International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Prediction of Patient Readmission via Machine Learning Algorithms

Samah Alajmani, Kamal Jambi

Abstract*: Predicting the probability of hospital readmission is one of the most vital issues and is considered to be an important research area in the healthcare sector. For curing any of the diseases that might arise, there shall be some essential resources such as medical staff, expertise, beds and rooms. This secures getting excellent medical service. For example, heart failure (HF) or diabetes is a syndrome that could reduce the living quality of patients and has a serious influence on systems of healthcare. The previously mentioned diseases can result in high rate of readmission and hence high rate of costs as well. In this case, algorithms of machine learning are utilized to curb readmissions levels and improve the life quality of patients. Unluckily, a comparatively few numbers of researches in the literature endeavored to address this issue while a large proportion of researches were interested in predicting the probability of detecting diseases. Despite there is a plainly visible shortage on this topic, this paper seeks to spot most of the studies related to predict the probability of hospital readmission by the usage of machine learning techniques such as such as Logistic Regression (LR), Support Vector Machine (SVM), Artificial Neural Networks (ANNs), Linear Discriminant Analysis (LDA), Bayes algorithm, Random Forest (RF), Decision Trees (DTs), AdaBoost and Gradient Boosting (GB). Specifically, we explore the different techniques used in a medical area under the machine learning research field. In addition, we define four features that are used as criteria for an effective comparison among the employed techniques. These features include goal, data size, method, and performance. Furthermore, some recommendations are drawn from the comparison which is related to the selection of the best techniques in the medical field. Based on the outcomes of this research, it was found out that (bagging and DT) is the best technique to predict diabetes, whereas SVM is the best technique when it comes to prediction the breast cancer, and hospital readmission.

Keywords: Hospital readmission; machine learning; decision trees; logistic regression, support vector machine; Linear Discriminant analysis; artificial neural networks; Bayes algorithm; random forest; AdaBoost; gradient boosting.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Samah Alajmani*, Faculty of Computing and Information Technology, Computer Science Department, King Abdulaziz University, Jeddah, Saudi Arabia. Email: salajmani@gmail.com
Prof. Kamal Jambi, Professor, Faculty of Computing and Information Technology, Computer Science Department, King Abdulaziz University, Jeddah, Saudi Arabia. Email: kjambi@kau.edu.sa

I. INTRODUCTION

Hospital readmission is one of the most important issues in the healthcare sector. Recently, readmissions to the hospitals have been at the center of healthcare studies and debates due to their known widespread negative effect on the finances and patient loads in healthcare systems globally [1]. Increased readmission to hospitals disturbs the normal life of patients and results in adverse outcomes for the healthcare systems [2]. In the US, for example, the Medicare Payment Advisory Committee has stated that during the discharge period of 30 days, 17.6% of patients were readmitted. This resulted in Medicare expenses of $17.9 billion annually through 76% of the patients could have been possibly prevented from readmission [1], [3]. The significant features to avoid admission to the hospital is to decrease the illness of patients, control the budget of healthcare and improve the results of patients [4]. Also, in economic terms, the intensity of the readmission issues can be quantified as the projected expenditure of unplanned readmissions is related to patients who are severely ill, elderly, and tolerating several diseases that are chronic, are accountable for a large percentage of this cost. In actual fact, over 27% of these readmissions can be prevented [1],[4] . In addition to this, the Patient Protection and Affordable Care Act was introduced to curb the rates of hospital readmissions. This act fined hospitals having unnecessary readmission at least 3% of Medicare compensation. In spite of the efforts being done, the intensity of rates of readmission is expected to increase in the coming years [1]. A system for decreasing payments of readmission has been initiated by the Centers for Medicare and Medicaid Services (CMS) through which if there is readmission of patients within a time period of 30 days after discharge, the hospital will be fined [5] . The executives of the hospital understand that the pressure for the rate of readmission will keep increasing for the coming years which will result in a rise of financial fines. At several hospitals, measures to decrease readmissions have been studied comprehensively and have been put into operation [6].

At present world-wide, several diseases that are heart failure (HF), chronic disease, (like diabetes) and etc. have been widely spread and the number of patients is constantly growing. These diseases have a negative impact on systems of the hospital in terms of the growing rate of hospital readmission and very high the cost.

