Elementaria: Journal of Educational Research
Journal website: https://elementaria.my.id/

E-ISSN: 2988-5949
DOI: https://doi.org/10.61166/elm.v2i1.66

Vol. 2 No. 1 (2024)
pp. 96-108

Research Article

Machine Learning for Database Management and Query Optimization

M.M.F. Fahima, A.H. Sahna Sreen, S.L. Fathima Ruksana, D.T.E. Weihena, M.H.M. Majid

1. Sri Lanka Institute of Information Technology, Sri Lanka; fahimamarzook@seu.ac.lk
2. Sri Lanka Institute of Information Technology, Sri Lanka; sahna@seu.ac.lk
3. Sri Lanka Institute of Information Technology, Sri Lanka; fathimaruksanasl@gmail.com
4. Sri Lanka Institute of Information Technology, Sri Lanka; dimaathathsarani97@gmail.com
5. South Eastern University of Sri Lanka, Sri Lanka; majidmhm9696@gmail.com

Received	: July 21, 2024	Revised	: August 04, 2024
Accepted	: August 11, 2024	Available online	: August 12, 2024

How to Cite: M.M.F. Fahima, A.H. Sahna Sreen, S.L. Fathima Ruksana, D.T.E. Weihena, & M.H.M. Majid. (2024). Machine Learning for Database Management and Query Optimization. *Elementaria: Journal of Educational Research*, *2*(1), 96–108. https://doi.org/10.61166/elm.v2i1.66

Machine Learning for Database Management and Query Optimization

Abstract. In the present day, Traditional database management methods are becoming more inadequate for effective data processing as the volume of data created by systems grows. Machine learning approaches have shown promise in optimizing database queries and enhancing database administration functions such as query optimization, workload management, indexing, and data quality assurance to solve this problem. We investigate the different machine learning algorithms used for query optimization and database management in this comprehensive literature review. Our review shows that machine learning approaches such as Deep Learning (DL), Reinforcement learning (RL), supervised learning, natural language processing (NLP), and unsupervised learning, among others, may be employed for query analysis, execution, and assessment. It is feasible to increase query performance and react to changing conditions by introducing machine learning techniques into database management systems.

Keywords: Teenager Query Optimization; Machine Learning; Artificial Intelligence; Database Management; Database Management Methods

Abstrak. Di masa kini, metode manajemen basis data tradisional semakin tidak memadai untuk pemrosesan data yang efektif seiring dengan meningkatnya volume data yang dihasilkan oleh sistem. Pendekatan pembelajaran mesin (machine learning) telah menunjukkan potensi dalam mengoptimalkan kueri basis data dan meningkatkan fungsi administrasi basis data seperti optimisasi kueri, manajemen beban kerja, pengindeksan, dan jaminan kualitas data untuk mengatasi masalah ini. Kami meneliti berbagai algoritma pembelajaran mesin yang digunakan untuk optimisasi kueri dan manajemen basis data dalam tinjauan literatur yang komprehensif ini. Tinjauan kami menunjukkan bahwa pendekatan pembelajaran mesin seperti Deep Learning (DL), Reinforcement Learning (RL), supervised learning, natural language processing (NLP), dan unsupervised learning, antara lain, dapat diterapkan untuk analisis kueri, eksekusi, dan penilaian. Dengan memasukkan teknik pembelajaran mesin ke dalam sistem manajemen basis data, dapat ditingkatkan kinerja kueri dan kemampuan untuk merespons kondisi yang berubah.
Kata Kunci: Optimisasi Kueri Remaja; Pembelajaran Mesin; Kecerdasan Buatan; Manajemen Basis Data; Metode Manajemen Basis Data

INTRODUCTION

The usage of data has significantly increased in recent years. In this digital era, the volume of data generated by various systems and applications continues to grow rapidly (Sarker, 2021; Senbekov et al., 2020). Whether it is data from social media, online transactions, IoT sensors, or other sources, the amount of data being produced is enormous. While large-scale data storage has become cheaper and more accessible, the main challenge today lies not only in storage capacity but also in managing and accessing this data efficiently (Nasir et al., 2022; Stergiou et al., 2020). This challenge is compounded by the fact that data is not only increasing in volume but also in variety and velocity. In this situation, traditional database management systems (DBMS) are increasingly struggling to handle and process large amounts of data efficiently and effectively.

