Turkish Journal of Computer and Mathematics Education Vol.12 No.10 (2021), 5277-5284

Research Article

Predicting a Risk of Diabetes at Early Stage using Machine Learning Approach

Sanskruti Patel[1], Rachana Patel[2], Nilay Ganatra[3], Atul Patel[4]

[1]Fculty of Computer Science and Applications, Charotar University of Science and Technology, Changa
[2]Fculty of Computer Science and Applications, Charotar University of Science and Technology, Changa
[3]Fculty of Computer Science and Applications, Charotar University of Science and Technology, Changa
[4]Fculty of Computer Science and Applications, Charotar University of Science and Technology, Changa
[1]sanskrutipatel.mca@charusat.ac.in,[2] rachanapatel.mca@charusat.ac.in, [3]nilayganatra.mca@charusat.ac.in, [4]atulpatel.mca@charusat.ac.in

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: Diabetes, known also as Diabetes mellitus, is a prolonged disease that distresses a way body uses to take up sugar or glucose. It can be described as a condition occurs when the blood sugar of a body is too high. The cases of diabetes are rapidly raising in low- and middle-income countries as compare to high income countries. Advancement in the field of Information and Communication Technology (ICT) are revolutionizing many sectors including health care and medical technologies. It plays a critical role in improvising services offered to patients and hospitals. Machine learning is a field of Artificial Intelligence that makes computer system to learn from data, identify the patterns and take decisions without human intervention. Machine learning can be applied on medical datasets to detect and diagnose diseases in an effective and accurate manner. This research focuses on applying machine learning classifiers on the publicly available dataset that contains signs and a symptom of either a person is diabetic or non-diabetic. For that, a dataset for early-stage diabetes risk prediction is acquired from UCI machine learning repository. The well-known machine learning classifiers i.e., Naive Byes, Random Forest, Support Vector Machine and Multilayer Perceptron are experimented on the dataset and the results are analysed. Finally, the result shows that the Random Forest provides the highest values i.e. 0.975 for precision, recall and F-measure respectively. Multiplayer perceptron also works well with 0.96 precision value, 0.963 recall value and 0.964 F-measure value, respectively.

Keywords: Diabetes, Machine Learning, Supervised Learning, Classification

1. Introduction

Diabetes, commonly known as Diabetes mellitus, is a chronic disease that affects a way body uses to take up sugar or glucose. It is a condition occurs when the blood sugar of a body is too high. Blood glucose is one of the main sources of energy for a human body [1]. In human body, pancreas is an organ which generates a special hormone, called Insulin. Pancreas releases Insulin into our blood stream and it helps the glucose get into the cells. A situation when the pancreas is unable to produce the Insulin or when the human body is unable to use the Insulin properly is called the diabetes. Diabetes may cause many health complications like heart attack, blindness, kidney failure, stroke etc. Moreover, people suffered from diabetes fill increase hunger, blurry vision, fatigue weight loss etc.

As per WHO factsheet [2], due to diabetes, a 5% increase is observed during 2000 and 2016 in premature mortality. Also, in 2014, 8.5% of adults aged 18 years and older are suffered from diabetes. It is the direct cause of death of 1.6 million people in 2016, emerged as a seventh leading cause of death. In 1980, 108 million people had diabetes that is increases to 422 million in 2014. The cases of diabetes are rapidly raising in low- and middle-income countries as compare to high income countries. Diabetes can be treated with medications and other remedies and its consequences can be avoided.

The diabetes can be classified into three categories: gestational diabetes, type-1 diabetes and type-2 diabetes [3]. Gestational diabetes is a special type of diabetes that occurs during pregnancy. There are mostly no reported symptoms found but it can be diagnosed during prenatal screening. Woman suffering with gestational diabetes may face complication during pregnancy and at the time of delivery. Also, the woman and their children are facing the high risk of having type 2 diabetes in future. Type 1 diabetes is a condition where the pancreas stops to produce insulin, or a little insulin was produced. It is diagnosed in paediatric population and known as juvenile diabetes. The patient suffering from type 1 diabetes is very less in number as compare to type 2 diabetes. When human body not able to produce the enough insulin or even it fights with insulin is known as type 2 diabetes. It affects the way the human body deals with blood sugar i.e., blood glucose. The majority of the type 2 diabetes patients are middle-aged or even older. It is the most common type of diabetes and it is also known as lifestyle disease.

