STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 11, Issue 2, 2024
May–August
DOI Journal: 10.37591/JoAIRA

https://journals.stmjournals.com/joaira

Review JoAIRA

Data Handling Algorithms for the Healthcare System for the Prediction of Diabetes in Health Data Science (HDS): A Review Report

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

Abstract

In recent years, diabetes has become the biggest disease in different countries around the world. This disease is caused by adulteration in food ingredients, unhealthy food habits, a lack of physical exercise, and changing the lifestyle every time without a routine chart. The main objective of this review paper is to provide a proper understanding of the machine learning algorithm used in the healthcare system to handle diabetic patients' data. We use a previous research article for the idea of machine learning (ML) algorithm implementation in diabetic prediction to find more accuracy. The handling of diabetes patients' data is very difficult in the healthcare system, so use the modern technology of computer science as data science. Data science technology is used for handling data in healthcare systems, so the term is introduced as health data science (HDS). In this survey paper, we review different data handling algorithms as ML algorithms for handling diabetic patient data. This paper presents diabetes prediction based on previous research, which is discussed in the literature review section of this paper. In previous work, different ML algorithms such as Support vector machine (SVM), K-nearest neighbor (KNN), Logistic regression (LR), Random forest (RF), Decision tree (DT), Deep neural network (DNN), and Naïve Bayesian classifiers were used for handling the diabetes data, but different challenges faced researchers, so we focused on the challenges of previous research works. In this review paper, we focus on the research challenges of previous research and set research goals on behalf of research gaps for the next research directions. The future work of this paper is to analyze different data handling algorithms for the prediction of diabetes in different cases.

Keywords: Data science, health data science (HDS) machine learning (ML), supervised learning algorithm, data handling algorithms, Support vector machine (SVM), K-nearest neighbor (KNN), Random forest (RF), Decision tree (DT), DNN, healthcare system, diabetic prediction

***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: January 31, 2024
Accepted Date: February 06, 2024
Published Date: June 20, 2024

Citation: Vinay Bhatt, Mayank Kumar. Data Handling Algorithms for the Healthcare System for the Prediction of Diabetes in Health Data Science (HDS): A Review Report. Journal of Artificial Intelligence Research & Advances. 2024; 11(2): 1–10p.

INTRODUCTION

Data science is a useful technology that combines different multidisciplinary subjects such as statistics, mathematics; machine learning algorithms, etc. for handling the data in different multidisciplinary fields with different phases [1]. When the different phases of data science as data collection, processing, data management, analysis, and visualizations [2] are used in healthcare systems, the new concept of data science is introduced as health data science (HDS). In this review paper, the focus is on the technology of data science applied to healthcare systems for the prediction of diabetes using data handling algorithms. The data handling algorithms are the

same as machine learning algorithms in data science [3]. The paper presents the scenarios of diabetic prediction using a data handling algorithm in HDS under the healthcare system or hospital-based research on behalf of historical data on diabetes. This survey paper presents the previous work of machine learning in the direction of diabetic prediction using different algorithms of machine learning and the focus of this paper is the next work in the direction of diabetic prediction on behalf of previous work.

The following Figure 1 shows the basic concept of idea for the next research in diabetic prediction using data handling algorithm in HDS [4]. The branch of data science is HDS which can implement a large amount of healthcare data to improve the healthcare system procedures [5]. This is mainly used in healthcare systems for providing accurate data, the high-quality structure of data, decreasing the waiting time, and opening facilities [6]. HDS is a category of multidisciplinary areas which extended by algorithm development, statistical implication, and related technology to create an approach regarding data [7]. This paper presents the systematic survey of bulkily diabetic patients data under prediction of different patients which gain from healthcare system as hospitals on behalf of previous literature reviews.

Figure 2 shows the data handling model for handling the diabetic data with four main steps of research for prediction diabetes using ML algorithms, first is collecting the diabetic data from healthcare, second is phases of data science are used for clean and quality of data, third is used machine learning algorithm [8] as well as data handling algorithm for handle the diabetic data for finding the accuracy [9] and final step is predicting the data as accurate or not.

