

# Comparing Different Machine Learning Techniques in Predicting Diabetes on Early Stage <sup>†</sup>

Shweta Yadu <sup>1,\*</sup> , Rashmi Chandra <sup>2</sup> and Vivek Kumar Sinha <sup>3</sup> 

<sup>1</sup> Department of Computer Science Engineering, Raipur Institute of Technology, Raipur 492101, Chhattisgarh, India

<sup>2</sup> Department of Computer Science and Engineering, Chhattisgarh Swami Vivekanand Technical University, Newai, Bhilai 491107, Chhattisgarh, India; 0101chandra.rashmi@gmail.com

<sup>3</sup> Department of Computer Science and Engineering, Noida International University, Greater Noida 203201, Uttar Pradesh, India; vivek.sinha@niu.edu.in

\* Correspondence: y.shweta1316@gmail.com

<sup>†</sup> Presented at the 2nd Computing Congress 2023, Chennai, India, 28–29 December 2023.

**Abstract:** One of the diseases that is constantly spreading and is estimated to cause a significant number of deaths worldwide is diabetes mellitus. It is determined by the quantity of a blood sugar molecule made from glucose. The possibility of this disease has been predicted using a variety of methods. To forecast diabetes at an early stage, adequate and clear data on diabetic individuals are needed. In this study, 520 records from a hospital in Bangladesh with 16 different characteristic numbers were used to make predictions. At UCI, this dataset is accessible to everyone. We used Random Forest, Ada Booster, KNN, and Bagging algorithms after feature selection. Through 10-fold cross-validation, it was discovered that the Random Forest method had the best test accuracy, scoring 97.03% correctly and 95.03% correctly.

**Keywords:** Ada boost; bagging; KNN; machine learning; predictive analysis; random forest



**Citation:** Yadu, S.; Chandra, R.; Sinha, V.K. Comparing Different Machine Learning Techniques in Predicting Diabetes on Early Stage. *Eng. Proc.* **2024**, *62*, 20. <https://doi.org/10.3390/engproc2024062020>

Academic Editors: Geetha Ganesan, Xiaochun Cheng and Valentina Emilia Balas

Published: 22 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Diabetes is one of the illnesses that is now growing at the fastest rate. The World Health Organization estimates that 422 million people worldwide have diabetes. Additionally, it states that non-communicable diseases account for almost 41 million preventable deaths annually, or nearly 71% of all fatalities worldwide. By 2030, non-communicable illnesses will be responsible for 52 million annual deaths if the problem is not addressed. The most common non-communicable diseases are diabetes and hypertension, which account for around 46.2% and 4% of all mortality, respectively [1]. These conditions are frequently caused by an excess of blood glucose, a sugar molecule made from carbs. With the help of the hormone insulin generated by the pancreas, food is broken down into its smallest molecules and nutrients, such as glucose, which are then taken up by all cells with the goal of creating energy. When the body does not create enough insulin, cells cannot absorb glucose [2], which can happen occasionally. It is difficult to diagnose this disease in its early stages because it is mostly a lifestyle-related condition. Usually advanced by the time it is found, it can only be treated with medication, with some patients also requiring insulin injections to regulate their blood sugar levels. Long-term uncontrolled blood sugar levels can cause serious organ damage, including diabetic retinopathy, which impairs vision, diabetic neuropathy, which harms the nerves, diabetic foot, as well as harm to the heart, pancreas, kidneys, and many other important organs [3]. A balanced diet and way of living can help someone manage their blood sugar levels. Since excessive blood sugar levels can gravely affect a person's body, those who have been diagnosed with diabetes must maintain a healthy lifestyle in addition to taking medicine to control their blood sugar. Regular health checks to check for any unexpected changes in the body's blood sugar levels

are the best way to manage a chronic condition like diabetes. Diabetes can be challenging to identify in its early stages, and it can be difficult to predict when it will first appear, even with all these precautions [4]. Making a medical diagnosis is a labor-intensive process that can be fairly difficult. ML in healthcare systems is not just used for diagnosis; it can also be used to forecast drug effects, manage medical data, support doctors, and make decisions, among other things. Healthcare systems built on machine learning can help clinicians obtain results very quickly. Healthcare professionals and information technology experts are now working in the field of ML-based healthcare systems to accelerate processing and provide better results [5]. The term “machine learning” (ML) refers to a variety of statistical methods that let computers learn from their experiences without having to be explicitly programmed. Applications of machine learning (ML) are profoundly changing the healthcare industry.