Advancement in the field of Information and Communication Technology (ICT) are revolutionizing many sectors including health care and medical technologies. It plays a critical role in improvising services offered to patients and hospitals. It provides efficient and effective ways for treatment, diagnosis, information access and storage, drug delivery etc. In last few decades, many advanced devices and technology are invented for medical

sector. ICT in medical sector provides many advantages like improvising quality of patient care, reduction in administration cost, accuracy and timely services etc [4].

Artificial Intelligence (AI) is a field of ICT that concerned with building computer machines in a way that they can perform tasks without human intervention. It is the field that provides the algorithms that can mimic the human intelligence. Machine learning is a field of AI that makes computer system to learn from data, identify the patterns and take decisions without human intervention [5]. Machine learning can be applied on medical datasets to detect and diagnose diseases in an effective and accurate manner.

Timely diagnosis of the diabetes is considered as a challenging problem and there are many parameters that effects on prediction of the diabetes at an early stage. This research focuses on applying machine learning classifiers on the publicly available dataset that contains signs and symptoms of either a diabetic or non-diabetic patient. The result of experiment shows the possibility to occur diabetes at an early stage. The aim of this paper is to propose an automated way to identify the risk of diabetes at an early stage using machine learning classifiers. For that, various state-of-the-art machine learning classifiers are experimented on the dataset and results are analysed.

2. Literature Survey

Several researchers have worked for applying machine learning methods and algorithms on medical datasets.

Kavakiotis et al. [6] used various data mining and machine learning methods for identifying diabetes and employed various algorithms to study. Shetty et al. [7] applied various techniques like Naïve Bayes and K-nearest neighbour for predicting diabetes. Kareem et al. [8] performed a comparative analysis for predicting diabetes at early stage. To find the optimum classifier, they have used various machine learning classifiers like multilayer perceptron and radial basis function. T. Mahboob Alam et al. [9] proposed a predictive model for early prediction of diabetes. For that, they have experimented K-means clustering, random forest and artificial neural network. M. Kavitha and S. Subbaiah [10] implemented classification algorithms in their research and predicted diabetes disease.

K. M. Almustafa [11] proposed a prediction model for heart disease and performed a sensitivity analysis. Various classifiers like Decision Table, Naive Bayes, J48 etc. are used to classify the heart disease. C. Krittanawong et al. [12] utilized machine learning algorithms for prediction of cardiovascular disease. They found that SVM is one of most promising classifiers for predicting the cardiovascular disease. X. Tian et al. [13] used machine learning algorithms to predict HBsAg seroclearance. They have found XGBoost as a most appropriate classifier during their research. D. Sisodia and D. S. Sisodia [14] developed a system for prediction of diabetes using classification algorithms. They have also used Receiver Operating Characteristic (ROC) curves to verify the results obtained during an experiment.

3. Background

A training of computer algorithms which improves automatically through the use of data and experience is known as Machine Learning (ML) [15]. Machine Learning (ML) is the subset of Artificial Intelligence (AI). Machine Learning is a field of AI that allows the computer system to learn automatically and improve through experience without need of being explicitly programmed [16]. It mainly focuses on the development of programs which can manipulate data and used it for learning without instructions. The fundamental aim of the machine learning algorithm is to learn the intrinsic patterns of the data provided without any human intervention. The data used by the Machine Learning algorithms to build a model is known as training data [17]. Training data is used to make prediction or decision without being explicitly programmed. With availability of the large amount of data, with Machine Learning algorithms, it is possible to develop predictive models that can manipulate, understand and analyse complex data to identify needful insight and provides more precise results [18]. The Machine Learning algorithms are important for the reasons like massive data, improve decision making, hidden patterns and trends in data and solve difficult problem as explained in figure 1.

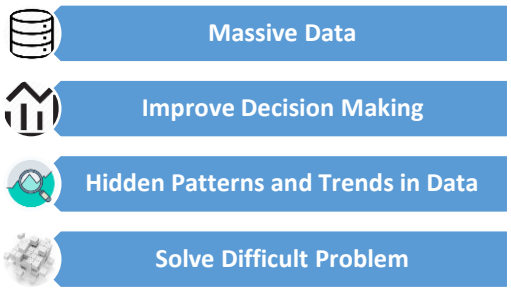

Figure 1. Importance of Machine Learning Algorithms

Massive Data: With the excessive amount of data, Machine Learning algorithms can be applied on the data to structure, analyse and fetch required insight from the data. It is use to solve complex problems by eliminating the need of human expertise.