DATA HANDLING ALGORITHMS IN HEALTH DATA SCIENCE

The data handling algorithm is an algorithm which is based on machine learning (ML) techniques implemented in health data science to provide the solution for handling the bulky health-related data (predicted diabetic patients data) produced by hospitals or healthcare systems means every machine learning algorithm that handles the healthcare data known as data handling algorithms [10, 8]. These algorithms are separated into four techniques unsupervised, supervised, reinforcement, and deep learning techniques [6, 7]. There are some important data handling algorithms are used for diabetic prediction in HDS for the next research directions (Table 1).

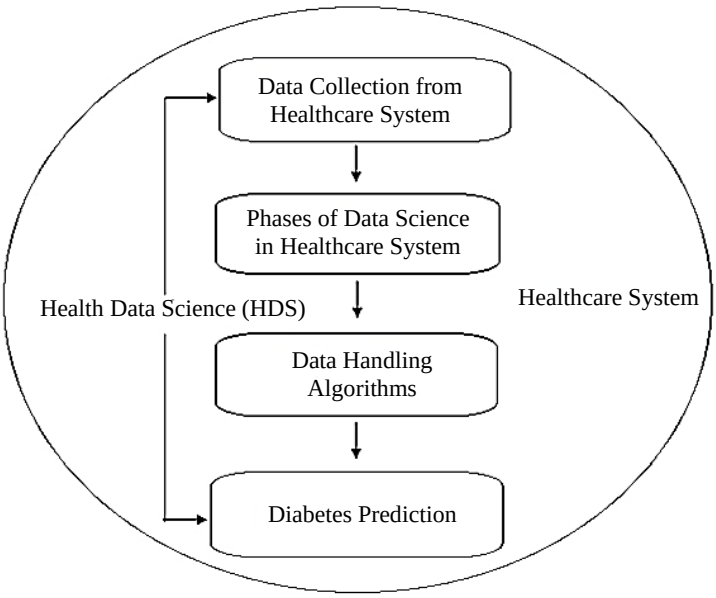

Figure 1. Diabetes prediction model using data handling algorithms.

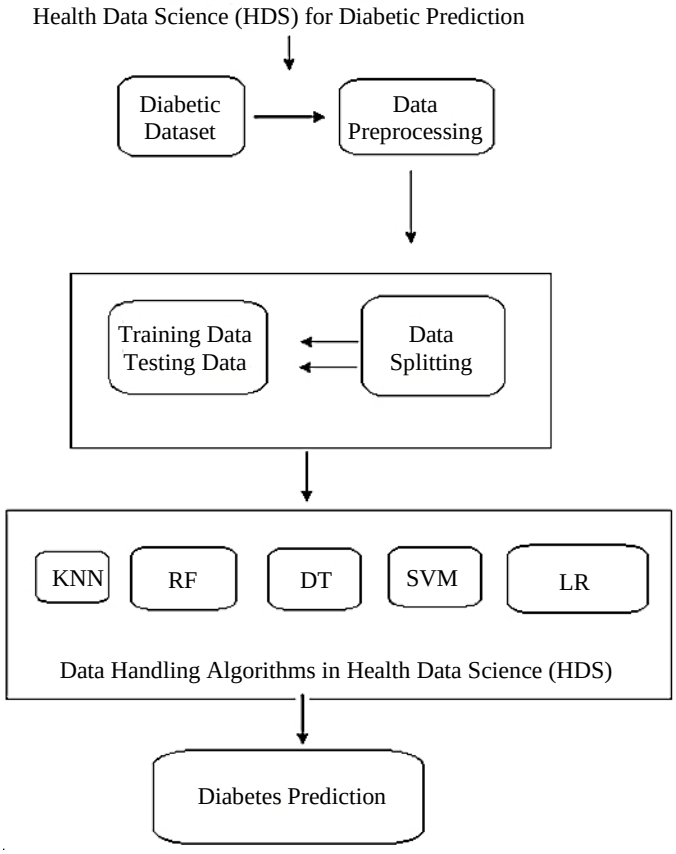

Figure 2. Data handling model in health data science (HDS).

Table 1. Data handling algorithms in HDS.