These patients regularly go to the hospital, demanding persistent planning to ensure that the necessary services like hospital beds, adequate medical staff, and rooms are accessible for appropriate service quality. It is therefore of utmost significance to forecast the possibility of a given patient's readmission.

In reality, admitting the patient again through a discharge time of one month i.e. 30 days indicates an eminent healthcare quality evaluation of top importance. The aim is to focus on this issue [7].

Amongst the utmost significant branches of artificial intelligence is machine learning which offers techniques and approaches for learning from experience [8]. It is mostly used by researchers to perform complicated experiments involving tasks of statistical analysis [9]. It is a broad field involving many disciplines comprising, but not restricted to, algebra, data processing, knowledge analytics, statistics, cognitive science, control theory, information theory, the complexity of computations and philosophy. This discipline performs as a key role to discover important and worthy information from databases that may hold details of medical records, applications of loans, financial transactions, supply maintenance, etc. [10]. Machine learning techniques were used in healthcare for prediction for example prediction of diseases. In addition to this, nowadays, the healthcare sector will be promising to use these techniques in the prediction of hospital readmission. So, in this paper will discuss ten techniques that are used in healthcare in prediction of diseases and hospital readmission.

Regulation of remaining this paper looks likes: Section II provides an overview of machine learning techniques that are included in this research. Section III presents related works. Section IV provides discussion and results. Lastly, Section VI summaries the research and provides potential trends for future work.

II. BACKGROUND

A discussion of ten prevalent techniques of machine learning that were employed in this study is in this section.

A. Logistic Regression Analysis

Regression is a concept that depends on the statistic to find the link between a dependent variable (fixed variable) and an independent variable (set of varying variables). The primary objective of methods to analyze regression is to research and evaluate the correlation between a set of attributes. Logistic regression (LR) suggests models that are quantitative in which any of the dependent variable or outcome variable can be categorical data rather than continuous data [11].

B. Support Vector Machine

Support Vector Machine (SVM) considers an authoritative technique of classification which can rely on the theory of statistical learning. This model was suggested by Vapnik in 1995 [12] . Also, it is a collection of methods of supervised learning, this algorithm has the ability to apply analysis for both classification and regression. Given a simple training sample as a set of two classes, the main goal of this SVM is to discover the optimum separating hyperplane between those two classes. For the best popularization hyperplane must not lie nearer to the points of data which are belonged to other class. Selecting hyperplane must be away from points of the data from every class. Finally, points which are located closest to the classifier margin are support vectors [13].

C. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is the algorithm of important analysis of data that has been extensively used previously for identification, supervised classification and reduction of dimensionality [14]. In addition to this, LDA is an effective approach for improving the discriminative ability among various classes so as to the discriminative capability critically identifies the classification performance, adoption of LDA for classifying strongly confusable patterns [15].

D. Bayes Algorithm

One of these algorithms is the classifier of Naïve Bayesian (NB) depends on the implementation of Bayes theorem in order to predict a specific event's most likely to participate in a group of likely classes. As it believes in individuality between variables used in the mechanism of classification, it is defined as naïve [16], [13].

The following equation, using Bayes theorem, shows how to find out the posterior probability P(c|x) using likelihood P(x|c), class prior probability P(c), and predictor probability P(x) [17]:

$$P\ (c \mid x) = (\ P(x \mid c)\ .\ P(c))\ /\ P(x) \qquad (1)$$

In the equation stated above,

- Provided with the predictor (x, attributes), the class's posterior probability (c, target) is P(c|x).
- The probability of predictor particular to a class is P(x|c)
- The Class's prior probability is P(c).
- Predictor prior probability is P(x).

E. Artificial Neural Network

Artificial Neural Network (ANN) is related to deep learning, a machine-learning technique. This technique relies on the system of human beings' brain which comprised of neural networks. This model is widely implemented in several types of research as it is skilled in model systems that are not linear (systems with unidentified or challenging relation within variables) [18]. Multi-Layer Perceptron (MLP) is an example of ANN. It is made of 3 layers: hidden layer, input layer and output layer. For the processing of data, the neurons apply functions that are not linear [19].