The UCI laboratory dataset will be used in this study’s general framework for predicting the progression of diabetes, and four different machine learning algorithms will be implemented and compared to determine which one has the highest accuracy.

The construction of this research paper is as follows: Background information on diabetics is presented in Section 1, and previous ML models that have been used to predict diabetes in the past are then presented in Section 2. The proposed method, the dataset description, and the preprocessing that went into this study are all shown in Section 3. Section 4 presents all of the experimental results and comparisons with pertinent literature. Section 5 presents the conclusion.

## 2. Related Work

The 768-record Pima Native Dataset and a range of machine learning algorithms are used by Malini M et al. [6] to facilitate the prediction of diabetes. In the suggested technique for classification and ensemble learning, classifiers from SVM, KNN, Random Forest, decision tree, logistic regression, and gradient boosting are utilized. The highest classification accuracy of 78% was achieved using logistic regression.

The technique proposed by M Asiful Huda et al. [7] outperforms the current findings in terms of recall and precision. The suggested method applies classification algorithms to a few features from a dataset on diabetic retinopathy, such as optical disc diameter, lesion-specific features (microaneurysms, exudates), or the presence of hemorrhages. The features are then collected and used in the final decision-making procedure to establish the presence or absence of diabetic retinopathy. The decision is then made utilizing the support vector machine, logistic regression, and decision tree algorithms. Compared to past experiments, the model’s accuracy rate is substantially higher at 88%.

Ophthalmologists can forecast DR with the help of a computer-aided classification system for exudates suggested by Smitha S Prem and Umesh A.C [8]. The classifications were established using the wavelet decomposition coefficient and LBP, which provide texture information and frequency information, respectively, in an image. The effectiveness of classifiers is assessed using a variety of supervised classification techniques. The KNN classifier has improved the performance of the proposed model with an accuracy of 94% for LBP features using the DIARETDRBI dataset with 89 images, while ANN has improved performance with 100% accuracy for wavelet features.

Salliah Shafi Bhat and Gufran Ahmed Ansari [9] use a machine learning technique to detect diabetes and recommend a healthy diet for diabetic patients using a diet recommendation system (DRS). Numerous machine learning approaches, such as the probabilistic-based naive Bayes (NB), the function-based multilayer perception (MLP), and the decision tree-based Random Forests (RF), are used to develop the machine learning model for the diagnosis of diabetes. Random Forests (RF), the classifier with the best accuracy, achieves 93%.

In Usama Ahmed et al.’s article [10], using a mixed strategy, a machine learning model for predicting diabetes is given. The conceptual framework is based on support vector

machine (SVM) and artificial neural network (ANN) models. These models examine the dataset in order to determine whether a diabetes diagnosis is accurate or not. These models' results act as the fuzzy model's input membership function, which ultimately determines whether or not a diabetes diagnosis is made. With a prediction accuracy of 94.87%, the suggested fused ML model exceeds the previously revealed methods.

A hybrid model based on the top three findings was constructed in this study by Sarra Samet [11], who employed six supervised machine learning classification approaches in combination to diagnose diabetes early on. The research employs the Pima Indians Diabetes Database, which is accessible through UCI's machine learning repository. They are all evaluated based on a variety of metrics. With a 90.62% accuracy rate, the hybrid model stands out against other cutting-edge methods.

Minhaz Uddin Emon [12] used feature extraction to find some features in an effort to predict diabetic retinopathy. The data required for this inquiry were provided via the UCI machine learning repository. In order to assess the performance, sensitivity, selectivity, true positive (tp), false negative (fn), and receiver operating characteristic (roc) curves, this dataset was explored using several machine learning (ML) methodologies. Naive Bayes, sequential minimal optimization (SMO), logistic regression, stochastic gradient descent (SGD), Bagging classifier, J48 classifier, decision tree classifier, and random forest classifier are a few of the machine learning techniques used in this study. The overall model that performs the best is logistic regression.