S.N.	Data Handling Algorithms	Descriptions
1	Support vector machine (SVM)	This is a vector-based network technique that aims to find hyperplanes under two separate categories of data sets.
2	K-nearest neighbor (KNN)	The non-parametric supervised learning algorithm which predicts the individual data point under groups of data solves both problems as classification and regression.
3	Logistic regression (LR)	The classification task-based supervised algorithm is used for analysis and correlation between independent variable sets and binary dependent variables.
4	Decision tree (DT)	This algorithm is a tree structure-based supervised learning that solved both types of problems as classification and regression. In the decision tree algorithm internal nodes indicate the features, branches indicate rules, and leaf nodes indicate the results of algorithms.
5	Deep neural network (DNN)	DNN is a feed forward-based deep learning algorithm that processes the data from the input to the output layer without going backward and connected to the layer in the forward direction and not touching a node again.
6	Random forest (RF)	This supervised learning algorithm is used to improve the performance by solving the complex problems of the model using multiple classifier combinations. The working of the RF algorithm is taking the outputs of multiple decision trees and producing a single output.
7	Naïve Bayesian classifier	This algorithm is a class of supervised learning algorithms used for text classification by using the Naïve Bayes theorem.

LITERATURE REVIEW

In this section, discuss on review of literature on behalf of previous work published in reputed journals [11]. This literature survey includes every work related to diabetic prediction-based research

in the form of review and analytical-based research papers and every work mentions with outcomes and weaknesses of these previous papers, in this survey paper section as the literature review section shown in Table 2 as:

Table 2. Literature survey for next research.

Reference	Research Category	Research Contributions	Research Gaps
Kakoly I.J. et al., (2023) [12]	Risk factor of data-driven diabetes prediction using ML algorithms	Proposed the study of ML applications for the data-driven type 2 diabetic prediction with risk factors using two feature selection techniques principal component analysis (PCA) and information gain (IG) under clinical and non-clinical aspects. The 82.2 % accuracy for diabetic prediction is the outcome of this research work [12].	The complicated calculation of computational complexity under advanced optimization techniques-based applications.
Arumugam K. et al., (2023) [13]	ML algorithms for multiple disease prediction	Proposed the three ML algorithms as SVM, Naïve Bayes, and decision tree for multiple disease prediction as diabetes and heart diseases. The outcome of this work is SVM and Naïve Bayes is outstanding performance than decision tree for the prediction of diabetes [13].	--
Febrian M. E. et al., (2023) [14]	Supervised learning for diabetes prediction	Proposed the implementation of two supervised MLs as KNN and Naïve Bayes for the observation of accuracy under the prediction of diabetes. The outcome of this work is performance of Naïve Bayes is better than KNN [14].	Advanced algorithm of neural networks is used for diabetic prediction for accuracy using additional techniques such as particle swarm optimization.
Kulkarni A. R. et al., (2023) [15]	Pre-diabetes and non-invasively diabetes detection by ML algorithm using electrocardiogram	Proposed the implementation of an ML algorithm for diabetic prediction under two categories: pre-diabetes and non-invasively diabetes detection through electrocardiogram (ECG). The technique of deep learning is better than machine learning for diabetic prediction [15].	--
Tasin I. et al., (2023) [16]	Techniques of explainable Artificial intelligence (AI) and machine learning for diabetic prediction	Proposed the different classification techniques of ML as random forest, decision tree, SVM, KNN, and linear regression algorithms for best diabetes prediction results. The decision tree-based XGBoost technique is better than other techniques for diabetes prediction [16].	Unavailability of insulin feature-based Real-Time Machine Learning (RTML) dataset.
Chou C.Y. et al., (2023) [17]	ML methods for onset diabetes prediction	Proposed the review on ML methods for diabetic prediction in eight different cases as ages, diabetes pedigree function, insulin level, BMI, diastolic blood pressure, sebum thickness, plasma glucose level, and number of pregnancies. The outcome of this study is two class-boosted decision trees and two class decision jungles are best for diabetic prediction [17].	Deep learning and unsupervised learning methods are used in the proposed model for future research directions.
Hama Saeed M.H. (2023) [18]	Upsampling techniques with machine learning algorithms for type 2 diabetes classification	Proposed the four machine learning algorithms for type 2 diabetes classifications under two categories of diabetes datasets: The Behavioral Risk Factor Surveillance System (BRFSS) and Photographic and Imaging Manufacturers Association (PIMA) dataset by using the upsampling technique with the ETC model. The accuracy of the BRFSS dataset is 0.99% and the PIMA dataset is 0.97% [18].	Deep learning techniques are used for diabetes prediction under data fusion-based hybrid models.

