STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 11, Issue 1, 2024
DOI Journal: 10.37591/JoAIRA

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Review JoAIRA

Applications of Machine Learning Algorithms in Health Data Science (HDS) for Next Research Directions: A Survey Report

Vinay Bhatt[1,*], Mayank Kumar[2]

Abstract

In present time, data science is in big trend under computer science. The functioning of this technology is purely based on other advance technology known as machine learning (ML). Data science and machine learning are subsets of artificial intelligence (AI). When a process of data science is used in healthcare systems, the new system is known as health data science (HDS). HDS is a branch of data science used for handle the large amount of data in healthcare system. Recently, data science is used for handle and analysis the large volume of data (structured or unstructured) with accuracy by using the different techniques with algorithms of machine learning. This survey paper presented the ML applications in data science using different previous research. In this paper, firstly discuss on the introduction of paper with related information, secondly, discuss on review of literature on behalf of previous research, thirdly, discuss on machine learning with its techniques and examples, fourthly, discuss on stages of data science, fifthly, discuss on weakness or research gaps of previous research works according to literature review and finally discuss on proposed work for next research directions using observations to research gaps.

Keywords: AI, ML, data science, health data science, supervised learning, unsupervised learning, reinforcement learning, deep learning, deep-reinforcement learning, ANN

INTRODUCTION

Data science is a progressive technology in the present time and handled by other advanced technology known as machine learning (ML). Machine learning is a subpart of artificial intelligence (AI). Due to ML handle the different types of data using different techniques and algorithms so this technology used in data science technology. Machine learning is an advance a group of awareness and recognition as a technology that can evaluate huge amounts of data and computerizes the responsibilities of data scientists [10]. Via connecting involuntary compilations of common techniques that have changed predictable arithmetical advances, machine learning has changed the method of data extraction and analysis works [11]. In scheming efficient and express algorithms as well as data-driven forms on behalf of real-time data processing and machine learning can convey correct outcomes with analysis [12]. Machine learning is an important part of data science for handling the large amount of structured and unstructured data [13, 14]. Machine learning is

***Author for Correspondence**
Vinay Bhatt
E-mail: vinay10191@gmail.com

[1]Research Scholar, Department of Computer Science and Engineering Asian International University, Imphal West, Manipur, India
[2]Associate Professor, Department of Computer Science and Engineering Asian International University, Imphal West, Manipur, India

Received Date: December 29, 2023
Accepted Date: January 04, 2024
Published Date: January 17, 2024

Citation: Vinay Bhatt, Mayank Kumar Applications of Machine Learning Algorithms in Health Data Science (HDS) for Next Research Directions: a Survey Report. Journal of Artificial Intelligence Research & Advances. 2024; 11(1): 17–22p.

© STM Journals 2024. All Rights Reserved

dividing into five categories according to types of data which handle by algorithms. Data science decides the algorithms of machine learning for working with data. Following figure 1 shows the stages of machine learning with data science.

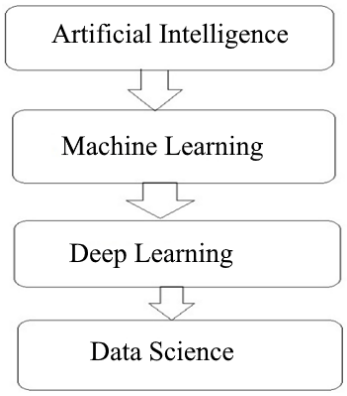

Figure 1. Stage of ML and data science under AI.

LITERATURE REVIEW

In this section, discuss the review of literature for next research directions on behalf of previous research, show in Table 1 as:

Table 1. Review of literature.

