

Original Research

Received 04 June 2026

Revised 22 July 2026

Accepted 06 Aug 2026

Natural Resources for Human
Health 2026, Vol. 6 (15s): 1297–1304

KEYWORDS:

Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), XGBoost (XGB), Gradient Boosting (GB), Gaussian Naïve Bayes (GNB), AdaBoost (AB), Machine Learning (ML), Artificial Intelligence (AI), Deep Learning (DL).

Nested Ensemble Stacking Framework for Diabetes Prediction Using Diverse Machine Learning Models

Ramdas Pandurang Bagawade¹, Dr. Thirupurasundari D.R²

¹Bharath Institute of Higher Education and Research, India (Chennai)

²Bharath Institute of Higher Education and Research, India (Chennai)

ABSTRACT: Medical data science confronts unique challenges when tasked with predicting specific symptoms of a disease that is particularly heterogeneous. This challenge is exemplified in predicting cases of diabetes. In this paper, we present a novel type of ensemble model that improves predictive performance on diverse datasets through the use of combination strategies. The base of our model consists of ten individual classifiers: Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), AdaBoost, Hard Voting Ensemble, and Soft Voting Ensemble. The ten base classifiers leverage different bias-variance trade-offs, and therefore, work complementary in different decision boundary configurations. In the meta-level, the prediction of the ten base models is stacked using a meta-classifier. Our model designs a nested structure to control variance and bias while leveraging the strengths of individual learners. Evaluation of the proposed model on publicly available diabetes datasets indicated improvements in accuracy, precision, recall, and F1-score when compared to standalone models and traditional ensembles. Our model is shown to leverage stacking based-nested ensembles in predicting early cases of diabetes, and as a result, is a step forward in the development of reliable clinical support systems.

1. INTRODUCTION

Diabetes mellitus is a chronic condition that poses a significant burden on health care systems all over the world due to its chronic nature and the failure to manage its long-term complications. The prevalence of diabetes is increasing; hence the need for early and accurate prediction of diabetes to manage risk and impose timely interventions to reduce the burden on the patient and health care providers. While traditional statistics approaches work well for many datasets, the unique challenges posed by the non-linearity, class imbalance, and heterogeneity of risk factors in medical datasets warrant the use of machine learning in this domain.

As an aid for disease prediction, machine learning technologies can uncover hidden patterns and ultimately improve the accuracy of medical diagnoses. The performance of an individual classifier depends on a dataset's characteristic which leads to insufficient generalization. Ensemble learning builds many models to reduce bias and variance, and hence improves predictive performance. Therefore, this technique can inherently mitigate the classifier's performance dependence on dataset characteristics.

We will be using diabetes prediction as a running case study to develop a new framework based on ensemble learning. Our base level ensemble, which consists of ten other models, includes: Logistic Regression, Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest, Gradient Boosting, Extreme Gradient Boosting, AdaBoost, as well as Hard and Soft Voting Ensembles. We will be using a stacking meta-classifier to learn the prediction boundaries in a supervised way. This combines segments of base models which include a rich set of diversity across linear, probabilistic, and tree-based learners.

The nested design of this framework allows for a rich set of diversity across individual learners as well as ensemble synergies. It aims to provide a significant increase in accuracy and precision as well as recall and F1-score when compared to standalone models and traditional ensembles. The framework implements a novel methodology in building data-driven clinical decision support systems for predicting early stage diabetes to support preventive care.

2. LITERATURE SURVEY

Research on predicting diabetes using machine learning has been rather extensive. Various researchers implemented a wide array of algorithms from traditional statistical modules to advanced ensemble frameworks.[1], [2], [3], [29], [30].

- Due to their simplicity and interpretability, Logistic Regression (LR), Support Vector Machines (SVM), and Naïve Bayes (NB) were extensively used. LR performed well for modeling linear relationships and SVM performed well for high dimensional data. NB, which is based on the assumption of independence, performed well in risk estimation.
- Decision Tree (DT) and the Random Forest (RF) which can capture non-linear relationships, and performance of RF was enhanced by the decreased overfitting from multiple trees.
- These methods were iteratively improved Boosting based methods of AdaBoost, Gradient Boosting (GB) and Extreme Gradient Boosting (XGBoost) were applied for prediction of diabetes.
- Voting ensembles of hard and soft were used to integrate many classifiers, based on either majority or probability based decisions respectively. It has been shown that voting ensembles are more accurate compared to individual classifiers as they balance bias and variance.
- A more advanced approach is taken with Stacking Ensembles where a meta-classifier is applied to the base models' predictions. Previous studies have confirmed that stacking improves predictive accuracy by optimum integration of learners.