F. Instant-Based Learning: K-Nearest Neighbor

Algorithm of K-Nearest Neighbor (KNN) is also named 'Memory-Based Learning'. It is the widely used method for classification based on Instance-Based Learning [16] . It is a technique that is substantially employed for samples classification. The distance from training samples number N can be calculated by means of KNN method [20]. This technique could be one of the public and classical machine learning techniques. KNN might be used in different problems of classification and prediction.

Retrieval Number: F7770038620/2020©BEIESP
DOI:10.35940/ijrte.F7770.038620
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

This technique can work on the distance functions basis. In this case, there are two various types of distance functions are extremely applied
first one will be the distance of Euclidean and the second one will be the distance of absolute. In the classification case, the independent data instance can be compared to the whole data set. In addition, based on the matrix of computed distance,
the data instances class labels are forecasted [21].

G. Decision Tree

Decision tree (DT) considers one of the popular methods utilized in machine learning. It depends on the classification by means of values of the attribute for making decisions. It consists of a set of roots, nodes, branches, and leaves [22] . In this technique, the classification of data can be from the node of root until the node of the leaf where decisions not made. It is usually used in research operations, especially in the analysis of decisions to assist and specify a strategy which will most probably reach the objective. Moreover, a transformation of the decision tree to a set of rules could be easily through using draw mapping from the node of root until reach the leaf nodes step by step and understanding of a group of rules, also can support both types of data either continued data or categorical data. In the end, following these steps could accomplish suitable consequences [8], [10]. A lot of algorithms have been recommended in the literature, for employing decision trees. CART (Classification and Regression Tree) is a significant algorithm. It is utilized for the variables that are categorical and continuous [20].

H. Ensemble-Based Learning

Techniques of ensemble-based learning create their predictions relied on a multi-classifiers' outputs combination. Ensemble learners involve boosting methods, for example, Stochastic Gradient Boosting, AdaBoost) and bagging methods, for example, Random Forest [16]. Boosting is a machine learning method on the basis of the conception of constructing a prediction rule that is extremely precise by merging several comparatively weak and inexact rules [23]. It considers a general approach to improve any learning algorithm's performance [24]. Bagging is focused on a bootstrapping methodology that creates several classification trees by continuously choosing random subsets from the training data [25].

- Bagging Techniques: Random Forests (RFs). RFs are a technique of ensemble learning which has the ability to support both approaches classification and regression. Through increasing several classification trees inside the phase of training, it expands the simple idea of a single classification tree. Every tree inside the forest creates its reply (voting for a class) to categorize an instance, the model selects the class which has gained the maximum votes compared to every tree inside the forest. One significant feature of RF compared to conventional decision trees is the protection versus overfitting that makes the technique has the ability to provide high performance [26], [27].

- Boosting Techniques: AdaBoost and Stochastic Gradient Boosting. Algorithm of AdaBoost and algorithm of Stochastic Gradient Boosting algorithm consider algorithms of ensemble-based that depend on the Boosting concept [16]. The AdaBoost algorithm has the capability to create poor learners' groups by preserving a group of weights over data of training and corrects them after every cycle of weak learning. Although increasing the training samples weights will occur when are incorrectly classified via present weak learner, decreasing the sample weights will happen when are properly classified [23]. However, gradient boosting gives significance to the difficult or misclassified instances by means of the strong learner's pseudo-residuals (the remaining errors). In every iteration, errors are calculated, and a weak learner is attuned to them. Next, the decrease in the whole error of the strong learner is the influence of the weak learner to the strong learner [16].

III. RELATED WORK

This section presents and compares different related works that have the same techniques proposed previously, and it discusses whether these techniques were used in the same application of prediction of hospital readmission or in different application, such as prediction of diseases. Since this section is considered as the core of the current research, we explain our strategy through the following flowchart, which illustrated in Figure 1.