Maintaining the integrity and reliability of databases has become more important in addressing these challenges (Abiodun et al., 2022; Garg & Goel, 2022; Lin et al., 2020). A trustworthy database is at the core of many business and technological applications today, where critical decisions often depend on the data stored (Diène et al., 2020; Upadhyay, 2020). However, simply expanding storage capacity or enhancing database infrastructure is no longer sufficient. Maintaining a reliable database requires a more intelligent and efficient approach that focuses not only on the technical aspects of storage but also on managing and optimizing queries to ensure that the data produced is relevant and timely (Alzahrani et al., 2022; Rudniy, 2022; Zou et al., 2022). Therefore, more advanced and dynamic solutions are needed to ensure that databases can continue to meet evolving demands.

Database query optimization is one of the critical aspects that must be addressed in modern data management (Boehm et al., 2022; Kossmann et al., 2022). In this context, query optimization refers to the effort to execute database queries in the most efficient way possible, given the available resources (Grzegorowski et al., 2021). This efficiency is measured not only by processing speed but also by the relevance and quality of the output generated (Madaan et al., 2024; Marion & Fixson, 2021). However, optimizing queries in a complex database is not easy, as many factors need to be considered, such as data structures, indexing, caching, parallelism, and network latency. All of these require a careful approach and often require manual adjustments by experienced database administrators.

As technology has advanced, machine learning (ML) has emerged as a potential solution to various challenges in database management, including query optimization. Machine learning offers a more adaptive and intelligent approach to data management, capable of learning from experience and historical data to make better decisions (Adi et al., 2020; Sarker, 2023; Zohuri & Rahmani, 2023). In the context of database management, machine learning can be applied to various tasks, from workload management, index selection, to data quality control (Kreuzberger et al., 2023; Whang et al., 2023). This approach enables database systems to adapt to changing conditions and continually improve query performance over time.

Machine learning techniques such as computer vision, natural language processing (NLP), reinforcement learning (RL), and deep learning (DL) have shown great potential in supporting database management activities (Cai et al., 2022; Naeem et al., 2020; Taye, 2023). For instance, NLP can be used to better understand the context of queries, allowing the system to respond in a more relevant and targeted manner. Meanwhile, RL can be used to optimize query execution by learning the best strategies based on feedback from previous executions (Ramadan et al., 2022; Wang et al., 2021). Deep learning can also be used to recognize patterns in data and predict query results with higher accuracy.

Additionally, one important application of machine learning in database management is in workload management (Gao et al., 2020; Liu et al., 2020). Workload management involves efficiently allocating resources to execute queries and other tasks within the database system (Perron et al., 2020; Uzzaman et al., 2024). By using ML techniques, systems can predict resource needs for each query and allocate resources optimally, thereby reducing processing time and improving overall system efficiency. The proper use of indexing can also be facilitated with the help of ML, where the system can automatically select and apply the most appropriate index based on data usage patterns.

Moreover, data quality control is becoming increasingly important in the era of big data (Fan & Geerts, 2022; McGilvray, 2021). Poor data quality can lead to incorrect conclusions and inaccurate decisions (Matheus et al., 2020). Machine learning techniques can be used to automatically detect anomalies, errors, and data completeness (Al-amri et al., 2021; Nassif et al., 2021). By leveraging models trained to recognize certain patterns in data, systems can proactively identify and address data quality issues before they affect analytical results.

The aim of this research is to explore and analyze how machine learning techniques can be applied in database management, particularly in the context of query optimization and overall database management. By doing so, it is hoped to provide new insights into how this technology can be used to enhance efficiency, accuracy, and relevance in managing increasingly complex and dynamic data.

METHOD

This research aims to analyze the contributions and findings of various relevant studies. In order to achieve this goal, various sources have been used, including leading academic platforms such as Science Direct, IEEE Explore, Research Gate, and Google Scholar. From various relevant publications, this research examines techniques such as neural networks, optimization algorithms, and machine learning approaches.