Reference	Research Category	Research Contributions	Research Gaps
Alwakid G. et al., (2023) [19]	Prediction of diabetic retinopathy using deep learning	Proposed the study of diabetic retinopathy (DR) prediction-based deep learning model under the five stages of DR as moderate, none, proliferate, mild with severe in both cases as without enhancement and contrast limited adaptive histogram equalization (CLAHE) filtering algorithm for image improvement. The outcome of this study is accuracy of Case 1 is 80.87% and Case 2 is 98.7% [19].	The additional augmentation approach for analysis of new data set using Visual Geometry Group (VGG) and DenseNet.
Tan K.R. et al., (2023) [20]	ML models for diabetes prediction	Proposed the method of machine learning as random forest (RF) for the prediction of diabetes with better performance under two types of diabetes populations [20].	A need for improvements to random forest for more validation.
Rastogi R. & Bansal M., (2023) [21]	Diabetes prediction model based on data mining techniques	Proposed the different techniques of data mining for diabetes prediction using special machine learning methods such as LR, SVM, Naïve Bayes, and random forest. The accuracy is higher of logistic regression (LR) than other ML techniques [21].	--
Kee O.T. et al., (2023) [22]	Diabetes prediction model using ML under cardiovascular complications	Proposed the review on ML-based diabetes prediction model in the case of cardiovascular complications. The outcome of this review is neural networks are best for the prediction of cardiovascular diseases between diabetic patients [22].	The performance analysis between gradient boosting machine and neural network under the same dataset.
Elsayed N.A. et al., (2023) [23]	Standard of care in diabetes with diagnosis and classifications	Proposed the review on classifications and diagnosis of diabetes on behalf of the American diabetes association (ADA) for standards of care in diabetes with three stages of diabetes autoimmunity, non-autoimmune, and monogenic diabetes [23].	
Nti I.K.et al. (2022) [24]	Big data analytics with machine learning	Proposed the assessment on ML under analytics of big data with a focus on the next work, a prospect with applications for different research directions. The outcome of this survey is some important algorithms of machine learning such as artificial neural network (ANN), DNN, CNN, decision tree, SVM, and ensemble learning techniques mostly used in different recent research areas like healthcare (for predicted diseases by ML), cyber security (for cyber-attack detection), communication networks (for mobile cellular networks), etc. and the result of this review is ML algorithms is maximum used in healthcare system [24].	The weakness of this research is data handling issues in diverse data sources when collecting incorrect and noisy data.
Gourisaria M.K. et al., (2022) [25]	Diabetes mellitus diagnosis of data science appositeness for healthcare systems	This research work uses four different models of machine learning as supervised (SVM, KNN, NB, RF, DT), unsupervised Function point (FP growth), reinforcement (P and N), and deep learning (ANN) under two different analytical methods as PCA and linear discriminant analysis (LDA) for handle the datasets. The outcome of this research is perfectly predicted Unified Configuration Interface (UCI) data sets by KNN, ANN, and DT compared to other techniques [25].	The problem statement of this research is data handling and analysis of a combination data sets of two different data sets with a focus on glucose levels using tongue images of patients.
Maier-Hein L. et al. (2022) [26]	Surgical data science	Proposed the review on data science with general and machine learning which is used in healthcare to improve the quality of healthcare by data modeling, data analysis, capture data,	The problem statements of this review failed to see the high-quality