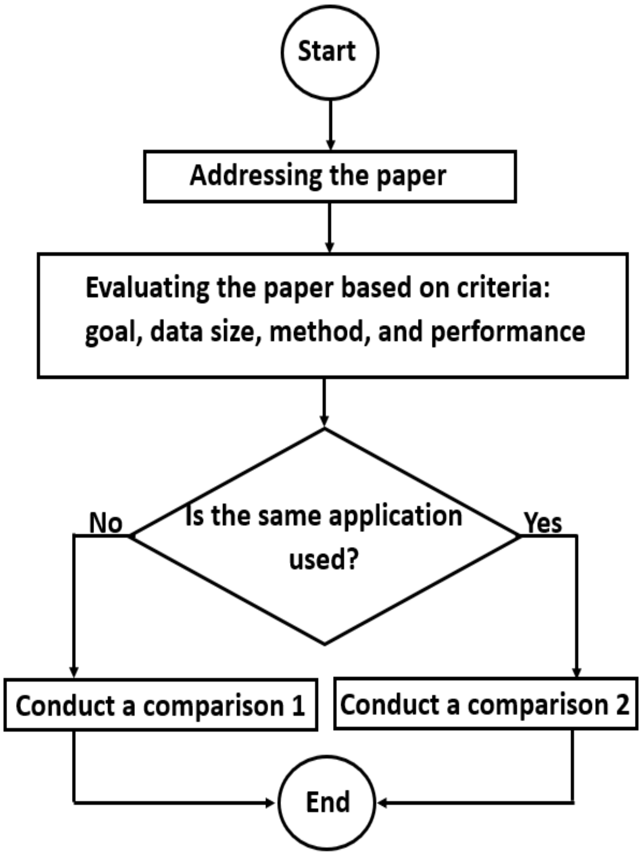

Fig. 1. Flowchart of suggested strategy.

In order to make the comparison sufficiently effective, it should be conducted depending on a strong criteria. In this study, we select specific features (as criteria) for the comparison, they are the goal, data size, method, and performance. In Table 1, the selected criteria are summarized.

Table 1: Criteria of the comparison between similar studies that used machine learning techniques

No.	*Criteria*	*Definition*
1	Goal	indicates to predict of diseases or hospital readmission.
2	Data size	refers to the number of instances stored in the used dataset
3	Method	indicates to the machine learning algorithms that are used for achieving the goal.
4	Performance	refers to the evaluation of the results in terms of the accuracy or ROC curves.

Many investigators tried to use techniques of machine learning in predictions diseases and problems of healthcare, but few researchers employed these techniques to predict hospital readmission. For example, there studies employed machine learning algorithms for predicting diseases such as chronic diseases (e.g. diabetes), breast cancer, thyroid, and etc. The following researchers' group performed machine learning to predict diabetes with different features. Nai-Arun and Sittidech [28] proposed the usage of KNN, DT, and NB with both boosting and bagging to classify diabetes. They utilized an authentic data set from a hospital in Thailand called Sawanpracharak Regional Hospital. The outcomes exposed that bagging along with DT was the preeminent model with the greatest accuracy of 95.312%. Orabi et al. [29] suggested that a system by summing a randomization code with a regression method to predict diabetes with respect to the age, by accuracy rate of 84%. Singh [21] approach was using Bays classifier, KNN, and DT (C4.5) to classify diabetes in medical applications. The research was conducted on the database of Pima Indians Diabetes. Based on the outcomes, Bays classifier obtained the best accuracy of 84%-88%. Also, Xu et al. [30] their suggestion was a prediction model centered on type II diabetes which targets to analyze particular easily obtainable indicators (such as weight, hip, age, waist, etc.). They employed these models RF, AdaBoost, ID3 algorithm, and NB. Depending on their experiments, the best accuracy was 85.00% by RF, this means RF can effectively forecast diabetes risk. Sowjanya et al. [31] suggested that a solution based on mobile/android application. They used DT (J48), SVM, MLP, and NB to classify the gathered data. Consequences showed that assessed with other algorithms, the J48 algorithm gave improved results with ROC areas 92.8%. Thus, using the factual dataset obtained from a well-known hospital in Chhattisgarh, India, J48 was applied to plan the machinery for the forecast of diabetes' mobile application. Another research executed three classification algorithms SVM, NB, and DT to reveal diabetes at an initial phase. Tests were conducted on the Pima Indians Diabetes Database (PIDD) that was obtained from the UCI machine learning repository. Outcome acquired indicates that compared to others, NB had the greatest accuracy of 76.30% [13]. Perveen et al. [32] followed the bagging ensemble learning and AdaBoost using DT (J48) to classify diabetes mellitus patients by means of risk factors of diabetes. The outcome of the experiment proved that the performance of the AdaBoost ensemble technique, on the whole, had improved results as compared to the standalone J48 and bagging.