S. Jyotheeswar and K.V. Kanimozhi [13] presented a study that used innovative decision trees (DT) and SVM to detect diabetic retinopathy (DR), as opposed to support vector machines. To forecast diabetic retinopathy, the new decision tree (N = 10) and support vector machine (N = 10) algorithms were employed. More than 50,000 digitized retinal images from the Kaggle fundus image dataset were used to identify diabetic retinopathy. support vector machine only managed an accuracy of 85.2%, whereas innovative decision tree managed a precision of 92.8%. ( $p = 0.03$ ) is the difference between DT and SVM that is statistically significant. The innovative decision tree method outperforms support vector machine for detecting diabetic retinopathy.

In this study, M. Paliwal and P. Saraswat [14] use controlled machine learning techniques on real data from 520 diabetic patients and probable diabetes patients ranging in age from sixteen to ninety. The Naive Bayes classifier, Light-GBM, and support vector machine (SVM) are some examples of these techniques. The performance of the support vector machine has the highest accuracy when comparing classification and recognition accuracy.

### 3. Methodology

Figure 1 shows how the envisioned system is laid out. An early-stage diabetes risk prediction dataset with patient records was used in the proposed method. The dataset is subjected to Random Forest, Ada boosting, KNN, and Bagging to produce an effective technique.

#### A. Data Collection

Islam et al. [15] created the early-stage diabetes risk prediction dataset (UCI Machine Learning Repository, 2020). Information was gathered from the patient files at the Sylhet Diabetes Hospital in Sylhet, Bangladesh. Diabetes is associated with 520 incidences and 16 characteristics. One continuous characteristic and fifteen categorical attributes are present. Table 1 Dataset description and Table 2 Description of attributes.

**Table 1.** Dataset description.

|                         | Number of Attributes | Number of Instances |
|-------------------------|----------------------|---------------------|
| Diabetics Patients Data | 16                   | 520                 |

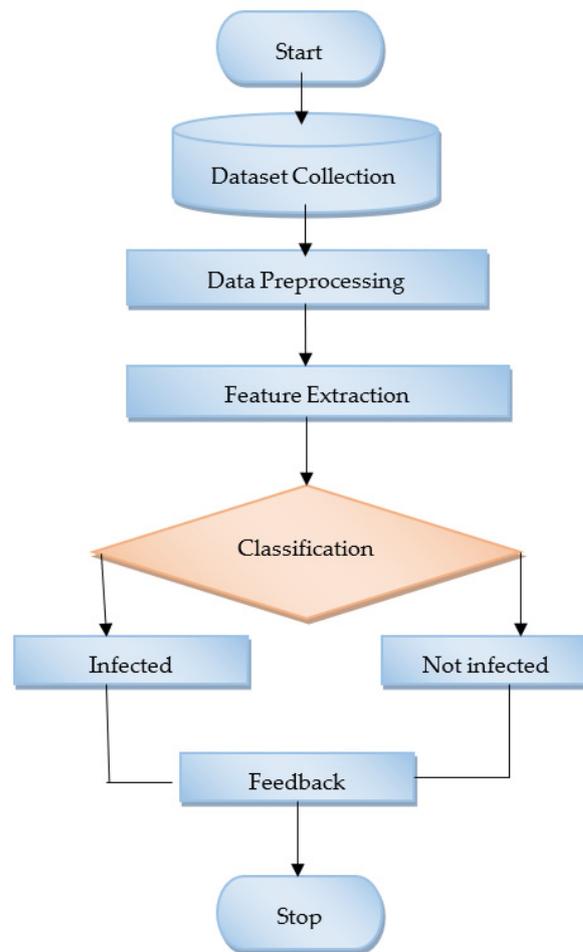


Figure 1. Proposed system architecture.