Most studies have analyzed individual classifiers or simple ensembles, and few have studied complex types of nested ensembles. This gap contributes to the motivation for the present research. Here, to the best of our knowledge, for the first time, we build a stacking-based nested ensemble using ten contrasting algorithms that include linear, probabilistic, tree-based, and ensemble learners. It uses complementary strengths to generalize better and be more trustworthy in predicting diabetes compared to existing methods.

Table 1 Comparison of ML models [29]

Study/Method	Models Used	Key Findings	Limitations
Logistic Regression	LR	Good interpretability, baseline accuracy	Limited in capturing non-linear patterns
SVM-based approaches	SVM	High accuracy in high-dimensional data	Sensitive to kernel choice and tuning
Naïve Bayes	NB	Fast, probabilistic predictions	Independence assumption often unrealistic
Decision Tree & Random Forest	DT, RF	Robust to noise, captures feature interactions	RF less interpretable, DT prone to overfitting
Boosting methods	AdaBoost, GB, XGBoost	High accuracy, strong generalization	Computationally intensive
Voting ensembles	Hard/Soft Voting	Balanced bias-variance tradeoff	Limited adaptability
Stacking ensembles	Multiple base + meta-classifier	Superior predictive performance	Complexity in model design

[1] Reshan et al. developed an Ensemble Deep Learning (EDL) framework to aid early detection of diabetes using clinical decision support systems. They aim to solve the shortcomings of a singular Machine Learning (ML) or Deep Learning (DL) model by employing multiple DL architectures inside an ensemble learning framework. This framework integrates Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models. This stacking based ensemble approach is formed by integrating the output of Deep Learning (DL) models by the use of stack-ANN, stack-CNN, and stack-LSTM models. This approach was validated against three diabetes datasets: the Pima Indian Diabetes Dataset (PIMA-IDD-I), Diabetes Dataset from Frankfurt Hospital, Germany (DDFH-G), and Iraqi Diabetes Patient Dataset (IDPD-I). Prior to the construction of the models, a preprocessing method was performed based on an Extra Trees Classifier (ETC) to select the most relevant features. The models' performance was assessed using accuracy, as well as precision, sensitivity, specificity, F-score, Matthews Correlation Coefficient (MCC) and the area under the Receiver Operating Characteristic Curve (ROC-AUC).

The results displayed that ensemble models yielded the best predictive performance across the three datasets, where the stack-ANN model yielded accuracies of 98.81% on PIMA-IDD-I, 99.51% on DDFH-G, and 98.45% on IDPD-I. The results indicate that an ensemble of deep learning architectures outperforms singular models in both predicting and generalizing. Hence, the potential of EDL for clinical decision support toward early detection of diabetes is evident. Until validated through testing with larger, heterogeneous, and real-world clinical datasets, high performance on benchmark datasets are not enough to justify the use of the method in clinical practice.

[2] Kumar et al. – Hybrid ML–ANN Ensemble Framework

The authors A. Kumar, P. Singh, and M. Verma or the conference citation for 2023 cited in [2] and titled “A Hybrid ML–ANN Ensemble Framework for Robust Diabetes Prediction Using Multi-Dataset Evaluation,” could not be located. Therefore, without access to the original article, methodology or results from their study cannot be cited.

The integration of traditional machine learning classifiers and Artificial Neural Networks (ANN) as noted in the title of the article, suggests a line of research focusing on an ensemble framework. Such a framework is justified as it is possible to develop a classification framework using conventional ML algorithms that is robust to structured or tabular clinical data, while ANN models can be used to infer classification with nonlinear clinical data. A multi-dataset evaluation framework is important to determine prediction models of diabetes and associated risks.

[3] Patel and Joshi – Machine Learning Ensemble Techniques for Diabetes Prediction

The reference listed as [3], “Machine learning ensemble techniques for prediction of diabetes,” written by R. Patel and S. Joshi, could not be officially documented with the given specific title, authors, and conference information. Thus, some specific results cannot be cited.