Some studies attempted employing machine learning according to detect breast cancer and thyroid disease. For example, Hussain et al. [33] used SVM with various kernel functions and MLP to detect breast cancer. Their findings revealed that SVM with a kernel with poor efficiency was better than MLP. Also, Asri et al. [34] claimed that on the datasets of Wisconsin Breast Cancer (original), a comparison of efficiency among different algorithms NB, DT (C4.5), K-NN, and SVM was performed. Their primary goal was to evaluate the accuracy in the classification of the data in regard to performance and efficiency. The outcome proved that the greatest accuracy of 97.13% was given by SVM. Ionita and Ionita [35] executed that the DT (CART) and TreeNet gradient boosting model to classify the patients suffering from thyroid diseases. Their outcomes confirmed that CART obtained the highest accuracy of 96.86%. In addition to this, Oztekin et al. [36] applied MLP, LR, and DTs to improve the forecast of results after combined transplantation of heart-lung. The performance of the predictive MLP and LR model was the best accuracy of 86%. On the other hand, Brindise and Steele [37] sought to use the SimpleLogistic model and BayesNet model in the complex issue of predicting pre-discharge for hospital readmission. Experimental results demonstrated that the best outcome was in SimpleLogistic, with a ROC evaluation of 74.4%. Table 2 summaries the literature review that has the same techniques with different applications in healthcare.

Comparatively few researchers discussed the problem of prediction of hospital readmission, such as Zolfaghar et al. [38] stated that there were two models Multicare Health System (MHS) model and the Yale model. They used RF and LR to forecast the chances of readmission in 30-day for congestive HF cases, this means prediction Risk of readmission (RoR). Also, their experiment based on an imbalance solution (OS) or with no OS. The outcome for the MHS model was 58.39% accuracy of LR with OS, 77.88% for LR with no OS, 77.90% for RF with no OS, 77.96% for RF with OS. On the other hand, the best accuracy was for the Yale model 78.03% for LR with no OS. Hosseinzadeh et al. [39] applied two models RF and NB to predict readmission, where they obtained a lot of data from the administration from the claim form. Consequences confirmed that NB obtained the best ROC of 84%. Another study discussed the prediction of readmission for HF patients by means that comparing the efficiency of various algorithms of machine learning. They used LR, Poisson Regression, RF, RF to LR, RF to SVM and Boosting. Results proved that as compared to LR, the greatest operating machine learning model, RFs, showed the progress of 17.8% for forecasting 30-day all-cause readmission [40]. Zheng et al. [5] suggested that different techniques to forecast the risk of readmission of (HF) patients, for example, NNs, SVM and RF. Outcomes proved that the SVM model achieved the highest accuracy of 78.4%. Finally, Alajmani and Elazhary in [7] discussed the issue of predicting the probability of hospital readmission and applied the following DT, SVM, MLP, NB, and LR.

Their findings revealed that SVM obtained the best accuracy of 95.22%. All literature surveys related to predicting the possibility of hospital readmission are summed up in Table 3.

to their known widespread negative effect on the finances and patient loads in healthcare systems globally [1]. Increased readmission to hospitals disturbs the normal life of patients and results in adverse outcomes for the healthcare systems [2].

Table 2: Summary of Related Work (Same Techniques with Different Applications)

Author	Goal	Data size	Method	Performance
Nai-Arun and Sittidech [28]	Prediction of diabetes	48,763 records diabetes dataset	NB, KNN, DT Bagging and Boosting	Bagging with DT (accuracy 95.312%)
Orabi et al. [29]	Prediction of diabetes	N/A diabetes dataset	DT	(accuracy 84 %)
Singh [21]	Prediction of diabetes	N/A diabetes dataset	KNN, Bays classifier and DT (C4.5)	Bayer classifier (accuracy 84%-88%)
Xu et al. [30]	Prediction of type II diabetes	403 testers diabetes dataset	RF, NB, ID3 and AdaBoost	RF (accuracy 84.13%)
Sowjanya et al. [31]	Prediction of diabetes	145 instances diabetes dataset	DT (J48), NB, SVM and MLP	J48 (ROC 92.8%)
Sisodia and Sisodia [13]	Prediction of diabetes	768 instances diabetes dataset	DT, SVM and NB	NB (accuracy 76.30%)
Hussain et al. [33]	Prediction of breast cancer	N/A cancer dataset	SVM and MLP	SVM (accuracy 97.51%)
Asri et al. [34]	Prediction of breast cancer	699 instances breast cancer dataset	SVM, DT (C4.5), NB and KNN	SVM (accuracy 97.13%)
Oztekin et al. [36]	Prediction of results following combined transplantation of heart-lung	16,604 cases hurt-lung dataset	DT, NNs (MLP) and LR	MLP and LR (accuracy 86%)
Ionita and Ionita [35]	Prediction of thyroid diseases	756 records thyroid dataset	DT (CART) and Gradient Boosting (TreeNet)	CART (accuracy 96.86%)
Perveen et al. [32]	Prediction of diabetes	667,907 records diabetes dataset	AdaBoost and Bagging using DT (J48)	AdaBoost (accuracy N/A)
Brindise and Steele [37]	Prediction of pre-discharge of hospital readmission	145,000 encounters	LR (SimpleLogistic) and BayesNet	SimpleLogistic (ROC 74.4%)