References	Research Category	Research Contributions
Rao, D. M. S., & Sridhathri, D. S, (2023) [1]	Diabetes prediction using ML algorithms	Proposed the research on diabetic prediction using five classification machine learning algorithms as XG Boost, decision tree, KNN, naïve bayes and random forest under precision. The outcome of this work is XG Boost algorithm is superior then other algorithms under precision [1].
AC, R., & Murthy, D. (2023) [2]	Approach of machine learning for diabetic prediction	Proposed the diabetic prediction model based on logistic regression based ML approach. The outcome is higher accuracy of data prediction using feature selection and regression technique [2].
Wee, B. F. et al., (2023) [3]	Deep learning and machine learning methods for diabetic detection	Proposed the review on different algorithms of machine learning and deep learning algorithm for diabetic detection. The outcome of this review is both algorithms are good in different fields according to purpose [3].
Madhu, B. et al., (2023) [4]	Diabetes risk prediction using ML algorithms	Proposed to different machine learning models as KNN, logistic regression, decision tree, random forest, ada boost classifier, XG boost and navies bayes classifier for predict to accuracy under diabetic prediction from PIMA Indian diabetes datasets. The outcome of this research is XG boost algorithm is highly accurate than other algorithms [4].
Sharma, A., & Mishra, P. K. (2022) [5]	Performance of ML for breast cancer diagnosis using optimized feature selection algorithm	Proposed the work on healthcare system for the observation to predict diseases like breast cancer using ML algorithms like DT, LR, KNN, ANN and RF under optimized feature selection algorithm [5].
Zhang, Q. et al., (2022) [6]	Data science approaches for COVID-19	Proposed the survey of data science for the research on COVID-19 using related parameters, mental health observation, diagnosis and risk measurement, digital make contact with tracing , communal media analytics with resource distribution and drug improvement [6].
Kumar S. et al., (2022) [9]	Big data analytics with machine learning on sustainable finance	Proposed the study of big data analytics with machine learning under sustainable finance research using related parameters like climate financing, social responsive financing, green

© STM Journals 2024. All Rights Reserved

		financing, carbon financing with impact investing for manage the profit and return with unifying policies [7].
Zeng, Z. et. al., (2022) [8]	ML methods for transcriptomics data analysis	Propose the implementation of ML and statistical methods like ANN, GCN, HMRF and SVCA for analysis to transcriptomics data with different data sets and summarizations [8].
Khan K. et al., (2022) [9]	ML application for concrete research	Propose the review on concrete research with different categories like conventional concrete, fiber reinforced, geo-polymer and recycle aggregate using different ML methods like supervised (task driven), unsupervised (data driven) and reinforcement (learn from error) learning [9].
Martins, R. M. & Gresse Von Wangenheim C., (2022) [10]	Machine learning application in teaching field	Proposed the survey on ML for teaching and used the case study in high school with regarding the strategy and technology of content with learning the concept of ML with algorithm and tasks for handling the project based problems [10].
He, Y. et al. (2022) [11]	ML in geochemistry and cosmochemistry	Proposed the implementation of ML methods for discover the hidden big data which related to Cosmo chemistry and geo chemistry using different process like water and soil quality prediction, sediment identification and digital mapping [11].
Gandomi, A. H. et al., (2022) [12]	ML in big data analytics	Proposed the review on ML for big data analytics with including process of handling the data like examine, analyzing and varied of data [12].
Liu T. et al., (2022) [13]	Deep learning for medical image analysis under COVID-19	Proposed the implementation of deep learning methods like CNN algorithm for medical image analysis in form of CT scan of COVID patients under COVID-19 [13].

DATA SCIENCE AND HEALTH DATA SCIENCE (HDS)

Data science is a modern technology which combines different subjects like math, statistics, specialized programming, data analysis, artificial intelligence (AI) and machine learning (ML) [15]. When an every processes of data science is used in healthcare system for handle the large amount of healthcare data the system is known as health data science (HDS) [17]. HDS is a branch of data science which implements on healthcare system for handle the bulk amount of patient data in healthcare system. Data science technology is based on the principle of machine learning and an algorithm which is used for finding the hidden pattern from raw data [16, 18]. Data science has four following stages, show in Table 2 as:

Table 2. Stages of data science.

S.N.	Data Science Stages	Description
1	Data ingestion	In this stage, collect the data in form of structured (customer data) and unstructured data (audio, video, log files) from different data sources like social media, websites etc.
2	Data storage and processing	In this stage, clean the data, transforming, duplicating and combine the data for storage to data warehouse using the ETL process like extract, transform and loads.
3	Data analysis	In this stage, analysis the data in form of pattern, values of distributions and range using predictive modeling like machine learning or deep learning for accuracy.
4	Communicate	In this stage, present the data in form of visualization like reports.

MACHINE LEARNING ALGORITHMS

Machine learning (ML) is an advanced field of computer science which is part of artificial intelligence (AI) [19, 20]. ML is a branch of AI using for increase the growth of data science using algorithm and related data for improve the human learning and data accuracy [21]. There are five types of machine learning, show in Table 3 as.