Reference	Research Category	Research Contributions	Research Gaps
		and organization of data and implementation on surgical data science (SDS). The outcome of this work is easily tools available for work in healthcare fields [26].	data sets on a large scale.
Mohamed I. et al. (2022) [27]	ML algorithms for Readmission prediction of COPD patients	Propose the ML algorithms as ANN, decision tree, and SVM with comparison under the accuracy of predictability with hospital readmission prediction used for the improvement of COPD (Chronic Obstructive Pulmonary Disease) prediction. The outcome of this work is a high accuracy of prediction at 91% using proposed ML algorithms [27].	The weakness of this research work is records of COPD patients are limited.
Krishnamoorthi R. et al. (2022) [28]	Diseases prediction framework using ML techniques	This work focuses on identifying classification problems in accuracy under ML. The outcome of this work is 83% accuracy of prediction with the lowest error rate by using logical regression (LR) techniques and the performance of LR is better than other proposed ML algorithms [28].	The weakness of this research is works only on predictive structured data.
Awotunde J. B. et al. (2022) [29]	Data-centric operations in unsupervised learning and natural computing on smart healthcare	This research focuses on the particle swarm optimization (PSO) model for analysis of data on behalf of cardiovascular disease (CVD) using different algorithms of ant colony optimization (ACO), genetic algorithm (GA) of natural computing, and firefly swarm optimization (FSO) with unsupervised learning techniques as clustering and association rule. The outcome of this research is accuracy of the proposed model is better than another ML-based model [29].	The drawback of this research is difficult to control regular traditional methods due to complex behaviors.
Hassan M. et al. (2022) [30]	Big data analytics and genomics for personalized medicine	This paper focuses on the observation of bio-molecular data using four important parameters of dimensions: veracity, volume, variety, and velocity with observation uses four methods as content of generated data, scale, forms, and rate. The outcome of this study is the techniques of big data analytics are mostly used in handling large amounts of genomics data in healthcare systems for helping to discover new approaches to new drug findings in healthcare [30].	Missing the techniques of deep learning for handling a large amount of genomics big data by using data-driven approaches.
Oyeleye M. et al., (2022) [31]	ML techniques for the prediction of heart rates	This research indicates to performance comparison between proposed algorithms on behalf of heart rate prediction. The result of this paper is the performance of KNN, LSTM, and RF is not good for heart rate prediction while SVM is the best prediction model for heart rate prediction [31].	The drawback of this paper is data analytical technique is limited.
Zubair Md et al. (2022) [32]	K-means clustering algorithm-based efficient data modeling	Proposed the implementation for improvement of PCA model with handling to data-driven model using K-means clustering algorithm in COVID-19 data set. The outcome of this work is traditional K-means clustering is better than other existing models [32].	The weakness of this research is solution of the data data-driven model is complicated due to using of a real dataset.
Abdollahi J. & Nouri-Moghaddam B., (2022) [33]	Hybrid stacked ensemble for diabetic prediction using genetic algorithm	Proposed the combination techniques for diabetic prediction using two techniques a genetic algorithm and hybrid stacked ensemble algorithms. The outcome is higher performance in the diagnosis of the diseases [33].	Prediction of special diseases under online medical services.
Sann R. et al., (2022) [34]	Big data analytics for online reviews	Proposed the survey of big data analytics application for online complaints prediction	Due to data limitation, the finding of accuracy

Reference	Research Category	Research Contributions	Research Gaps
		against the behavior of the hospitality industry [34].	is difficult for a range of acceptable.
Sotto S. D. & Viviani M., (2022) [35]	Lack of health information detection on social websites	Proposed the data science approach for handling the lack of health information detection in social networking websites. The feature increases when classical ML classifications are used in the health training model [35].	The effects on the quality of results under depth analysis of health data.
Cerrato P. et al., (2022) [36]	Evaluation of algorithmic accuracy and equity	Proposed the survey of new platform development for evaluation of algorithm equity and accuracy under healthcare system with different digital tools such as radiology, oncology, dermatology, and ophthalmology [36].	The testing is missing in dataset population composition using algorithm validation techniques.

Table 3. Challenges and goals for diabetic prediction in HDS under healthcare system.