By reviewing the literature from these various sources, this research firmly gains an in-depth understanding of the application of these techniques in solving database management problems, especially in the context of query optimization and workload management. The results of this literature analysis provide an important basis for exploring the potential and challenges in applying machine learning technology in the field of database management.

RESULT AND DISCUSSION

Query Optimization

Query optimization is a very important concept for learning how to operate database queries effectively and efficiently within the available resource set (Krishnan et al., 2018; Ramadan et al., 2022). The relevance, speed, and overall value of the output are significantly affected by query optimization. Responsibility for systems that retrieve data and database management systems is critical in this context. However, in query optimization there are various factors that must be considered, including query semantics, data structure, indexing, caching, parallelism, and network latency (Marcus & Papaemmanouil, 2016). These considerations make it a complex challenge, considering that each factor can affect query performance differently.

Traditionally, query optimization is carried out using cost-based optimizers (Yang et al., 2022). These conventional techniques in query optimization rely on heuristics, statistics, and rule-based approaches to find optimal execution strategies (Karvelas et al., 2023; Rahman et al., 2024). These systems, which are driven by heuristics, require significant work to create and maintain in the first place, as well as considerable time to adapt to a particular database (Peres & Castelli, 2021). Two problematic features of most existing query optimization techniques are: first, they consist of complex and well-structured heuristics that have been developed over many years by developers. These algorithms often require additional adjustments by knowledgeable DBAs (Database Administrators) to increase query speed. Second, they employ a "fire and forget" strategy, meaning that the optimization process does not take into account the observed performance of the execution plan, thereby preventing query optimizers from learning from their mistakes (Marcus & Papaemmanouil, 2018).

The application of machine learning in query optimization addresses this problem by studying actual query performance and improving the recommendations provided about the best path for the query (Matošević et al., 2021). In this way, machine learning mimics the patterns of neural networks so that it can learn from experience (Blazek & Lin, 2021; Kufel et al., 2023). In contrast to traditional techniques, machine learning allows systems to dynamically adjust and improve optimization strategies based on collected performance data. This helps in overcoming the shortcomings of previous methods which tend to be unable to identify recent changes in the database and learn from past mistakes (Marcus & Papaemmanouil, 2018). Machine learning enables continuous adjustment and adaptation to new conditions, which can significantly increase the effectiveness and efficiency of query optimization.

How to Use Machine Learning to Improve Query Optimization?

Machine Learning for Query Analysis

The first step in query optimization is to study the query in depth and understand its context and meaning. Techniques in natural language processing (NLP), such as entity recognition, stemming, tokenization, parsing, and lemmatization, are very useful for achieving this. These techniques help in extracting terms, expressions, and ideas from a query and mapping them to relevant fields and data sources. By using these techniques, we can understand the query better and relate it to the appropriate data. This process is important to ensure that queries can be processed correctly and efficiently by the database system.

Additionally, machine learning methods such as query reformulation and expansion, query recommendation, and categorization can be very useful in dealing with difficult, ambiguous, or unclear queries. These techniques can be applied to improve, refine, or simplify queries, providing more options or guidance to users. For example, query reformulation can help turn less specific queries into more focused ones, while recommendations and categorization can provide suggestions or group queries based on similarities or specific goals (Milicevic et al., 2015). Thus, these techniques improve the quality and relevance of search results, thereby improving the user experience in using the database system.

Machine Learning for Query Execution

Another procedure in query optimization is planning and executing the query, which includes determining the best strategy, algorithm, and order of operations. Reinforcement learning (RL) methodologies like Q-learning (Krishnan et al., 2018; Li et al., 2019; Marcus & Papaemmanouil, 2018). Methods such as policy gradients and deep Q-networks can be useful. Through an iterative learning process called reinforcement learning (RL), a computer acting as an agent chooses actions on a regular basis and gets feedback on how well those actions work. Through multiple rounds (episodes), deep reinforcement learning approaches train a neural network model to maximize the performance of its assigned actions (policy) (Casals et al., 2023; Marcus & Papaemmanouil, 2018).

