol. 7553, pp. 379–386). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33266-1_47

Cen, J., & Li, Y. (2022). A rumor detection method from social network based on deep learning in big data environment. *Computational Intelligence and Neuroscience*, *2022*, 1–8. https://doi.org/10.1155/2022/1354233

Chang, W., Sun, D., & Du, Q. (2023). Intelligent sensors for POI recommendation model using deep learning in location-based social network big data. *Sensors, 23*(2), 850. https://doi.org/10.3390/s23020850

Chen, M. (n.d.). Marginalized denoising autoencoders for domain adaptation.

Chen, X.-W., & Lin, X. (2014). Big data deep learning: Challenges and perspectives. *IEEE Access, 2*, 514–525. https://doi.org/10.1109/ACCESS.2014.2325029

Cheung, W. K., et al. (2024). Interpolation-split: A data-centric deep learning approach with big interpolated data to boost airway segmentation performance. *Journal of Big Data, 11*(1), 104. https://doi.org/10.1186/s40537-024-00974-x

Cui, H., Si, B., Wang, J., Zhao, B., & Pan, W. (2024). Short-term origin–destination flow prediction for urban rail network: A deep learning method based on multi-source big data. *Complex & Intelligent Systems, 10*(4), 4675–4696. https://doi.org/10.1007/s40747-024-01391-6

Dean, J., et al. (n.d.). Large scale distributed deep networks.

Fan, L., Wang, H., Zhao, Y., & Xin, K. (2024). Application of deep learning in public network security management. *Transactions on Computer Science and Intelligent Systems Research*, *8*, 111–118. https://doi.org/10.62051/jycc7c87

Fang, Z., & Wang, S. (2024). Boosting financial market prediction accuracy with deep learning and big data: Introducing the CCL model. *Journal of Organizational and End User Computing*, *36*(1), 1–25. https://doi.org/10.4018/JOEUC.3584544

Fang, Z., Jin, C., & Liu, C. (2024). The impact of built environment in Shanghai neighborhoods on the physical and mental health of elderly residents: Validation of a chain mediation model using deep learning and big data methods. *Buildings, 14*(11), 3575. https://doi.org/10.3390/buildings14113575

Gao, W., Li, X., Wang, Y., & Cai, Y. (2022). Medical image segmentation algorithm for three-dimensional multimodal using deep reinforcement learning and big data analytics. *Frontiers in Public Health, 10*, 879639. https://doi.org/10.3389/fpubh.2022.879639

Glorot, X., Bordes, A., & Bengio, Y. (2011). Domain adaptation for large-scale sentiment classification: A deep learning approach. *arXiv*.

Gong, X., Wu, X., & Zhou, X. (2023). Deep learning-based security situational awareness and detection technology for power networks in the context of big data. *Applied Mathematics and Nonlinear Sciences*, *8*(1), 2939–2956. https://doi.org/10.2478/amns.2023.1.00012

Gong, Y., Wang, C., Li, J., Mahyuddin, M. N., & Seman, M. T. A. (2024). Application of deep learning in iron ore sintering process: A review. *Journal of Iron and Steel Research International*, *31*(5), 1033–1049. https://doi.org/10.1007/s42243-024-01197-3

Goswami, S., & Kumar, A. (2022). Survey of deep-learning techniques in big-data analytics. *Wireless Personal Communications, 126*(2), 1321–1343. https://doi.org/10.1007/s11277-022-09793-w

Jin, H., Jiao, T., Clifton, R. J., & Kim, K.-S. (2022). Dynamic fracture of a bicontinuously nanostructured copolymer: A deep-learning analysis of big-data-generating experiment. *Journal of the Mechanics and Physics of Solids, 164*, 104898. https://doi.org/10.1016/j.jmps.2022.104898

Kumar, R. H., & Sunitha, G. (2024). Big data analytics in healthcare environment using chaotic red deer optimizer with deep learning for disease classification model. *Multimedia Tools and Applications, 83*(32), 77697–77715. https://doi.org/10.1007/s11042-024-18239-3

Li, M., & Wang, Y. (2020). An energy-efficient silicon photonic-assisted deep learning accelerator for big data. *Wireless Communications and Mobile Computing, 2020*, 1–11. https://doi.org/10.1155/2020/6661022

Li, W., Hsu, C.-Y., & Tedesco, M. (2024). Advancing Arctic sea ice remote sensing with AI and deep learning: Opportunities and challenges. *Remote Sensing*, *16*(20), 3764. https://doi.org/10.3390/rs16203764

Li, Y., Huang, C., Ding, L., Li, Z., Pan, Y., & Gao, X. (2019). Deep learning in bioinformatics: Introduction, application, and perspective in big data era. *Bioinformatics*. https://doi.org/10.1101/563601

Li, Z., Chen, B., Wu, S., Su, M., Chen, J. M., & Xu, B. (2024). Deep learning for urban land use category classification: A review and experimental assessment. *Remote Sensing of Environment, 311*, 114290. https://doi.org/10.1016/j.rse.2024.114290

Lin, X., Zhang, Y., & Wang, J. (2022). Digital library information integration system based on big data and deep learning. *Journal of Sensors, 2022*, 1–8. https://doi.org/10.1155/2022/9953787