Improve Decision Making: Machine Learning algorithms can be applied on huge data to make the better decision compared with traditional decision-making process.

Hidden Patterns and Trends in Data: Identifying uncovers patterns and obtaining key insight from the data is the basic function of the Machine Learning. Predictive models along with usage of statistical techniques allow Machine Learning algorithms of manipulate and explore data deeply.

Solve Difficult Problem: Machine Learning algorithms are capable of solving complex problems like weather forecasting**,** stock price prediction, and linkage of genes, self-driving cars and many more.

Based on the method of training, the Machine Learning algorithms are broadly classified into two categories: Supervised Learning and Unsupervised Learning.

A. Supervised Learning

Supervised Learning is the process of training Machine with well labelled data. In supervised learning, algorithm is provided with the labelled input examples along with their desired outputs [19]. Supervised learning is the category of the machine learning algorithm which required a guide. The labelled data is used to train the Machine learning algorithm to understand the hidden patterns of the data. The labelled dataset is known as training dataset. After providing sufficient training to the model, the model applies learned knowledge on the real world. Supervised learning problems include classification and regression. The classification problem goal is to identify the group to which input belong to. Like classification, regression is mapping input with equivalent output. However, outputs of the classifications are discrete values while for regression are a numerical value.

B. Unsupervised Learning

In unsupervised learning Machine Learning model is trained using unlabelled data and let model to learn the information form the data without any explicit guidance [19]. In this category, machine learns from the data for which output is not known. Clustering is the type of unsupervised learning technique. For clustering problem, only the input data is provided to without explicit output data. Unsupervised machine learning model is capable of grouping data cluster wise, but it won't tell the type of the data [20]. Unsupervised learning eliminates the human need for labelling data.

4. Methods and Materials

4.1 Methods

For an experiment, we have considered four machine learning classifiers that include Naïve Bayes, Multilayer Perceptron, Support Vector Machine (SVM), and Random Forest.

4.1.1 Naïve Bayes

The group of classification algorithms based on Bayes Theorem are commonly known as Naïve Bayes. The Naïve word in the algorithm presents the general understanding that the presence of a particular feature in the class is completely independent or unrelated to the occurrence of some other feature in the similar class as displayed in figure 2[21]. Bayes Theorem is well known mathematical formula which is used to determine the conditional probability of an event, based on the previous knowledge about the condition related to the event [22].

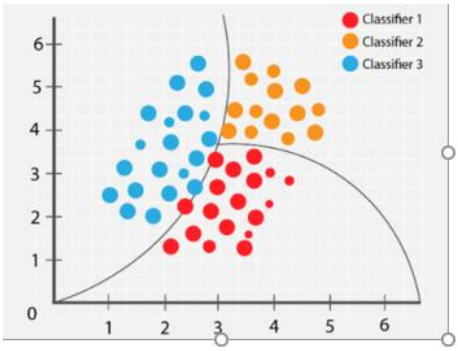

Figure 2. Naïve Bayes Classifier [21]

Mathematically Bayes Theorem can be represented as below.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Here, P(A|B) is the posterior probability of the target class. The prior probability of the class is determined by P(A). The prior probability of predictor is defined by P(B). The likelihood which is the probability of predictor given class is defined by P(B|A).

4.2.2 Support Vector Machine (SVM)

A Support Vector Machine, known most commonly as SVM, is classification algorithm used for two-group classification problems. It falls under the supervised learning category. After providing sufficient training using labelled data for each class, it can be able to categorize new unseen data. The aim of the support vector machine algorithm is to identify the hyperplane in an N-dimensional space. Here N represents the number of features. These features uniquely separate the data points [23]. Two classes' data points can be separated based on number possible hyperplanes; however, objective is to determine the plan with maximum margin. Increased margin distance helps in classifying future data points with more confidence. 'Support Vectors' is the idea used by SVM to identify the closest points to the hyperplane. The working of SVM is illustrated using figure 3.

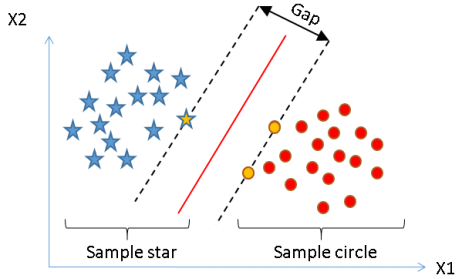

Figure 3. Support Vector Machine [23]

4.3.3 Artificial Neural Network (ANN)

Artificial Neural Network imitates the functioning of biological nervous system of human body. Similar to human brain ANN algorithms learn through experiences and examples. It learns from the past examples and experiences. It then applies learned knowledge during testing [24]. ANN can make computer more powerful in solving the problems which are not known to the human. ANN helps in the problems like regression and classification. The conventional structure of ANN is described in figure 4.