S.N.	Challenges	Goals
1	The data is incorrect and noisy due to diverse data sources, and it is very difficult to handle the data.	Performance comparison between LR, RF, DT, and SVM under anomaly detection for easily handled from single sources.
2	The machine learning techniques for data analysis of the diabetes dataset have been analyzed only according to blood glucose level, which is not sufficient.	Comparative analysis between KNN and Naïve Bayes classifier two features of diabetes like hemoglobin and blood sugar.
3	The test of diabetes PIDD-based predictive structured dataset.	Performance analysis between KNN, RF, SVM, and LR using a dataset of diabetes after smoking and drinking, physical inactivity, and family diabetes history.
4	Handling large amounts of data is the biggest problem in the healthcare system.	Proposed the DNN-based model in deep learning techniques for handling large amounts of diabetes data.

CHALLENGES AND GOALS

When completing the reviews of literature for the next research directions on behalf of previous research papers, face the different challenges for the goals of new and next research directions, as shown in Table 3.

CONCLUSION AND FUTURE SCOPE

This review paper focuses on the technique of machine learning for the prediction of diabetes in data science and is applied to healthcare systems. This paper proposes a survey of data handling algorithms in data science in the healthcare field and a new branch presented as well as HDS for handling the huge amount of diabetic predicted data in hospitals or healthcare systems. The aim of this paper is the proper review of the next research direction in the field of prediction and handling large amounts of diabetic's patient's data using different data handling algorithms as well as machine learning algorithms for analysis of the performance of machine learning algorithms. The review paper proposes seven selected ML algorithms with brief introductions as SVM, KNN, LR, DT, RF, Naïve Bayesian classifier, and DNN for data handling in HDS under the healthcare system on behalf of the prediction of diabetes for the next research directions.

The proposed paper is based on theoretical approaches of data handling algorithms in HDS, only proposed the review of the literature and basic information of diabetic prediction on behalf of theory based on previous research papers, focus on research gaps and challenges of previous work for next research directions. So, the next research work is on diabetic prediction-based research in the direction of practical implementation using ML algorithms as well as data handling algorithms and ML tools. The future scope of the survey paper is to focus on the achievements of data science in different problems of healthcare systems.

The future work of this review paper is the implementation of data handling algorithms as machine learning algorithms with ML tools in four cases of diabetes as first case is anomaly detection of diabetes, the second case is analysis of hemoglobin and fast blood sugar (FBS) level in a predictive dataset of diabetes, the third case is performance analysis under F1-score based on smoking habit, drinking habit, family diabetic history, and physical inactivity attributes in a dataset of diabetes and finally the fourth case of this direction is proposed the new ML approaches to handle the large predictive dataset of diabetics.