Table 3. ML types with examples.

© STM Journals 2024. All Rights Reserved

S.N.	ML Types	Description	Example
1	Supervised Learning	In this technique, involved training machine with lot of training data for specific task.	Decision tree, , logistic regression, support vector machine, K-nearest neighbor (KNN)
2	Unsupervised Learning	This technique is opposite to supervised learning means without any training machine and any training data. This technique is used for anomalies finding with clustering data.	K-means clustering,
3	Reinforcement Learning	This technique is used in research and development. In this technique, the output is depending on present input state and next input is depending on the output of prior input.	Q-Learning, markov decision process
4	Deep Learning	This technique is used for build neural network which function and structure is based on human brain.	ANN, CNN
5	Deep Reinforcement Learning	This technique is combining to deep learning and reinforcement learning which is used for builds robotics, smart healthcare system and game.	Deep Q learning

RESEARCH GAPS OF PREVIOUS RESEARCH ON BEHALF OF LITERATURE REVIEW

When review of literature is completed in this paper, discuss on the limitations or research gaps for work in next research directions on behalf of literature review show in Table 4 as:

Table 4. Research gaps for next research.

S.N.	Research Gaps
1	The performance of decision tree, KNN, naïve bayes and random forest is slower than XG Boost under precision for diabetic prediction [1].
2	The problem is regression and feature selection methods applies in only one approach of machine learning as logistic regression for diabetic prediction [2].
3	The problem is work of algorithm is based on dataset [3].
4	The performance of comparative algorithms is not good than XG Boost for accuracy of diabetic prediction [4].
5	The outcome of predictive results is not 100% accurate in terms of accuracy under ML based techniques [5].
6	The study of data science approaches which mentions in review is not proper fit for handle the COVID-19 pandemic [6].
7	The systematically study of big data analytics with ML under sustainable finance in small scale with including limited literature review [7].
8	Challenge the data analysis when increase data complexity [8].
9	The performance of ML methods is good for small input factors according to research [9].
10	The mostly problem is how to teach ML to students [10].
11	Miss the up to date ML methods for handle the different types of data with regarding to cosmo chemistry and geo chemistry [11].
12	Different types of problems faced in different types of big data analytics based research like fraud detection, medical informatics, national intelligence and marketing [12].
13	Mostly problem is handling the large imaging data sets [13].

PROPOSED WORK FOR NEXT RESEARCH ON BEHALF OF RESEARCH GAPS

In this section, discuss the further works for next research directions on behalf of research gaps which produced by literature review, show in Table 5 as.

Table 5. Proposed work for next research directions.

S. N.	Proposed Work
1	Further work on improve the performance of classification based machine learning algorithms as naïve bayes, random forest, KNN and decision tree for diabetic prediction.

© STM Journals 2024. All Rights Reserved

2	Further work on comparative analysis between different algorithms of machine learning for diabetic prediction using feature selection and regression methods.
3	Further work on collect to clear dataset of diabetic patients from healthcare system for analysis to accuracy using ML algorithms.
4	Further work on improve to performance of diabetic prediction accuracy using ML algorithms.
5	In future, work on neuro-fuzzy with combination of machine learning and deep learning for accuracy in resourceful diagnosis.
6	Further work on new approaches of data science with combined to machine learning for handle the COVID-19 pandemic infections.
7	Further study on sustainable finance using under different factors like moderating, dependent and independent variables, with different relationships like positive, negative, curvilinear and linear by big data analytics and ML in large scale .
8	Further work on develop a new sequencing protocol for reducing data complexity in spatial transcriptomics data.
9	Next work on performance improvement of ML when input factors increase.
10	Further work on build the supportive programming environment on behalf of machine learning models.
11	Propose the new approaches of ML which combine to deep learning for up to date ML.
12	Further work on big data analytics with advance ML like representation learning distributed learning, active with transfer learning and parallel learning.
13	Next work on deep learning with data science for handle the large amounts of data.

CONCLUSION

As we know, data science is a recent technology in computer science which is applicable in different fields. This technology is used for data accuracy in research fields using data analysis techniques by different types of related data like structured and unstructured data. When work on data science, use machine learning (ML) based techniques and algorithm for handle the data. This review paper, discuss the briefly introduction of machine learning with its techniques, stages of data science, review of literature, research gaps which produced by literature survey and proposed work for next research directions according to research gaps by review of literature.