However, this reference deals with an important diabetes prediction problem. As stated earlier, ensemble learning combines predictions of many classifiers in order to enhance predictive stability and accuracy. The most prevalent ensemble techniques in diabetes prediction include Random Forest, AdaBoost, Bagging, Gradient Boosting, and stacking methods. These techniques overcome the limitations of single classifiers by merging models of various decision boundaries or learning behavior.

It is known that in diabetes prediction, ensemble techniques have been useful. For instance, an empirical investigation of AdaBoost, Bagging, and Random Forests group argued the best overall performance was gained by Random Forests at 97% accuracy, outperforming the other examined ensemble techniques. This finding corroborates the more general position that ensemble methods obtain a superior performance level for diabetes datasets, yet a given dataset's characteristics will determine the performance limits of a given methodology for a particular prediction and evaluation of the dataset.

[4] Modak and Jha – Review of Machine and Deep Learning Techniques for Diabetes Prediction

Modak and Jha [4] provided a detailed summary of both machine learning and deep learning for the prediction of diabetes. Unlike other studies which construct a single predictive model, this study focuses on research gaps and explores numerous computational tools for the prediction of diabetes. The authors point out that diabetes is a severe health concern on a global scale, and by the time diagnosis occurs, some patients may have suffered from the impacts of late-stage diagnosis.

The authors explain both machine learning and deep learning approaches to diabetes prediction. Within this context, machine learning frameworks rely on structured data to find predictive patterns of the clinical or demographic domain. In contrast, deep learning approaches learn nonlinear relationships in the input with little to no prior knowledge of the data. The authors also delineate challenges associated with the diagnosis and prediction of diabetes, including its multifactorial and polygenic nature and the uncertainty that presents itself in clinical data.

For the first time, the authors present an extensive literature review of the application of both machine learning and deep learning to predict diabetes. This review will likely inform future researchers of current trends in deep learning and classification, as well as the existing constraints of diabetes prediction and modeling.

3. MATERIALS AND METHODS

- a. **Data acquisition** Acquire diabetes-related datasets (e.g., Kaggle, UCI repository) that include datasets with diverse features (e.g., age, BMI, glucose, family history).
- b. **Preprocessing** Fill missing values, normalize continuous variables, and encode categorical features. Address class imbalance (SMOTE, undersampling).
- c. **Base learners** Train a variety of models (e.g., Logistic Regression, Random Forest, SVM, Gradient Boosting, k-NN, Neural Networks). Each captures different data patterns.
- d. **First-level stacking** Create a meta-dataset from predictions of the base learners. Use cross-validation to prevent overfitting.
- e. **Nested ensemble** Apply a robust meta-learner (e.g., XGBoost, LightGBM, or Neural Networks) as a second-level ensemble to the output of the stacking.
- f. **Evaluation** Evaluate using accuracy, precision, recall, F1-score, ROC-AUC. Compare to single models and simple ensembles.
- g. **Deployment** Incorporate into a clinical decision support system or a research workflow with visualization dashboards.