REFERENCES

1. Gharaibeh M, Alzu'bi D, Abdullah M, Hmeidi I, Al Nasar MR, Abualigah L, Gandomi AH. Radiology imaging scans for early diagnosis of kidney tumors: A review of data analytics-based machine learning and deep learning approaches. Big Data Cogn Comput. 2022;6:29. DOI: 10.3390/bdcc6010029.
2. Akpan IJ, Aguolu OG. The application of data science techniques and algorithms in women's health studies. medRxiv. 2022.
3. Sanchez-Pinto LN, Luo Y, Churpek MM. Big data and data science in critical care. Chest. 2018;154:1239–1248. DOI: 10.1016/j.chest.2018.04.037, PubMed: 29752973.
4. Khan FA, Zeb K, Al-Rakhami M, Derhab A, Bukhari SAC. Detection and prediction of diabetes using data mining: A comprehensive review. IEEE Access. 2021;9:43711–43735. DOI: 10.1109/ACCESS.2021.3059343.
5. Sisodia D, Sisodia DS. Prediction of diabetes using classification algorithms. Procedia Comput Sci. 2018;132:1578–1585. DOI: 10.1016/j.procs.2018.05.122.
6. Kaur G, Chhabra A. Improved J48 Classification Algorithm for the Prediction of Diabetes. Int J Comput Appl. 2014;98:13–17. DOI: 10.5120/17314-7433.
7. Selvin E, Steffes MW, Gregg E, Brancati FL, Coresh J. Performance of A1C for the classification and prediction of diabetes. Diabetes Care. 2011;34:84–89. DOI: 10.2337/dc10-1235, PubMed: 20855549.
8. Alam TM, Iqbal MA, Ali Y, Wahab A, Ijaz S, Baig TI, Hussain A, Malik MA, Raza MM, Ibrar S, Abbas Z. A model for early prediction of diabetes. Inform Med Unlocked. 2019;16:100204. DOI: 10.1016/j.imu.2019.100204.
9. Srivastava S, Sharma L, Sharma V, Kumar A, Darbari H. Prediction of diabetes using artificial neural network approach. In: Engineering Vibration, Communication and Information Processing: ICoEVCI 2018. Springer: Singapore; 2019. pp. 679–687. DOI: 10.1007/978-981-13-1642-5_59.
10. Sharma A, Mishra PK. Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. Int J Inf Technol. 2021;1–4.
11. Sharma A, Mishra PK. State-of-the-art in performance metrics and future directions for data science algorithms. J Sci Res. 2020;64:221–238. DOI: 10.37398/JSR.2020.640232.
12. Kakoly IJ, Hoque MR, Hasan N. Data-driven diabetes risk factor prediction using machine learning algorithms with feature selection technique. Sustainability. 2023;15:4930. DOI: 10.3390/su15064930.
13. Arumugam K, Naved M, Shinde PP, Leiva-Chauca O, Huaman-Osorio A, Gonzales-Yanac T. Multiple disease prediction using machine learning algorithms. Mater Today: Proc. 2023;80:3682–3685. DOI: 10.1016/j.matpr.2021.07.361.
14. Febrian ME, Ferdinan FX, Sendani GP, Suryanigrum KM, Yunanda R. Diabetes prediction using supervised machine learning. Proc Comput Sci. 2023;216:21–30. DOI: 10.1016/j.procs.2022.12.107.
15. Kulkarni AR, Patel AA, Pipal KV, Jaiswal SG, Jaisinghani MT, Thulkar V, Gajbhiye L, Gondane P, Patel AB, Mamtani M, Kulkarni H. Machine-learning algorithm to non-invasively detect diabetes and pre-diabetes from electrocardiogram. BMJ Innovations. 2023;9:32–42. DOI: 10.1136/bmjinnov-2021-000759.
16. Tasin I, Nabil TU, Islam S, Khan R. Diabetes prediction using machine learning and explainable AI techniques. Healthcare Technol Lett. 2023;10:1–10. DOI: 10.1049/htl2.12039, PubMed: 37077883.