Table 3: Summary of Related Work (Same Techniques with Same Applications)

Author	Goal	Data size	Method	Performance
Zolfaghar et al. [38]	Prediction of readmission	6739 patients HF dataset	RF and LR	LR with no OS (accuracy 78.03%)
Hosseinzadeh et al. [39]	Prediction of readmission	619,274 hospital discharges	NB and DT	NB (ROC 84.00%)
Mortazavi et al. [40]	Prediction of readmission	1653 patients HF dataset	LR, Poisson Regression, RF, RF to LR, RF to SVM and Boosting	RF (17.8% improvement over LR)
Zheng et al. [5]	Prediction of readmission	1641 instances HF dataset	NN, RF and SVM	SVM (accuracy 78.4%)
Alajmani and Elazhary [7]	Prediction of readmission	3090 instances diabetes dataset	LR, MLP, SVM, DT and NB	SVM (accuracy 95.22%)

IV. RESULT AND DISCUSSION

This study endeavors to clarify the following:

1) Elucidating different machine learning techniques that were broadly applied in the healthcare sector.
2) This paper predominantly concentrated on these algorithms namely; artificial neural network, k-nearest neighbor, linear discriminant analysis, logistic regression analysis, support vector machine, random forest, naive Bayesian classifier, decision tree, AdaBoost, and gradient boosting.
3) Essentially focusing on the usage of the above techniques in operations of prediction particularly in prediction of possibility hospital readmission.
4) Each application has a specific dataset and various features, so the performance differs from one application to another.
5) Depending on the objective the used techniques vary for predicting process.
6) The performance can be seen as the major metric to decide which method can be used and provide effective results.

V. CONCLUSION

To summarize as it is known that the hospital readmission readmits the patient within 30 days to the hospital. This paper discussed the importance of hospital readmission in the healthcare sector. Also, it clarified that machine learning can be useful in the healthcare sector in operations of prediction whether in the prediction of diseases or hospital readmission.

This paper revealed different techniques of machine learning that were widely used in this area. It was crystal clear that the performance of techniques differs from one scenario to another. Despite the fact that this research elucidated that there were multiple studies which used machine learning techniques in the healthcare domain, there were comparatively few studies which discussed the prediction of hospital readmission by means of machine learning techniques. Furthermore, this paper clarifies some benefits that can be reflected in the following recommendations: Firstly, if the goal is the prediction of diabetes, it is recommended to use bagging and DT where the accuracy is 95.31%. Secondly, if the goal is the prediction of breast cancer, it is recommended to use SVM where the accuracy is 97.51%. Finally, if the goal is the prediction of hospital readmission, it is recommended to use SVM where the accuracy is 95.22%. Putting all the previous findings into consideration, the topic can be interested enough to encourage researchers to devote their researches and efforts towards this field. In future work, we intend to execute all these models of machine learning in the prediction of hospital readmission and assess which is the best technique for this prediction.