By learning from the feedback and benefits of past query executions, these reinforcement learning approaches can assist in modifying parameters and judgments as necessary. By using strategies like online learning and active learning to speed up query execution plans, machine learning may also help in adapting to changing data and circumstances. Plans in addition to transferrable skills. These techniques can assist in utilizing knowledge from other domains or activities, as well as modifying the model and approach in response to fresh information and feedback (Marcus & Papaemmanouil, 2018).

Machine Learning for Query Evaluation

The last step in query optimization is to assess the quality and relevance of the results. Machine learning can help with this by using supervised learning (SL) techniques (Uma at al., 2023). including classification regression (Roh at al., 2019). ranking, and clustering. Based on certain preset criteria or metrics, these strategies can assist in assigning scores, labels, rankings, or groupings to query results. Personalization, suggestion, and feedback are all strategies that machine learning may utilize to increase user experience and satisfaction. These strategies can assist in tailoring query results to the user's choices, requirements, and behavior, as well as soliciting their feedback and ratings. The following table shows some of the machine learning techniques and their application areas in DBMS.

Table 1. Machine Learning Techniques for Databases (Roh et al., 2019)

	Database Problem	Method
Offline NP Problem	Knob Space Exploration	Gradient-based
		Dense Network
		DDPG
	Index Selection	Q-learning

	View Selection	Q-learning
		DDQN
	Parity key selection	Q-learning
Online NP Problem	Join Order selection	Q-learning
		DQN
		MCTS
	Query rewriting	MCTS
Regression Problem	Cost estimation	Tree-LSTM
		Tree-ensemble
		Autoregressive
		Dense network
		Sum-product
	Index benefit estimation	Dense network
	View benefit estimation	Dense network
	Latency prediction	Dense network
		Graph embedding
	Learned index	Dense network
Prediction Problem	Trend prediction	Clustering based
	transaction scheduling	Q-learning

Workload Management

In the context of Workload management, or allocating resources to queries, is a technique used for improving the database management system (DBMS) performance. The use of machine learning (ML) methods, which predict query resource needs and allocate resources correctly, has improved workload management. When this technique trees is the use of reasoning trees, where a model is built to predict the resource requirements of a query based on its input parameters. A different approach is reinforcement learning, which entails teaching a model how to split resources according to the job at hand. Responsive workload management can save a Database Administrator's (DBA) time spent assessing performance and making changes. Use IBM Db2 to demonstrate this (Li et al., 2021).

Indexing

For to quicken up data retrieval, database developers use a structure for data known as an index. The type of index used will depend on the data architecture and the most frequently used queries (Marcus at al., 2016). Indexing is the process of creating indexes for table contents in databases to improve query performance. Techniques in machine learning, such as deep reinforcement learning, are utilized (Roberts at al., 2021; Sharma at al., 2021). Have been used in the procedure of training a model to perform automated index selection, which chooses the optimal indexing for a particular table in the database based on workload. One such technique involves training a model to predict the positive effects of introducing an index to a particular column in a database table using decision trees. Another approach is to use reinforcement learning, a method in which a model is taught to determine whether or not to build an index depending on the job at hand.

Data Quality Assurance

Data Quality management includes evaluating data abnormalities, mistakes, and omissions to identify the quality of data.ML techniques have been applied to ensure the quality of data by automatically identifying these problems through model training. Clustering is one such technique that entails teaching a model to identify outliers and combine related data points into groups. Another method is to use the identification of anomalies, which involves training a model to identify odd patterns in data that may point to mistakes or contradictions (Bai & Zhuo, 2020; Farias et al., 2016; Karakurt et al., 2017; Sharma, 2021).

We have found that there are many different machine learning approaches that can be applied to query optimization. Examples of these include decision trees, genetic algorithms, reinforcement learning (RL), deep learning, convolutional neural networks (NN), and neural networks (NN). The vast majority of research was done on relational database query optimization. Using decision trees, the best query execution method was found depending on the information being analyzed and the query structure. The best set of query parameters for optimization was found using genetic algorithms. The reinforcement learning method was employed to ascertain the optimal query execution strategy, while neural networks were utilized to forecast the optimal execution plan depending on the question's characteristics. The results show that machine learning methods can perform better than conventional optimization.