17. Chou CY, Hsu DY, Chou CH. Predicting the onset of diabetes with machine learning methods. J Pers Med. 2023;13:406. DOI: 10.3390/jpm13030406, PubMed: 36983587.
18. Hama Saeed MA. Diabetes type 2 classification using machine learning algorithms with up-sampling technique. J Electr Syst Inf Technol. 2023;10:1–0.
19. Alwakid G, Gouda W, Humayun M. Deep learning-based prediction of diabetic retinopathy using CLAHE and ESRGAN for enhancement. Inhealthcare. 2023;11. DOI: 10.3390/healthcare11060863, PubMed: 36981520.
20. Tan KR, Seng JJB, Kwan YH, Chen YJ, Zainudin SB, Loh DHF, Liu N, Low LL. Evaluation of machine learning methods developed for prediction of diabetes complications: A systematic review. J Diabetes Sci Technol. 2023;17:474–489. DOI: 10.1177/19322968211056917, PubMed: 34727783.
21. Rastogi R, Bansal M. Diabetes prediction model using data mining techniques. Meas Sensors. 2023;25:100605. DOI: 10.1016/j.measen.2022.100605.
22. Kee OT, Harun H, Mustafa N, Abdul Murad NA, Chin SF, Jaafar R, Abdullah N. Cardiovascular complications in a diabetes prediction model using machine learning: A systematic review. Cardiovasc Diabetol. 2023;22:13. DOI: 10.1186/s12933-023-01741-7, PubMed: 36658644.
23. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, Collins BS, Gaglia JL, Hilliard ME, Isaacs D, Johnson EL, Kahan S, Khunti K, Leon J, Lyons SK, Perry ML, Prahalad P, Pratley RE, Seley JJ, Stanton RC, Gabbay RA. Classification and diagnosis of diabetes: Standards of care in diabetes—2023. Diabetes Care. 2023;46(Supplement 1):S19–S40. DOI: 10.2337/dc23-S002.
24. Nti IK, Quarcoo JA, Aning J, Fosu GK. A mini-review of machine learning in big data analytics: Applications, challenges, and prospects. Big Data Min Analytics. 2022;5:81–97. DOI: 10.26599/BDMA.2021.9020028.
25. Gourisaria MK, Jee G, Harshvardhan GM, Singh V, Singh PK, Workneh TC. Data science appositeness in diabetes mellitus diagnosis for healthcare systems of developing nations. IET Commun. 2022;16:532–547. DOI: 10.1049/cmu2.12338.
26. Maier-Hein L, Eisenmann M, Sarikaya D, März K, Collins T, Malpani A, Fallert J, Feussner H, Giannarou S, Mascagni P, Nakawala H, Park A, Pugh C, Stoyanov D, Vedula SS, Cleary K, Fichtinger G, Forestier G, Gibaud B, Grantcharov T, Hashizume M, Heckmann-Nötzel D, Kenngott HG, Kikinis R, Mündermann L, Navab N, Onogur S, Roß T, Sznitman R, Taylor RH, Tizabi MD, Wagner M, Hager GD, Neumuth T, Padoy N, Collins J, Gockel I, Goedeke J, Hashimoto DA, Joyeux L, Lam K, Leff DR, Madani A, Marcus HJ, Meireles O, Seitel A, Teber D, Ückert F, Müller-Stich BP, Jannin P, Speidel S. Surgical data science – from concepts toward clinical translation. Med Image Anal. 2022;76:102306. DOI: 10.1016/j.media.2021.102306, PubMed: 34879287.
27. Mohamed I, Fouda MM, Hosny KM. Machine learning algorithms for COPD patients readmission prediction: A data analytics approach. IEEE Access. 2022;10:15279–15287. DOI: 10.1109/ACCESS.2022.3148600.
28. Krishnamoorthi R, Joshi S, Almarzouki HZ, Shukla PK, Rizwan A, Kalpana C, Tiwari B. A novel diabetes healthcare disease prediction framework using machine learning techniques. J Healthc Eng. 2022;2022:1684017. DOI: 10.1155/2022/1684017, PubMed: 35070225.
29. Awotunde JB, Adeniyi AE, Ajagbe SA, González-Briones A. Natural computing and unsupervised learning methods in smart healthcare data-centric operations. In: Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data. Cambridge, USA: Academic Press; 2022. pp. 165–190.
30. Hassan M, Awan FM, Naz A, deAndrés-Galiana EJ, Alvarez O, Cernea A, Fernández-Brillet L, Fernández-Martínez JL, Kloczkowski A. Innovations in genomics and big data analytics for personalized medicine and health care: A review. Int J Mol Sci. 2022;23:4645. DOI: 10.3390/ijms23094645, PubMed: 35563034.
31. Oyeleye M, Chen T, Titarenko S, Antoniou G. A predictive analysis of heart rates using machine learning techniques. Int J Environ Res Public Health. 2022;19:2417. DOI: 10.3390/ijerph19042417, PubMed: 35206603.

32. Zubair M, Iqbal MD, Shil A, Chowdhury MJM, Moni MA, Sarker IH. An improved K-means clustering algorithm towards an efficient data-driven modeling. Ann Data Sci. 2022;1–3.
33. Abdollahi J, Nouri-Moghaddam B. Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. Iran J Comput Sci. 2022;5:205–220. DOI: 10.1007/s42044-022-00100-1.
34. Sann R, Lai PC, Liaw SY, Chen CT. Predicting online complaining behavior in the hospitality industry: Application of big data analytics to online reviews. Sustainability. 2022;14:1800. DOI: 10.3390/su14031800.
35. Di Sotto SD, Viviani M. Health misinformation detection in the social web: An overview and a data science approach. Int J Environ Res Public Health. 2022;19:2173. DOI: 10.3390/ijerph190 42173, PubMed: 35206359.
36. Cerrato P, Halamka J, Pencina M. A proposal for developing a platform that evaluates algorithmic equity and accuracy. BMJ Health Care Inform. 2022;29. DOI: 10.1136/bmjhci-2021-100423, PubMed: 35410952.