Table 2. Description of attributes [16].

| S_No. | Features           |                 |         |             |
|-------|--------------------|-----------------|---------|-------------|
|       | Features Name      | Values          | Missing | Non-Numeric |
| 1.    | Age                | 1–100           | No      | No          |
| 2.    | Sex                | Male and Female | No      | Yes         |
| 3.    | Polydipsia         | Yes/No          | No      | Yes         |
| 4.    | Polyuria           | Yes/No          | No      | Yes         |
| 5.    | Sudden weight loss | Yes/No          | No      | Yes         |
| 6.    | Polyphagia         | Yes/No          | No      | Yes         |
| 7.    | Weakness           | Yes/No          | No      | Yes         |
| 8.    | Genital thrush     | Yes/No          | No      | Yes         |
| 9.    | Visual blurring    | Yes/No          | No      | Yes         |
| 10.   | Irritability       | Yes/No          | No      | Yes         |
| 11.   | Itching            | Yes/No          | No      | Yes         |
| 12.   | Partial paresis    | Yes/No          | No      | Yes         |
| 13.   | Delayed healing    | Yes/No          | No      | Yes         |
| 14.   | Muscle stiffness   | Yes/No          | No      | Yes         |
| 15.   | Obesity            | Yes/No          | No      | Yes         |
| 16.   | Alopecia           | Yes/No          | No      | Yes         |
| 17.   | Class              | Yes/No          | No      | Yes         |

## B. Data Preprocessing

1. Importance Graph analysis: Important graphs have been drawn and analyzed to determine which features play an important role in prediction. They graphically describe the feature importance. A technique called “feature importance” values input features according to how well they predict return labels. The reliability and effectiveness of a predictive model in practice can be improved by using feature importance scores, which play a significant role in predictive modeling projects and provide insight into the data, information into the model, and a basis for feature selection.
2. Filter Method for Feature Selection: A filter method has been applied in the dataset to remove redundant data. Instead, using classification algorithms, filter approaches analyze features based on data qualities. Information theory, correlation, distance, consistency, fuzzy sets, and rough sets can all be used as the foundation for filter measures. First, features are chosen and sorted, batch-wise in the multivariate case (handling redundancy naturally) and independently of feature space in the univariate case. The highest ranked features are chosen in the second stage using a performance criterion [17].

### a. Training Dataset

In this approach, the dataset has been split into training and testing data. The data in the proposed dataset have been divided into training and testing data in an 8:2 ratio.

### b. Applied Models

Machine learning algorithms have shown great success in the issues of diabetes prediction. How a machine learning system might be used to identify diabetics is explained in this study. This system makes use of the 16 properties found in the UCI Machine Learning Repository, which is openly accessible. Three different architectures, including Random Forest, Ada boosting, KNN, and Bagging are examined as the core of our research. Detailed explanations of the predefined architecture are provided below.

- Random Forest: The Random Forest algorithm, first proposed by Bierman [18,19], consists of a number of independent classifiers for tree structures, each of which makes a classification prediction. Based on the classification predictions with the highest number of votes from each classifier, the output is predicted. The accuracy increases linearly with the number of trees in the forest, which also removes overfitting problems [20]. This simple machine learning technique typically yields excellent results without hyper-tuning. When using the Random Forest method, the classifier will not overfit the model if there are enough trees in the forest, which is a severe problem that can sabotage results. Missing data issues can be resolved by the Random Forest classifier, which can also be more beneficial for categorical values [21].
- KNN: The term K Nearest Neighbors (KNN) refers to the sample’s K closest neighbors. The idea behind the approach is that you can view the category of K known instances that are closest to the unknown instance when it is necessary to discover the category of an unknown instance [22]. The category that makes up the greatest percentage of the K instances is counted and is assumed to be the category of an unidentified case. Different K values have a significant impact on classification when chosen. To determine the categorization, the distance between each instance and the sample point must be determined. Three steps make up the specific implementation: locating the sample’s K closest neighbors first; then Third, choose

the category with the highest percentage of categories in the closest neighbor as the classification category [23] by calculating the proportion of nearest neighbor categories. KNN algorithms use data to categorize new data elements entirely based on similarity metrics. The class with the closest neighbors is given the statistics [24].