Benefits and Challenges in Applying Machine Learning to Query Optimization and Database Management

Machine learning can be an effective method for query optimization because it has the potential to streamline, improve, and change the information retrieval process. By using machine learning, we can give our users better service and value while also increasing the efficacy, accuracy, and relevance of your inquiries (Ramadan at al., 2022; Yang, 2019). However, there are several obstacles and restrictions to machine learning, such as those related to data quality, interpretability, scalability, and privacy. Because of this, we must constantly weigh the benefits and drawbacks of machine learning and employ caution when applying it (Ramadan et al., 2022).

However, some of the older techniques can be of poor quality and result in decreased efficiency because they cannot account for the complex interconnections between query optimization aspects. Query optimization has been compared with machine learning methods. This article provides a thorough examination of recent work on the use of machine learning algorithms for query optimization. Several machine-learning techniques have been proposed for query optimization. Augmentation learning is one such approach that entails training a model to make decisions based on both advantages and disadvantages. Over time, the model learns to make judgments that result in better query performance (Marcus & Papaemmanouil, 2016).

CONCLUSION

Machine learning algorithms have demonstrated considerable potential in database administration and query optimization activities such as managing workloads, the use of indexing optimizing queries, and data quality assurance. Before ML is extensively used in commercial systems, several challenges need to be addressed. The need for enormous quantities of labeled data for ML model training is one of these challenges. Another pries the requirement for scalable and trustworthy real-world implementations of ML algorithms. Last but not least, integrating machine learning with traditional DBMS can be difficult and necessitate modifications to the fundamental software and architecture. Research should concentrate on creating machine learning (ML) algorithms that are scalable, reliable, and simple to combine with traditional DBMS. It's also crucial to offer guidelines and resources for training and deploying ML models in database management platforms. With the help of all these attempts, machine learning (ML) can process large amounts of data effectively and efficiently, which has the potential to alter databases and database administration.

REFERENCES

Abiodun, O. I., Alawida, M., Omolara, A. E., & Alabdulatif, A. (2022). Data provenance for cloud forensic investigations, security, challenges, solutions and future perspectives: A survey. *Journal of King Saud University-Computer and Information Sciences*, *34*(10), 10217–10245.

Adi, E., Anwar, A., Baig, Z., & Zeadally, S. (2020). Machine learning and data analytics for the IoT. *Neural Computing and Applications*, *32*, 16205–16233.

Al-amri, R., Murugesan, R. K., Man, M., Abdulateef, A. F., Al-Sharafi, M. A., & Alkahtani, A. A. (2021). A review of machine learning and deep learning techniques for anomaly detection in IoT data. *Applied Sciences*, *11*(12), 5320.

Alzahrani, A., Alyas, T., Alissa, K., Abbas, Q., Alsaawy, Y., & Tabassum, N. (2022).

Hybrid approach for improving the performance of data reliability in cloud storage management. *Sensors*, *22*(16), 5966.

Bai, Z., & Zhuo, R. (2020). Quality Management of Crowd Sensing Data Based on Machine Learning. *2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, 185–188.

Blazek, P. J., & Lin, M. M. (2021). Explainable neural networks that simulate reasoning. *Nature Computational Science*, *1*(9), 607–618.

Boehm, M., Kumar, A., & Yang, J. (2022). *Data management in machine learning systems*. Springer Nature.

Cai, Q., Cui, C., Xiong, Y., Wang, W., Xie, Z., & Zhang, M. (2022). A survey on deep reinforcement learning for data processing and analytics. *IEEE Transactions on Knowledge and Data Engineering*, *35*(5), 4446–4465.

Casals, D., Buil-Aranda, C., & Valle, C. (2023). *SPARQL query execution time prediction using Deep Learning*.