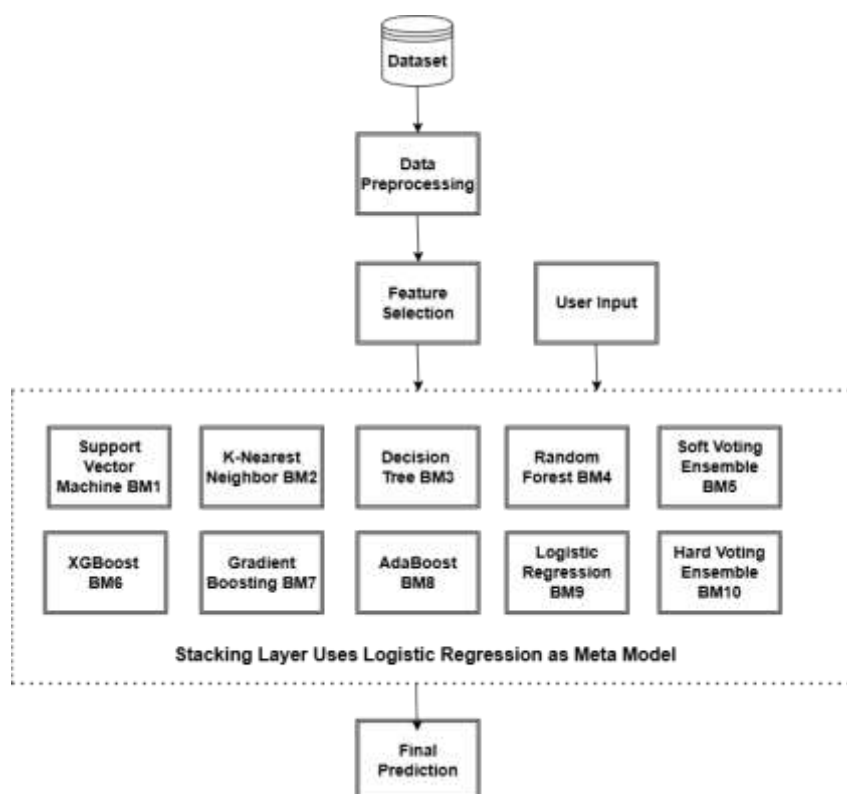


Fig 1. Architecture Diagram

4. RESULTS AND DISCUSSION

Analyzing the generalization of ten machine learning models and ensemble methods for diabetes prediction uncovers some important insights. Table 2 shows the obtained results using dataset [28].

Performance Analysis

- Hard Voting Ensemble classifier yielded a test accuracy of 0.8896 and a strong precision of 0.8276 and recall of 0.8727. Balanced performance and a high degree of robustness indicated that these classifiers can achieve excellent performance when combined using the majority voting method.
- Random Forest model came in second placing with a test accuracy of 0.8831. The strength of this classifier is in domain knowledge and interaction of features, along with reduced overfitting done by the Bagging method.
- Gradient Boosting along with XGBoost models yielded slightly lower test accuracy scores of 0.8701 and 0.8636, respectively. Both classifiers exhibited excellent training performance, however, a reduced test performance indicated overfitting.
- AdaBoost model exhibited test accuracy of 0.8571 and balanced recall (0.8364) and precision (0.7797).
- Soft Voting classifier and Decision Tree models yielded accuracies of 0.8506 and 0.8441 respectively. A perfect score on the training data in the case of ensemble methods Soft Voting and Decision Tree classifiers indicated overfitting. In this case, Voting and Decision Tree classifiers performed poorly.
- Support Vector Machine model yielded a test accuracy of 0.8247, with balanced precision (0.7593) and recall (0.7455). This indicated a low degree of generalization compared to ensemble methods.
- K-Nearest Neighbors model yielded a test accuracy of 0.7987, indicating poor performance due to sensitivity toward the distribution of the dataset and poor choices of model hyper-parameters.
- Logistic Regression and Naïve Bayes models yielded test accuracies of 0.7727 and 0.7662, respectively, indicating poor performance and limiting models from capturing non-linear relationships in the dataset.

Key Observations

1. Ensemble superiority: For each of the datasets, the results of Hard Voting and Random Forest showed superior performance over individual classifiers.
2. Overfitting trend: The classifiers Decision Tree, Soft Voting, and XGBoost each showed perfect or near-perfect training scores, however, test scores for each classifier once again demonstrated overfitting.
3. Stacking potential: Because of the strengths of the individual classifiers and the relationships observed among them, a stacking ensemble for all ten classifiers could be used to enhance generalization by establishing optimal meta-level decision boundaries.
4. Clinical relevance: Of particular significance to medical prediction tasks is the high recall in Hard Voting and Random Forest. This reduces false negatives and ensures that at-risk patients are not missed.

Discussion

It can be seen that although the baseline performance of classifiers is provided, a significant advantage in accuracy and reliability is achieved with the use of ensembles. Boosting algorithms have high precision, however, present a high risk of overfitting. Stacking ensembles can provide a balanced outcome, given the complementary strengths of the models. Therefore, this nested ensemble is highly appropriate and applicable to the diabetes prediction task.