Future Scope

Future scope of this paper is focus on advance algorithms of machine learning; advance tool and technology of data science for handle the data and produced to good accuracy.

Further work are focus on research gaps of this review paper, select any datasets like diabetes datasets, COVID-19 datasets, image dataset etc. in healthcare system which mentions in proposed work section for dataset collection and presents the new approach based work on health data science (HDS) for handle the data using machine learning algorithms.

REFERENCES

1. Rao, D. M. S., & Sridhathri, D. S. (2023). Diabetes mellitus prediction using ensemble machine learning techniques. In ITM Web of Conferences (Vol. 56, p. 05015). EDP Sciences.
2. AC, R., & Murthy, D. Diabetes Prediction Using Machine Learning Approach. 2023.
3. Wee, B. F., Sivakumar, S., Lim, K. H., Wong, W. K., & Juwono, F. H. (2023). Diabetes detection based on machine learning and deep learning approaches. Multimedia Tools and Applications, 1-33.
4. Madhu, B., Aerranagula, V., Mahomad, R., Ravindernaik, V., Madhavi, K., & Krishna, G. (2023). Techniques of Machine Learning for the Purpose of Predicting Diabetes Risk in PIMA Indians. In E3S Web of Conferences (Vol. 430, p. 01151). EDP Sciences.
5. Sharma, A., & Mishra, P. K. (2022). Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. International Journal of Information Technology, 14(4), 1949-1960.

© STM Journals 2024. All Rights Reserved

6. Zhang, Q., Gao, J., Wu, J. T., Cao, Z., & Dajun Zeng, D. (2022). Data science approaches to confronting the COVID-19 pandemic: a narrative review. Philosophical Transactions of the Royal Society A, 380(2214), 20210127.
7. Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2022). Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. Annals of Operations Research, 1-44.
8. Zeng, Z., Li, Y., Li, Y., & Luo, Y. (2022). Statistical and machine learning methods for spatially resolved transcriptomics data analysis. Genome biology, 23(1), 1-23.
9. Khan, K., Ahmad, W., Amin, M. N., & Ahmad, A. (2022). A Systematic Review of the Research Development on the Application of Machine Learning for Concrete. Materials, 15(13), 4512.
10. Martins, R. M., & Gresse Von Wangenheim, C. (2022). Findings on Teaching Machine Learning in High School: A Ten-Year Systematic Literature Review. Informatics in Education.
11. He, Y., Zhou, Y., Wen, T., Zhang, S., Huang, F., Zou, X., ... & Zhu, Y. (2022). A review of machine learning in geochemistry and cosmochemistry: Method improvements and applications. Applied Geochemistry, 105273.
12. Gandomi, A. H., Chen, F., & Abualigah, L. (2022). Machine learning technologies for big data analytics. Electronics, 11(3), 421.
13. Liu, T., Siegel, E., & Shen, D. (2022). Deep Learning and Medical Image Analysis for COVID-19 Diagnosis and Prediction. Annual Review of Biomedical Engineering, 24.
14. Krishnamoorthi, R., Joshi, S., Almarzouki, H. Z., Shukla, P. K., Rizwan, A., Kalpana, C., & Tiwari, B. (2022). A novel diabetes healthcare disease prediction framework using machine learning techniques. Journal of Healthcare Engineering, 2022.
15. Manley, K., Nyelele, C., & Egoh, B. N. (2022). A review of machine learning and big data applications in addressing ecosystem service research gaps. Ecosystem Services, 57, 101478.
16. Shahraki, A., Abbasi, M., Taherkordi, A., & Jurcut, A. D. (2022). A comparative study on online machine learning techniques for network traffic streams analysis. Computer Networks, 207, 108836.
17. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3), 1-21.
18. Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9, 381-386.
19. Jin, W. (2020, May). Research on machine learning and its algorithms and development. In Journal of Physics: Conference Series (Vol. 1544, No. 1, p. 012003). IOP Publishing.
20. Pandey, D., Niwaria, K., & Chourasia, B. (2019). Machine Learning Algorithms: A Review. Machine Learning, 6(02).
21. Ray, S. (2019, February). A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.
22. Divya, K. S., Bhargavi, P., & Jyothi, S. (2018). Machine learning algorithms in big data analytics. Int. J. Comput. Sci. Eng, 6(1), 63-70.

© STM Journals 2024. All Rights Reserved