Diène, B., Rodrigues, J. J. P. C., Diallo, O., Ndoye, E. L. H. M., & Korotaev, V. V. (2020). Data management techniques for Internet of Things. *Mechanical Systems and Signal Processing*, *138*, 106564.

Fan, W., & Geerts, F. (2022). *Foundations of data quality management*. Springer Nature.

Farias, V. A. E., Sousa, F. R. C., Maia, J. G. R., Gomes, J. P. P., & Machado, J. C. (2016). Machine learning approach for cloud nosql databases performance modeling. *2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, 617–620.

Gao, J., Wang, H., & Shen, H. (2020). Machine learning based workload prediction in cloud computing. *2020 29th International Conference on Computer Communications and Networks (ICCCN)*, 1–9.

Garg, M., & Goel, A. (2022). A systematic literature review on online assessment security: Current challenges and integrity strategies. *Computers & Security*, *113*, 102544.

Grzegorowski, M., Zdravevski, E., Janusz, A., Lameski, P., Apanowicz, C., & Ślęzak, D. (2021). Cost optimization for big data workloads based on dynamic scheduling and cluster-size tuning. *Big Data Research*, *25*, 100203.

Karakurt, İ., Özer, S., Ulusinan, T., & Ganiz, M. C. (2017). A machine learning approach to database failure prediction. *2017 International Conference on Computer Science and Engineering (UBMK)*, 1030–1035.

Karvelas, A., Foufoulas, Y., Simitsis, A., & Ioannidis, Y. E. (2023). Toulouse: Learning Join Order Optimization Policies for Rule-based Data Engines. *EDBT/ICDT Workshops*.

Kossmann, J., Papenbrock, T., & Naumann, F. (2022). Data dependencies for query optimization: a survey. *The VLDB Journal*, *31*(1), 1–22.

Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (mlops):

Overview, definition, and architecture. *IEEE Access*, *11*, 31866–31879.

Krishnan, S., Yang, Z., Goldberg, K., Hellerstein, J., & Stoica, I. (2018). Learning to optimize join queries with deep reinforcement learning. *ArXiv Preprint ArXiv:1808.03196*.

Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., Czogalik, Ł., Dudek, P., Magiera, M., & Lis, A. (2023). What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine. *Diagnostics*, *13*(15), 2582.

Li, G., Zhou, X., & Cao, L. (2021). Machine learning for databases. *Proceedings of the First International Conference on AI-ML Systems*, 1–2.

Li, G., Zhou, X., Li, S., & Gao, B. (2019). Qtune: A query-aware database tuning system with deep reinforcement learning. *Proceedings of the VLDB Endowment*, *12*(12), 2118–2130.

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., & Jenkyns, R. (2020). The TRUST Principles for digital repositories. *Scientific Data*, *7*(1), 1–5.

Liu, C., Feng, Y., Lin, D., Wu, L., & Guo, M. (2020). Iot based laundry services: an application of big data analytics, intelligent logistics management, and machine learning techniques. *International Journal of Production Research*, *58*(17), 5113–5131.

Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., Alon, U., Dziri, N., Prabhumoye, S., & Yang, Y. (2024). Self-refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems*, *36*.

Marcus, R., & Papaemmanouil, O. (2016). Workload management for cloud databases via machine learning. *2016 IEEE 32nd International Conference on Data Engineering Workshops (ICDEW)*, 27–30.

Marcus, R., & Papaemmanouil, O. (2018). Towards a hands-free query optimizer through deep learning. *ArXiv Preprint ArXiv:1809.10212*.

Marion, T. J., & Fixson, S. K. (2021). The transformation of the innovation process: How digital tools are changing work, collaboration, and organizations in new product development. *Journal of Product Innovation Management*, *38*(1), 192–215.

Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*, *37*(3), 101284.

Matošević, G., Dobša, J., & Mladenić, D. (2021). Using machine learning for web page classification in search engine optimization. *Future Internet*, *13*(1), 9.

McGilvray, D. (2021). *Executing data quality projects: Ten steps to quality data and trusted information (TM)*. Academic Press.

Milicevic, M., Baranovic, M., & Zubrinic, K. (2015). Application of machine learning algorithms for the query performance prediction. *Advances in Electrical and*

Computer Engineering, *15*(3), 33–44.