REFERENCES

1. S. B. Roy et al., "Dynamic Hierarchical Classification for Patient Risk-of-Readmission," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 2015-Augus, 2015, pp. 1691–1700.
2. C. K. McIlvennan, Z. J. Eapen, and L. A. Allen, "Hospital Readmissions Reduction Program," Circ. Hear. Assoc., vol. 131, no. 20, pp. 1796–1803, 2015.
3. C. Xiao, T. Ma, A. B. Dieng, D. M. Blei, and F. Wang, "Readmission Prediction via Deep Contextual Embedding of Clinical Concepts," PLoS ONE. Public Libr. Sci., vol. 13, no. 4, pp. 1–15, 2018.
4. C. Van Walraven, C. Bennett, A. Jennings, P. C. Austin, and A. J. Forster, "Proportion of Hospital Readmissions Deemed Avoidable: A Systematic Review," Cmaj, vol. 183, no. 7, pp. E391–E402, 2011.
5. B. Zheng, J. Zhang, S. W. Yoon, S. S. Lam, M. Khasawneh, and S. Poranki, "Predictive Modeling of Hospital Readmissions Using Metaheuristics and Data Mining," Expert Syst. with Appl., vol. 42, no. 20, pp. 7110–7120, 2015.
6. J. Futoma, J. Morris, and J. Lucas, "A comparison of models for predicting early hospital readmissions," J. Biomed. Inform., vol. 56, pp. 229–238, 2015.
7. S. Alajmani and H. Elazhary, "Hospital Readmission Prediction Using Machine Learning Techniques: A Comparative Study," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 4, pp. 212–220, 2019.
8. I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research," Comput. Struct. Biotechnol. J., vol. 15, pp. 104–116, 2017.
9. P. Chowriappa, S. Dua, and Y. Todorov, "Introduction to Machine Learning in Healthcare Informatics," in Machine Learning in Healthcare Informatics.Springer, 2014, pp. 1–23.
10. T. Mitchell, "Machine Learning, McGraw-Hill Higher Education," New York, 1997.
11. A. H. Karp, "Using Logistic Regression to Predict Customer Retention," in Proceedings of the Eleventh Northeast SAS Users Group Conference. http://www. lexjansen. com/nesug/nesug98/solu/p095. pdf, 1998.
12. V. Vapnik, "The nature of statistical learning theory," New York: Springer, 1995.
13. D. Sisodia and D. S. Sisodia, "Prediction of Diabetes Using Classification Algorithms," Procedia Comput. Sci., vol. 132, no. Iccids, pp. 1578–1585, 2018.
14. R. A. Fisher, "The Statistical Utilization of Multiple Measurements," Ann. Eugen. Online Libr., vol. 8, no. 4, pp. 376–386, 1938.
15. G.-D. Wu and Z.-W. Zhu, "Linear-Discriminant-Analysis-Based Type-2 Fuzzy Neural-Network for Speech Detection and Recognition," J. Inf. Sci. Eng., vol. 34, no. 6, pp. 1425–1440, 2018.
16. S. F. Sabbeh, "Machine-Learning Techniques for Customer Retention: A Comparative Study," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 2, pp. 273–281, 2018.
17. S. RAY, "6 Easy Steps to Learn Naive Bayes Algorithm with codes in Python and R," Analytics Vidhya, 2017. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/. [Accessed: 26-Oct-2019].
18. F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, and J. Havel, "Artificial Neural Networks in Medical Diagnosis," J. Appl. Biomed., vol. 11, no. 2, pp. 47–58, 2013.
19. N. Jothi, N. A. Rashid, and W. Husain, "Data Mining in Healthcare - A Review," Procedia Comput. Sci., vol. 72, pp. 306–313, 2015.
20. K. Shailaja, B. Seetharamulu, and M. A. Jabbar, "Machine Learning in Healthcare: A Review," 2018 Second Int. Conf. Electron. Commun. Aerosp. Technol., no. Iceca, pp. 910–914, 2018.
21. A. Singh, "Comparing Data Mining Algorithms for Diabetes Disease Prediction," Int. J. Contemp. Technol. Manag., 2018.
22. R. K. Sharma, V. Sugumaran, H. Kumar, and M. Amarnath, "A Comparative Study of Naive Bayes Classifier and Bayes Net Classifier for Fault Diagnosis of Roller Bearing Using Sound Signal," Int. J. Decis. Support Syst., vol. 1, no. 1, p. 115, 2015.
23. R. E. Schapire, "Explaining Adaboost," in Empirical inference. Springer, 2013, pp. 37–52.
24. Y. Freund and R. E. Schapire, "Experiments with A New Boosting Algorithm," in Icml. Citeseer, 1996, pp. 148–156.
25. G. Chirici et al., "Stochastic Gradient Boosting Classification Trees for Forest Fuel Types Mapping Through Airborne Laser Scanning and IRS LISS-III imagery," Int. J. Appl. Earth Obs. Geoinf., vol. 25, no. 1, pp. 87–97, 2013.
26. L. Breiman, "Random Forests," Mach. Learn. Springer, vol. 45, no. 1, pp. 5–32, 2001.
27. A. Hazra, S. K. Mandal, A. Gupta, A. Mukherjee, and A. Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques : A Review," Adv. Comput. Sci. Technol., vol. 10, no. 7, pp. 2137–2159, 2017.
28. N. Nai-Arun and P. Sittidech, "Ensemble Learning Model for Diabetes Classification," Adv. Mater. Res., vol. 931–932, pp. 1427–1431, 2014.
29. K. M. Orabi, Y. M. Kamal, and T. M. Rabah, "Early Predictive System for Diabetes Mellitus Disease," vol. 1, pp. 420–427, 2016.
30. W. Xu, J. Zhang, Q. Zhang, and X. Wei, "Risk Prediction of Type II Diabetes Based on Random Forest Model," in 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). IEEE., 2017, pp. 382–386.
31. K. Sowjanya, A. Singhal, and C. Choudhary, "MobDBTest: A Machine Learning Based System for Predicting Diabetes Risk Using Mobile Devices," in 2015 IEEE International Advance Computing Conference (IACC). IEEE., 2015, pp. 397–402.
32. S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," Procedia Comput. Sci., vol. 82, no. March, pp. 115–121, 2016.
33. M. Hussain, S. K. Wajid, A. Elzaart, and M. Berbar, "A Comparison of SVM Kernel Functions for Breast Cancer Detection," in 2011 Eighth International Conference Computer Graphics, Imaging and Visualization, 2011, pp. 145–150.
34. H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," Procedia Comput. Sci., vol. 83, no. Fams, pp. 1064–1069, 2016.
35. I. Ionita and L. Ionita, "Applying Data Mining Techniques in Healthcare," Stud. Informatics Control, vol. 25, no. 3, pp. 385–394, 2016.
36. A. Oztekin, D. Delen, and Z. (James) Kong, "Predicting the Graft Survival for Heart-Lung Transplantation Patients: An Integrated Data Mining Methodology," Int. J. Med. Inform., vol. 78, no. 12, pp. 84–96, 2009.