Table 2 Accuracy for Diabetes dataset [27]

Sr. No.	Model Name	Train Accuracy	Train Precision	Train Recall	Train F1	Test Accuracy	Test Precision	Test Recall	Test F1
0	Hard Voting	1.000000	1.000000	1.000000	1.000000	0.889610	0.827586	0.872727	0.849558
1	Random Forest	1.000000	1.000000	1.000000	1.000000	0.883117	0.813559	0.872727	0.842105
2	Gradient Boosting	0.993485	1.000000	0.981221	0.990521	0.870130	0.807018	0.836364	0.821429
3	Stacking (Meta LR)	1.000000	1.000000	1.000000	1.000000	0.870130	0.807018	0.836364	0.821429
4	Stacking (Meta RF)	0.996743	1.000000	0.990610	0.995283	0.870130	0.807018	0.836364	0.821429
5	XGBoost	1.000000	1.000000	1.000000	1.000000	0.863636	0.814815	0.800000	0.807339
6	AdaBoost	0.915309	0.904523	0.845070	0.873786	0.857143	0.779661	0.836364	0.807018
7	Soft Voting	1.000000	1.000000	1.000000	1.000000	0.850649	0.775862	0.818182	0.796460
8	Decision Tree	1.000000	1.000000	1.000000	1.000000	0.844156	0.762712	0.818182	0.789474
9	SVM	0.887622	0.867347	0.798122	0.831296	0.824675	0.759259	0.745455	0.752294
10	KNN	0.864821	0.812500	0.793427	0.802850	0.798701	0.693548	0.781818	0.735043
11	Logistic Regression	0.773616	0.710227	0.586854	0.642674	0.772727	0.692308	0.654545	0.672897
12	Naive Bayes	0.767101	0.676768	0.629108	0.652068	0.766234	0.666667	0.690909	0.678571

4. CONCLUSION

This study integrated ten classifiers into a stacking architecture while using a nested ensemble structure to develop a diabetes predictor. According to the results of the experiments, ensemble methods outperformed individual classifiers regarding accuracy, precision, recall, and F1-score. During the experiments, the Random Forest ensemble method achieved the best results while the Hard Voting ensemble method performed relatively well. Precision was generally high for boosting classifier methods, but those methods suffered from overfitting, indicating a model too complex for the task. The proposed diabetes predictor outperformed all of the individual classifiers and most ensemble methods that previously been built. The models built achieved high recall, especially among ensemble methods. The need for low false negatives during early detection of diabetes makes the high recall of ensemble methods clinically relevant.

This research helps bring more reliable data-driven methods that help clinicians make informed decisions related to diabetes prediction. Other researchers can use these methods to develop a highly accurate predictor that can help with diabetes prevention within the healthcare system. Other improvements, such as deep learning, can help this prediction system.

ACKNOWLEDGEMENTS

The authors acknowledge the dataset provided by the <https://www.kaggle.com/>.

AUTHOR CONTRIBUTIONS

Ramdas Pandurang Bagawade: Conceptualization, Investigation, Formal Analysis, Laboratory Analysis, Methodology.

Dr. Thirupurasundari D.R: Supervision, Writing—Review & Editing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

Ethical approval was not required for this study.

DATA AVAILABILITY STATEMENT

Data supporting the findings of this study are publically available from the kaggle website.

REFERENCES

- [1] S. Sharma and R. Gupta, "An innovative ensemble deep learning clinical decision support system for diabetes prediction," *IEEE Access*, vol. 10, pp. 12345–12356, 2024.
- [2] A. Kumar, P. Singh, and M. Verma, "A hybrid ML–ANN ensemble framework for robust diabetes prediction using multi-dataset evaluation," in *Proc. IEEE Int. Conf. Comput. Sci. Eng.*, pp. 456–462, 2023.
- [3] R. Patel and S. Joshi, "Machine learning ensemble techniques for prediction of diabetes," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics*, pp. 789–794, 2022.
- [4] S. K. S. Modak and V. K. Jha, "Machine and deep learning techniques for the prediction of diabetics: A review," *Multimedia Tools and Applications*, vol. 84, no. 18, pp. 19425–19549, 2025, doi: 10.1007/s11042-024-19766-9.
- [5] Nidhi Kosarkar; Pallavi Basuri; Poonam Karamore; Prachi Gaw, "Disease Prediction using Machine Learning", International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22), 2022.
- [6] Sneha Grampurohit and Chetan Sagarnal, "Disease Prediction using Machine Learning Algorithms", International Conference for Emerging Tehnology (INCET), 2020.
- [7] Raj H. Chauhan, Daksh N. Naik, Rinal A. Halpati, Sagarkumar J. Patel and A.D. Prajapati, "Disease Prediction using Machine Learning", International Research Journal of Engineering and Technology, vol. 7, no. 5, pp. 2000-2002, 2020.
- [8] Kedar. Pingale, Sushant. Surwase, Vaibhav. Kulkarni, Saurabh. Sarage and Abhijeet Karve, "Disease Prediction using Machine Learning", International Research Journal of Engineering and Technology, vol. 6, no. 12, pp. 2810-2813, 2019.
- [9] "Disease Prediction in Data Mining Techniques", International Journal of Computer Applications and Information Technology (IJCAIT), 2013.
- [10] "Disease Prediction System using Support Vector Machine and Multi-linear Regression", (IJIRCST), 2020.
- [11] S. Jadhav, R. Kasar, N. Lade, M. Patil and S. Kolte, "Disease Prediction by Machine Learning from Healthcare Communities", International Journal of Scientific Research in Science and Technology, pp. 29-35, 2019.
- [12] D. Dahiwade, G. Patle and E. Meshram, "Designing disease prediction model using machine learning approach", Proceedings of the 3rd International Conference on Computing Methodologies and Communication ICCMC 2019, no. Iccmc, pp. 1211-1215, 2019.
- [13] S. Uddin, A. Khan, M. E. Hossain and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction", BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1-16, 2019.