Naeem, M., Rizvi, S. T. H., & Coronato, A. (2020). A gentle introduction to reinforcement learning and its application in different fields. *IEEE Access*, *8*, 209320–209344.

Nasir, M. H., Arshad, J., Khan, M. M., Fatima, M., Salah, K., & Jayaraman, R. (2022). Scalable blockchains—A systematic review. *Future Generation Computer Systems*, *126*, 136–162.

Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine learning for anomaly detection: A systematic review. *Ieee Access*, *9*, 78658–78700.

Peres, F., & Castelli, M. (2021). Combinatorial optimization problems and metaheuristics: Review, challenges, design, and development. *Applied Sciences*, *11*(14), 6449.

Perron, M., Castro Fernandez, R., DeWitt, D., & Madden, S. (2020). Starling: A scalable query engine on cloud functions. *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 131–141.

Rahman, M. M., Islam, S., Kamruzzaman, M., & Joy, Z. H. (2024). Advanced Query Optimization In Sql Databases For Real-Time Big Data Analytics. *Academic Journal on Business Administration, Innovation & Sustainability*, *4*(3), 1–14.

Ramadan, M., El-Kilany, A., Mokhtar, H. M. O., & Sobh, I. (2022). RL_QOptimizer: A Reinforcement Learning Based Query Optimizer. *IEEE Access*, *10*, 70502–70515.

Roh, Y., Heo, G., & Whang, S. E. (2019). A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, *33*(4), 1328–1347.

Rudniy, A. (2022). Data Warehouse Design for Big Data in Academia. *Computers, Materials & Continua*, *71*(1).

Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, *2*(5), 377.

Sarker, I. H. (2023). Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects. *Annals of Data Science*, *10*(6), 1473–1498.

Senbekov, M., Saliev, T., Bukeyeva, Z., Almabayeva, A., Zhanaliyeva, M., Aitenova, N., Toishibekov, Y., & Fakhradiyev, I. (2020). The recent progress and applications of digital technologies in healthcare: a review. *International Journal of Telemedicine and Applications*, *2020*(1), 8830200.

Sharma, V. (2021). *Deep Learning Data and Indexes in a Database*. Utah State University.

Stergiou, C. L., Psannis, K. E., & Gupta, B. B. (2020). IoT-based big data secure management in the fog over a 6G wireless network. *IEEE Internet of Things Journal*, *8*(7), 5164–5171.

Taye, M. M. (2023). Understanding of machine learning with deep learning:

architectures, workflow, applications and future directions. *Computers*, *12*(5), 91.

Upadhyay, N. (2020). Demystifying blockchain: A critical analysis of challenges, applications and opportunities. *International Journal of Information Management*, *54*, 102120.

Uzzaman, A., Jim, M. M. I., Nishat, N., & Nahar, J. (2024). Optimizing SQL databases for big data workloads: techniques and best practices. *Academic Journal on Business Administration, Innovation & Sustainability*, *4*(3), 15–29.

Wang, C., Gruenwald, L., d'Orazio, L., & Leal, E. (2021). Cloud Query Processing with Reinforcement Learning-Based Multi-objective Re-optimization. *Model and Data Engineering: 10th International Conference, MEDI 2021, Tallinn, Estonia, June 21–23, 2021, Proceedings 10*, 141–155.

Whang, S. E., Roh, Y., Song, H., & Lee, J.-G. (2023). Data collection and quality challenges in deep learning: A data-centric ai perspective. *The VLDB Journal*, *32*(4), 791–813.

Yang, L., Yang, L., Pang, Y., & Zou, L. (2022). gCBO: A Cost-based Optimizer for Graph Databases. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 5054–5058.

Zohuri, B., & Rahmani, F. M. (2023). Artificial intelligence driven resiliency with machine learning and deep learning components. *Japan Journal of Research*, *1*(1).

Zou, B., You, J., Wang, Q., Wen, X., & Jia, L. (2022). Survey on learnable databases: A machine learning perspective. *Big Data Research*, *27*, 100304.