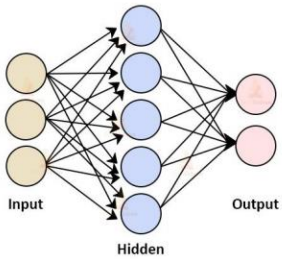

Figure 4. Artificial neural network (ANN) [25]

Neurons are arranged in the multiple layers in the Neural Network. Each layer of the network is connected with the layers on both sides in which on side layer engaged in receiving input signals which are useful for the network

to learn or process and other side layer as responses or output for the information [25]. There are multiple hidden layers between these two layers. The interconnections between all layers including hidden and output layer are known as weight. These weights indicate the strength in between the units. ANN is also referred to as multilayer perceptron network.

4.4.4 Random Forest

Another most frequently used supervised learning technique is Random Forest that supports regression and classification. In order to improve the performance, it combines the multiple classifiers. These classifiers are used to solve the difficult problem. The concept is called ensemble learning. On different subsets of the given dataset, Random Forest classifier contains a set of decision trees. By this way, it takes the average of each decision trees output to improve the prediction accuracy on the given dataset [26]. Hence, instead of relying on single decision tree for the prediction, Random Forest takes prediction from multiple trees and based on the maximum number of votes of prediction, and it predict the final output. The figure 5 depicts the working of random forest algorithm. The following steps illustrate the working of the random forest algorithm.

Step-1: Select random data points form the training dataset.

Setp-2: Develop the subset of decision tree associated with the select data points.

Step-3: Select the number for decision tree that you want to build.

Step-4: Repeat step 1&2

Step-5: Get the prediction of each decision tree for the given data points and assign the new data points to category that wins the majority votes.

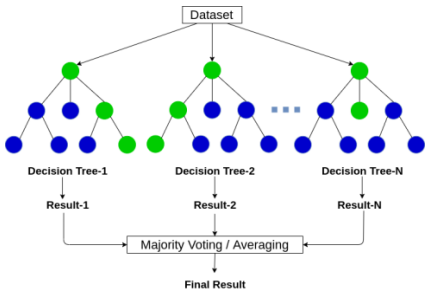

Figure 5. Random forest [26]

4.2 Materials

The dataset considered in this research is obtained from UCI machine learning repository [27] [28]. The dataset has been collected from the patients of Sylhet Diabetes Hospital. It is located in Sylhet, Bangladesh. The dataset was collected using direct questionnaires and it was approved by the doctor [29]. The dataset contains total 16 distinct attributes that help to predict the early-stage diabetes risk prediction. There are total 520 instances are available in the dataset. The dataset has the target attribute i.e., class that contains values either positive or negative. Among 520 instances, 320 are positive class and 200 are negative class. The table 1 contains the attribute information of the dataset.

Table 1. Attribute Information of the Dataset

Attribute	**Description**
Age	Age of a person. Positive value that ranges between 16-90
Gender	Male and Female
Polyuria	Yes or No
Polydipsia	Yes or No
Sudden weight loss	Yes or No
Weakness	Yes or No
Polyphagia	Yes or No
Genital thrush	Yes or No
Visual blurring	Yes or No
Itching	Yes or No
Irritability	Yes or No
Delayed healing	Yes or No
Partial paresis	Yes or No
Muscle stiffness	Yes or No

Alopecia	Yes or No
Obesity	Yes or No
Class	Positive or Negative

4.3 Performance Evaluation Metrics

State-of-the-art performance evaluation metrics are used during the research that helps to analyse the results obtained during an experiment. They are precision, recall, accuracy, F-Measure and kappa statistics [30].

The value of Kappa statistic is a metric used to compare the accuracy obtained with the accuracy expected. The formula is calculated as given below.

$$\text{Kappa} = (\text{observed accuracy} - \text{expected accuracy})/(1 - \text{expected accuracy})$$

Precision defines the ratio of true positives to the total number of predictions as positives. To calculate the precision, following formula is used.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

Recall provides a fraction of the actual positive cases which are correctly identified. It can be calculated using the following equation.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN})$$

F-measure is a combination of precision and recall into a single measure. It can be represented as the harmonic mean of precision and recall. The following equation is used to calculate the F-measure.

$$\text{F} - \text{measure} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

Another performance metrics used most commonly is accuracy. Accuracy is the number of correct predictions made to all prediction ratios. Accuracy can be calculated using the following formula.