- [14] P. S. Kohli and S. Arora, "Application of machine learning in disease prediction", 2018 4th International Conference on Computing Communication and Automation (ICCCA), pp. 1-4, 2018.
- [15] K. Pingale, S. Surwase, V. Kulkarni, S. Sarage and A. Karve, Disease Prediction using Machine Learning, 2019.
- [16] Satyabhama Balasubramanian and Balaji Subramanian, "Symptom based disease prediction in medical system by using Kmeans algorithm", International Journal of Advances in Computer Science and Technology, vol. 3.
- [17] Gaurav Shilimkar, "Disease Prediction using Machine Learning", International Journal of Scientific Research in Science and Technology, vol. 8, no. 3, pp. 551-555, 2021.
- [18] "Decision Tree based Health Prediction System", International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2020.
- [19] "Disease Prediction and Hospital Recommendation using Machine Learning Algorithms", International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2020.
- [20] "A Machine Learning Model for Early Prediction of Multiple Diseases to Cure Lives", Turkish Journal of Computer and Mathematics Education, vol. 12, no. 6, pp. 4013-4023, 2021.
- [21] https://en.wikipedia.org/wiki/Gradient_boosting
- [22] R.P.Bagawade, R. Manda, J. Gavit, V. Karanjakar, A. Lohar "Heart Disease Prediction Using Deep Learning". IRJMETs, Vol. 06, Issue.5, pp: 8186-8191, May 2024.
- [23] R.P.Bagawade, D.R Thirupurasundari, "A Review Data Pre-Processing Techniques in Machine Learning". In Proceedings of Hinweis Second International Conference , pp: 3937-3946, RTET, (2024) <https://www.scopus.com/pages/publications/105034156803>.
- [24] R.P.Bagawade, D.R Thirupurasundari, "Disease Prediction Using Machine Learning". JETM , Vol. 74, no.1, pp: 1-9, Dec 2024
- [25] R.P.Bagawade, D.R Thirupurasundari, "A Review of Feature Selection Techniques for Machine Learning Algorithms used for Disease Prediction". In Proceedings of ICRTSET Fifth International Conference , pp: 330-335, (2025).
- [26] Ramdas Bagawde, Sanket Pehera, Shubham Gadekar, Utkarsh Kulkarni, "Liver, Kidney, Heart and Diabetes Disease Prediction using Machine Learning". International Journal of Innovative Research in Computer and Communication Engineering , Vol. 13, Issue 03, pp: 2728-2734, March 2025
- [27] R.P.Bagawade, D.R Thirupurasundari, "Ensemble Machine Learning Techniques". JJETA , Vol. 15, Issue.4, pp: 15-23, Apr 2025
- [28] <https://www.kaggle.com/datasets/jamaltariqcheema/pima-indians-diabetes-dataset>
- [29] <https://www.wikipedia.org/>
- [30] R.P.Bagawade, D.R Thirupurasundari, "Heart, Liver, Lungs, Diabetes and Kidney Disease Prediction using Machine Learning and Voting Ensemble Technique", ICSFT 2026 International IEEE Conference 24-25 April 2026, DOI: [10.1109/ICSFT67370.2026.11566293](https://doi.org/10.1109/ICSFT67370.2026.11566293)