37. L. R. Brindise and R. J. Steele, “Machine Learning-based Pre-discharge Prediction of Hospital Readmission,” in 2018 International Conference on Computer, Information and Telecommunication Systems (CITS)., IEEE, 2018, pp. 1–5.
38. K. Zolfaghar, N. Meadem, A. Teredesai, S. B. Roy, S. C. Chin, and B. Muckian, “Big Data Solutions for Predicting Risk-of-Readmission for Congestive Heart Failure Patients,” in 2013 IEEE International Conference on Big Data.IEEE., 2013, pp. 64–71.
39. A. Hosseinzadeh, M. Izadi, A. Verma, D. Precup, and D. Buckeridge, “Assessing the Predictability of Hospital Readmission Using Machine Learning,” in Proceedings of the Twenty-Fifth Innovative Applications of Artificial Intelligence Conference Assessing, 2013, pp. 1532–1538.
40. B. J. Mortazavi et al., “Analysis of Machine Learning Techniques for Heart Failure Readmissions,” Circ. Cardiovasc. Qual. Outcomes. Am Hear. Assoc., vol. 9, no. 6, pp. 629–640, 2016.

AUTHORS PROFILE

Samah Hazzaa Alajmani received her Bachelor degree from the Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia. She earned a Master degree in Information Technology from the Queensland University of Technology, Brisbane, Australia. She is currently a lecturer at Taif University, Taif, Saudi Arabia, and is pursuing her PhD degree in King Abdulaziz University, Jeddah, Saudi Arabia.

Prof. Kamal Jambi received his M.S from Michigan State University, East Lansing in 1986 and Ph. D. from Illinois Institute of technology, Chicago in 1991, both in Computer Science. He is currently a Professor of Computer Science at King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include AI, Deep learning, Blockchain, Resilience and Bigdata.