Liu, B., Ding, M., Zhu, T., Xiang, Y., & Zhou, W. (2019). Adversaries or allies? Privacy and deep learning in big data era. *Concurrency and Computation: Practice and Experience*, *31*(19), e5102. https://doi.org/10.1002/cpe.5102

Liu, W., Zeng, G., & Hu, K. (2020). Growth scale prediction of big data for information systems based on a deep learning SAEP method. *IEEE Access, 8*, 62883–62894. https://doi.org/10.1109/ACCESS.2020.2966770

Qin, F. (2022). College English intelligent writing score system based on big data analysis and deep learning algorithm. *Journal of Database Management, 33*(5), 1–26. https://doi.org/10.4018/jdm.314561

Sani, A., et al. (2024). Deep learning techniques in data mining: A comprehensive overview. *International Journal of Innovative Science Research and Technology*, 1254–1270. https://doi.org/10.38124/ijisrt/IJISRT24SEP367

Su, X., & Yin, W. (2024). RETRACTED ARTICLE: Application of Internet big data analysis technology based on deep learning in tourism marketing performance evaluation. *Soft Computing, 28*(S2), 595–595. https://doi.org/10.1007/s00500-023-08482-5

Sun, H., Liu, Z., Wang, G., Lian, W., & Ma, J. (2019). Intelligent analysis of medical big data based on deep learning. *IEEE Access, 7*, 142022–142037. https://doi.org/10.1109/ACCESS.2019.2942937

Wang, C.-T., & Chen, B.-S. (2023). Drug discovery for periodontitis treatment based on big data mining, systems biology, and deep learning methods. *SynBio, 1*(1), 116–143. https://doi.org/10.3390/synbio1010009

Wang, T., et al. (2024). Deep learning and machine learning, advancing big data analytics and management: Object-oriented programming. *arXiv*. https://doi.org/10.48550/ARXIV.2409.19916

Wang, W., Lu, D., Zhou, X., Zhang, B., & Mu, J. (2013). Statistical wavelet-based anomaly detection in big data with compressive sensing. *EURASIP Journal on Wireless Communications and Networking, 2013*(1), 269. https://doi.org/10.1186/1687-1499-2013-269

Wang, Y. (n.d.). Application of big data technology platform based on deep learning in smart tax evaluation system.

Wong, J., & Zhang, Q. (2023). Deep learning of sparse patterns in medical IoT for efficient big data harnessing. *IEEE Access, 11*, 25856–25864. https://doi.org/10.1109/ACCESS.2023.3256721

Xu, H., Wang, J., & Liu, C. (2024). Algorithm for detecting abnormal electricity usage by customers based on electricity big data and deep learning. *Journal of Physics: Conference Series, 2814*(1), 012043. https://doi.org/10.1088/1742-6596/2814/1/012043

Yu, M., Zheng, X., Qin, P., Cui, W., & Ji, Q. (2024). Urban color perception and sentiment analysis based on deep learning and street view big data. *Applied Sciences, 14*(20), 9521. https://doi.org/10.3390/app14209521

Yu, Y. (2022). Big data network security defense mode of deep learning algorithm. *Open Computer Science, 12*(1), 345–356. https://doi.org/10.1515/comp-2022-0257

Yu, Y., Li, M., Liu, L., Li, Y., & Wang, J. (2019). Clinical big data and deep learning: Applications, challenges, and future outlooks. *Big Data Mining and Analytics, 2*(4), 288–305. https://doi.org/10.26599/BDMA.2019.9020007

Yu, Z. (2021). Big data technology and deep learning algorithm in civil engineering teaching. *Journal of Physics: Conference Series, 1852*(3), 032023. https://doi.org/10.1088/1742-6596/1852/3/032023

Zhao, Q., Peng, S., Wang, J., Li, S., Hou, Z., & Zhong, G. (2024). Applications of deep learning in physical oceanography: A comprehensive review. *Frontiers in Marine Science*, *11*, 1396322. https://doi.org/10.3389/fmars.2024.1396322

Zhong, W., Yu, N., & Ai, C. (2020). Applying big data based deep learning system to intrusion detection. *Big Data Mining and Analytics, 3*(3), 181–195. https://doi.org/10.26599/BDMA.2020.9020003

Zhou, G., Sohn, K., & Lee, H. (n.d.). Supplementary material: Online incremental feature learning with denoising autoencoders.

Zhou, L. (2024). Perspective chapter: Big data and deep learning in hydrological modeling. In *River Basin Management - Challenges and Coping Strategies [Working Title]*. IntechOpen. https://doi.org/10.5772/intechopen.1007640

Zhou, Y., Zhao, S., Wang, X., & Liu, W. (2018). Deep learning model and its application in big data. In A. Marcus & W. Wang (Eds.), *Design, User Experience, and Usability: Theory and Practice* (Vol. 10918, pp. 795–806). Springer. https://doi.org/10.1007/978-3-319-91797-9_55

Zhu, M., & Chen, Q. (2020). Big data image classification based on distributed deep representation learning model. *IEEE Access, 8*, 133890–133904. https://doi.org/10.1109/ACCESS.2020.3011127