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

Here, FP is the false positive and FN is the false negative, whereas TP is the true positive and TN is the true negative,

5. Experiment and Result Analysis

As mentioned above, the dataset has been acquired from UCI machine learning repository. It contains total 16 attributes describing the symptom and the class attributes specifies whether a person is positive or negative. It contains 520 instances approved by doctor. An experiment is conducted using WEKA; open-source software developed at University of Waikato. It provides tools for visualization, data pre-processing and several machine learning algorithms. It provides an interface for implementing machine learning techniques for real-world problems and applications. Moreover, several performance metrics are used to analyse the results obtained during experiment.

The problem mentioned here i.e., Diabetes classification is considered as a supervised learning problem. Also, the class label used here contains only two possible values. Therefore, it is a binary classification problem. Also, a 10 folds cross validation is selected during experiment. The performance of the various machine learning classification is described in the table 2 and table 3. Table 2 displayed the values obtained during experimented for TP rate, FP rate, Precision, Recall, F-measure and ROC.

Table 2. Performance Evaluation of Different Classifiers

Classifier	**Precision**	**Recall**	**F-measure**	**ROC**	**TP Rate**	**FP Rate**
Naïve Byes	0.878	0.871	0.872	0.945	0.871	0.120
Random Forest	0.975	0.975	0.975	0.998	0.975	0.027
Support Vector Machine	0.921	0.921	0.921	0.918	0.921	0.085
Multilayer Perceptron	0.964	0.963	0.964	0.994	0.963	0.038

From the table 2, it shows that Random Forest classifier works better compare to the other classifiers implemented. It provides 0.975 True Positive that is highest. The value of False Positive is 0.027 that is lowest compare to the other classifiers. Moreover, the value obtained for precision, recall and F-measure are 0.975 that is highest among all other classifiers. Also, it has been observed that Multilayer Perceptron also works well with 0.964 value for precision, 0.963 value for recall and 0.964 value for recall respectively.

The table 3 describes the values for the other performance evaluation metrics used that includes correctly classified instances, incorrectly classified instances, kappa statistic and root mean squared error.

Table 3. Corrected and Incorrected Instances with Kappa Statistic and RMSE

Classifier	**Kappa Statistic**	**Correctly Classified Instances**	**Incorrectly Classified Instances**	**Root Mean Squared Error**
Naïve Byes	0.734	87.1154	12.8846	0.3184
Random Forest	0.9472	97.5	2.5	0.1398
Support Vector Machine	0.8339	92.1154	7.8846	0.2808
Multilayer Perceptron	0.923	96.3462	3.6538	0.1638

The table 3 shows that the Random Forest again provides the highest value for correctly classified instances i.e., 97.5% and lowest value for incorrectly classified instances i.e., 2.5%. The value obtained for kappa statistic is 0.9472 and the value obtained for root mean squared error is 0.1398. Multilayer Perceptron is also performing well and provides 96.3462% correctly classified instances with 0.923 value for kappa statistic.

6. Conclusion

Machine learning approaches work well for diagnosis of various disease. It performs very well with medical datasets. Prediction of diabetes at early stage helps the patient in order to provide appropriate treatment. In this paper, the feasibility of machine earning classifiers for diabetes prediction is discussed. For that, four state-of-the-art machine learning classifiers i.e., Naïve Byes, random forest, Support Vector Machine and Multilayer Perceptron are discussed and implemented on the diabetes dataset obtained from UCI machine learning repository. From the results obtained, it has been observed that Random Forest provides highest results among all other classifiers. Also, multilayer perceptron performs well after a Random Forest for predicting diabetes at the early stage.