- Ada booster: The AdaBoost (adaptive boosting) technique was created by Yoav Freund and Robert Shapire in 1995 to create a strong classifier out of a collection of poor classifiers. The Boosting family of algorithms includes AdaBoost (Adaptive Boosting) [25]. This kind of learner focuses more on incorrectly classified samples during training, modifies the sample distribution, and repeats this process until the weak classifier has undergone a predetermined amount of training, at which point learning is complete [26]. By retaining a collection of weights across training data and adaptively adjusting them after each weak learning cycle, the AdaBoost algorithm produces a series of poor learners. The weights of the training samples that will be misclassified by the weak learner that is currently in use will be increased, while the weights of the training samples that the learner will correctly classify will be dropped [27].
- Bagging: Bagging is another example of an ensemble technique in which a group of weak learners is combined to produce a strong learner who performs better than one. It is a meta-estimator that uses random subsets of the original dataset to fit base classifiers, and then it sums up individual predictions to provide a final prediction [20].

#### 4. Result

Using four separate algorithms—Random Forest, KNN, Ada Booster, and Bagging—on the processed data of the UCI Machine Learning Repository dataset produced test accuracy values of 97.03, 92.19, 91.11, and 94.03 percent, respectively. We were able to achieve the best result, which was 97.03%, by using the Random Forest model architecture. Table 3 Obtain accuracy below shows the result.

**Table 3.** Obtain accuracy.

|                 | Training Models |             |        |         |
|-----------------|-----------------|-------------|--------|---------|
|                 | Random Forest   | Ada Booster | KNN    | Bagging |
| Obtain Accuracy | 97.03%          | 91.11%      | 92.19% | 94.03%  |

##### A. 10-Fold Cross-Validation Analysis

We assessed the performance of our models using fresh data that had not been used during training using the 10-fold cross-validation method. The average cross-validation score for the employed methodologies is displayed in Table 4. We used the Random Forest classification model and obtained a 10-fold cross-validation score of 95.03%.

**Table 4.** 10-fold cross-validation values.

|   | Training Models |             |        |         |
|---|-----------------|-------------|--------|---------|
|   | Random Forest   | Ada Booster | KNN    | Bagging |
| Accuracy attained through the use of 10-fold cross-validation | 95.03%          | 90.88%      | 91.39% | 92.94%  |

### B. Confusion Matrix Analysis

The confusion matrix gives us a thorough understanding of both the successes and failures of our classification model. The precision of a classification model is assessed using the  $N \times N$  matrix, where  $N$  is the total number of target groups. There are 16 overall attributes in the proposed system. The matrix contrasts the true goal evaluation with what the machine learning model had predicted. By comparing the model's classification of the various fault categories to their actual classification, this matrix shows how they differ [28]. The resulting four pieces of information illustrate this:

- True Positive (TP): Non-diabetic patients identified as non-diabetic.
- False Positive (FP): Misclassification of healthy patients as unhealthy.
- True Negative (TN): accurately identifying healthy patients as healthy.
- False Negative (FN): Incorrectly classifying diabetic patients as non-diabetic.

The Comparison Parameter Formula has shown other variations, as discussed in [28]. Table 4 shows the various comparison parameters.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}$$

## 5. Conclusions

It is becoming more likely for people of all ages to develop diabetes. According to studies, predicting diabetes in its early stages can be a crucial first step in treating the condition. The early prediction of this disease is improving thanks to the machine learning methods we have used. Earlier predictions might make it easier to cover the expense of the treatment, which is great. Finding the most effective method for prediction on this UCI dataset was the major goal of this study. We achieved the best results possible with Random Forest and Bagging. In the future, we will offer a free service through our online application to end users for the early-stage prediction of this disease. We have advised that a web-application system be developed for end-users.