References

1. "Diabetes," *Who.int*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes. [Accessed: 11-Apr-2021].
2. E. Crawford, "What is Diabetes?: Notes for a lesson for senior pupils," *J. Inst. Health Educ.*, vol. 1, no. 2, pp. 10–15, 1963.
3. "What is diabetes," *Idf.org*. [Online]. Available: https://www.idf.org/aboutdiabetes/what-is-diabetes.html. [Accessed: 11-Apr-2021].
4. B. Lindberg, C. Nilsson, D. Zotterman, S. Söderberg, and L. Skär, "Using information and communication technology in home care for communication between patients, family members, and healthcare professionals: A systematic review," *Int. J. Telemed. Appl.*, vol. 2013, p. 461829, 2013.
5. "Artificial intelligence and Machine learning made simple," *Marutitech.com*, 30-Sep-2016. [Online]. Available: https://marutitech.com/artificial-intelligence-and-machine-learning/. [Accessed: 11-Apr-2021].
6. 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.
7. D. Shetty, K. Rit, S. Shaikh, and N. Patil, "Diabetes disease prediction using data mining," in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017, pp. 1–5.
8. Kareem, L. Shi, 3lin Wei, and Y. Tao, "A comparative analysis and risk prediction of diabetes at early stage using machine learning approach," *International Journal of Future Generation Communication and Networking*, vol. 13, no. 3, pp. 4151–4163–4151–4163, 2020.

9. T. Mahboob Alam *et al.*, "A model for early prediction of diabetes," *Inform. Med. Unlocked*, vol. 16, no. 100204, p. 100204, 2019.
10. M. Kavitha and S. Subbaiah , "Implementing classification algorithms for predicting chronic diabetes diseases," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6S3, pp. 1748–1751, 2019.
11. K. M. Almustafa, "Prediction of heart disease and classifiers' sensitivity analysis," *BMC Bioinformatics*, vol. 21, no. 1, p. 278, 2020.
12. Krittanawong *et al.*, "Machine learning prediction in cardiovascular diseases: a meta-analysis," *Sci. Rep.*, vol. 10, no. 1, p. 16057, 2020.
13. Tian et al., "Using machine learning algorithms to predict hepatitis B surface antigen seroclearance," *Comput. Math. Methods Med*., vol. 2019, p. 6915850, 2019
14. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms*," Procedia Comput. Sci*., vol. 132, pp. 1578–1585, 2018.
15. 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
16. Akritidis and P. Bozanis, "A supervised machine learning classification algorithm for research articles," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing - SAC '13*, 2013.
17. "Medium," *Towardsdatascience.com*.[Online].Available: https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08. [Accessed: 11-Apr-2021].
18. K. M. Singh, A. Kumar, and R. K. P. Singh, "Role of information and communication technologies in Indian agriculture: An overview," *SSRN Electron. J.*, 2015.
19. Wikipedia contributors, "Machine learning," *Wikipedia, The Free Encyclopedia*, 10-Apr-2021. [Online].Available:https://en.wikipedia.org/w/index.php?title=Machine_learning&oldid=1017026799. [Accessed: 11-Apr-2021].
20. "Diagnosis of Diabetes Using Classification Mining Techniques"," *International Journal of Data Mining & Knowledge Management Process (IJDKP*, vol. 5, no. 1, 2015.
21. S. Ray, "6 Easy Steps to Learn Naive Bayes Algorithm." 2017.
22. F. R. F. Padao and E. A. Maravillas, "Using Naïve Bayesian method for plant leaf classification based on shape and texture features," in *2015 International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*, 2015.
23. K. M. Orabi, Y. M. Kamal, and T. M. Rabah, "Early predictive system for diabetes mellitus disease," in *Advances in Data Mining. Applications and Theoretical Aspects*, Cham: Springer International Publishing, 2016, pp. 420–427.
24. M. Biswas and R. Adlak, "Classification of galaxy morphologies using artificial neural network," in *2018 4th International Conference for Convergence in Technology (I2CT)*, 2018.
25. R. Dharwal and L. Kaur, "Applications of artificial neural networks: A review," *Indian J. Sci. Technol.*, vol. 9, no. 47, 2016.
26. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, pp. 93–104, 2012.
27. "UCI Machine Learning Repository," Uci.edu. [Online]. Available: https://archive.ics.uci.edu/ml/index.php. [Accessed: 11-Apr-2021].
28. "UCI Machine Learning Repository: Early stage diabetes risk prediction dataset. Data Set," Uci.edu. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset. [Accessed: 11-Apr-2021].
29. M. Gupta, D. Konar, S. Bhattacharyya, and S. Biswas, Eds., Computer Vision and Machine Intelligence in Medical Image Analysis: International Symposium, ISCMM 2019. Singapore: Springer Singapore, 2020.
30. White Paper, "Metrics for multi-class classification: An overview," *Arxiv.org. [Online]*. Available: http://arxiv.org/abs/2008.05756v1. [Accessed: 11-Apr-2021].