**Author Contributions:** Conceptualization, S.Y. and V.K.S.; methodology, S.Y.; software, S.Y.; validation, S.Y. and V.K.S.; formal analysis, R.C.; investigation, S.Y.; resources, S.Y.; data curation, S.Y.; writing—original draft preparation, S.Y.; writing—review and editing, S.Y. and V.K.S.; visualizations, S.Y.; supervision, R.C. and V.K.S.; project administration, S.Y.; funding acquisitions. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The simulation was performed in the RITEE Workstation, and external journals and links were only referred to for result validation.

**Acknowledgments:** The authors express their gratitude to the Raipur Institute of Technology, Raipur, Chhattisgarh, India, for providing the opportunity to pursue their master's studies and Teachers from department of computer science and engineering for their guidance and help.

**Conflicts of Interest:** The authors declare no financial interests or personal or business associations that could have influenced the work reported in this article.

## References

1. Fitriyani, N.L.; Syafrudin, M.; Alfian, G.; Rhee, J. Development of Disease Prediction Model Based on Ensemble Learning Approach for Diabetes and Hypertension. *IEEE Access* **2019**, *7*, 144777–144789. [\[CrossRef\]](#)
2. Banchhor, M.; Singh, P. Comparative study of ensemble learning algorithms on early stage diabetes risk prediction. In Proceedings of the 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 21–23 May 2021. [\[CrossRef\]](#)
3. Warsi, G.G.; Saini, S.; Khatri, K. Ensemble Learning on Diabetes Data Set and Early Diabetes Prediction. In Proceedings of the 2019 International Conference on Computing, Power and Communication Technologies (GUCON), New Delhi, India, 27–28 September 2019; pp. 182–187.
4. Ahmed, T.M. Developing a predicted model for diabetes type 2 treatment plans by using data mining. *J. Theor. Appl. Inf. Technol.* **2016**, *90*, 181–187.
5. Lohani, B.P.; Thirunavukkarasan, M. A Review: Application of Machine Learning Algorithm in Medical Diagnosis. In Proceedings of the 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 10–12 November 2021; pp. 378–381. [\[CrossRef\]](#)
6. Malini, M.; Gopalakrishnan, B.; Dhivya, K.; Naveena, S. Diabetic Patient Prediction using Machine Learning Algorithm. In Proceedings of the 2021 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 9–10 October 2021. [\[CrossRef\]](#)
7. Huda, S.M.A.; Ila, I.J.; Sarder, S.; Shamsujjoha, M.; Ali, M.N.Y. An Improved Approach for Detection of Diabetic Retinopathy Using Feature Importance and Machine Learning Algorithms. In Proceedings of the 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, 28–30 June 2019. [\[CrossRef\]](#)
8. Prem, S.S.; Umesh, A.C. Classification of Exudates for Diabetic Retinopathy Prediction using Machine Learning. In Proceedings of the 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 30–31 October 2020; pp. 357–362. [\[CrossRef\]](#)
9. Bhat, S.S.; Ansari, G.A. Predictions of diabetes and diet recommendation system for diabetic patients using machine learning techniques. In Proceedings of the 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 21–23 May 2021; pp. 1–5. [\[CrossRef\]](#)
10. Ahmed, U.; Issa, G.F.; Khan, M.A.; Aftab, S.; Khan, M.F.; Said, R.A.; Ghazal, T.M.; Ahmad, M. Prediction of Diabetes Empowered with Fused Machine Learning. *IEEE Access* **2022**, *10*, 8529–8538. [\[CrossRef\]](#)
11. Samet, S.; Laouar, M.R.; Bendib, I. Diabetes mellitus early stage risk prediction using machine learning algorithms. In Proceedings of the 2021 International Conference on Networking and Advanced Systems (ICNAS), Annaba, Algeria, 27–28 October 2021. [\[CrossRef\]](#)
12. Emon, M.U.; Zannat, R.; Khatun, T.; Rahman, M.; Keya, M.S.; Ohidujjaman. Performance Analysis of Diabetic Retinopathy Prediction using Machine Learning Models. In Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 20–22 January 2021; pp. 1048–1052. [\[CrossRef\]](#)
13. Jyotheeswar, S.; Kanimozhi, K.V. Prediction of Diabetic Retinopathy using Novel Decision Tree Method in Comparison with Support Vector Machine Model to Improve Accuracy. In Proceedings of the 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 7–9 April 2022; pp. 44–47. [\[CrossRef\]](#)
14. Paliwal, M.; Saraswat, P. Research on Diabetes Prediction Method Based on Machine Learning. In Proceedings of the 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 10–12 October 2022; pp. 415–419. [\[CrossRef\]](#)
15. Islam, M.M.F.; Ferdousi, R.; Rahman, S.; Bushra, H.Y. Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques. *Adv. Intell. Syst. Comput.* **2020**, *992*, 113–125. [\[CrossRef\]](#)
16. Cherrington, M.; Thabtah, F.; Lu, J.; Xu, Q. Feature selection: Filter methods performance challenges. In Proceedings of the 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 3–4 April 2019; pp. 1–4. [\[CrossRef\]](#)
17. Jin, Z.; Shang, J.; Zhu, Q.; Ling, C.; Xie, W.; Qiang, B. RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. In Proceedings of the 21st International Conference on Web Information Systems Engineering, WISE 2020, Amsterdam, The Netherlands, 20–24 October 2020. [\[CrossRef\]](#)
18. Zhang, L.; Cui, H.; Welsch, R.E. A Study on Multidimensional Medical Data Processing Based on Random Forest. In Proceedings of the 2020 5th International Conference on Universal Village (UV), Boston, MA, USA, 24–27 October 2020; pp. 1–5. [\[CrossRef\]](#)
19. Sapra, V.; Sapra, L.; Vishnoi, A.; Srivastava, P. Identification of Brain Stroke using Boosted Random Forest. In Proceedings of the 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM), Dehradun, India, 10–11 November 2022; pp. 1–5. [\[CrossRef\]](#)
20. Swarupa, A.N.V.K.; Sree, V.H.; Nookambika, S.; Kishore, Y.K.S.; Teja, U.R. Disease Prediction: Smart Disease Prediction System using Random Forest Algorithm. In Proceedings of the 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT), Visakhapatnam, India, 13–14 November 2021; pp. 48–51. [\[CrossRef\]](#)
21. Abu-Aisheh, Z.; Raveaux, R.; Ramel, J.Y. Efficient k-nearest neighbors search in graph space. *Pattern Recognit. Lett.* **2020**, *134*, 77–86. [\[CrossRef\]](#)

22. Zhu, Y.; Wang, J.; Li, X. Research on GA-KNN Image Classification Algorithm. In Proceedings of the 2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), Hamburg, Germany, 7–9 October 2022; pp. 278–282. [[CrossRef](#)]
23. Chethana, C. Prediction of heart disease using different KNN classifier. In Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; pp. 1186–1194. [[CrossRef](#)]
24. Harvey, P.K.; Brewer, T.S. On the neutron absorption properties of basic and ultrabasic rocks: The significance of minor and trace elements. *Geol. Soc. Spec. Publ.* **2005**, *240*, 207–217. [[CrossRef](#)]
25. Zhang, Y.; Ni, M.; Zhang, C.; Liang, S.; Fang, S.; Li, R.; Tan, Z. Research and application of adaboost algorithm based on SVM. In Proceedings of the 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 24–26 May 2019; pp. 662–666. [[CrossRef](#)]
26. Wang, R. AdaBoost for Feature Selection, Classification and Its Relation with SVM, A Review. *Phys. Procedia* **2012**, *25*, 800–807. [[CrossRef](#)]
27. Ariza-López, F.J.; Rodríguez-Avi, J.; Alba-Fernández, M.V. Complete control of an observed confusion matrix. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium 2018, Valencia, Spain, 22–27 July 2018; Volume 2018, pp. 1222–1225. [[CrossRef](#)]
28. Li, X.; Rai, L. Apple Leaf Disease Identification and Classification using ResNet Models. In Proceedings of the 2020 IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT), Shenzhen, China, 13–15 November 2020; pp. 738–742. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.