

Machine Learning Classifiers for Disease Diagnosis using Wrist Pulse Signal

Aathithya S V¹, Achanta Sai Krishna², Kuralanbu S³, Vimal Dharshan N⁴, R Shanmughasundaram⁵

Department of Electrical and Electronics Engineering

Amrita School of Engineering, Coimbatore

Amrita Vishwa Vidyapeetham, India

¹aathithyavellingiri@gmail.com, ²sai.krishna.achanta21@gmail.com, ³kuralanbuselvaraj@gmail.com,

⁴vimaldharshan808@gmail.com, ⁵r_shanmughasundaram@cb.amrita.edu

Abstract—Ayurveda, an ancient Indian medical tradition, uses a diagnostic method Nadi Pariksha evaluates the pulse to determine imbalances in doshas and constitution, providing information about a patient's state of health. Even though Nadi Pariksha has a centuries-old heritage, combining it with contemporary technologies brings new opportunities to reduce the time taken for the identification of disease and also with better accuracy. The dataset is collected from Git. This dataset serves as a critical component in training a machine learning model designed to discern patterns between pulse characteristics and various health conditions. Machine Learning techniques like Decision Tree, SVM, Logistic Regression, XGBoost and KNN are used. The XG Boost algorithm emerging as the preferred choice due to its superior accuracy (96%) when compared to other models, such as Decision Tree, SVM and Logistic Regression. Alongside with the help expert and hardware support like Arduino UNO, Photoplethysmography (PPG) sensor wrist pulse signals are manually collected by placing three fingers alongside on the radial artery, and the three points placed on the hand are known as doshas, i.e., vaatha, pittha and kapha and the abnormality in doshas is identified for that particular person. The results from the Machine Learning model and the Arduino UNO micro controller is compared and checked whether the results are matched.

Index Terms—Ayurveda, Nadi Pariksha, Dosha, Vaatha, Pittha, Kapha, Machine Learning Algorithms, SVM, Logistic Regression, Decision Tree, XG Boost, KNN

I. INTRODUCTION

Ayurveda, which means “the science of life,” is rooted in the belief that the universe is composed of five fundamental elements: fire, water, air, earth and ether. These elements are associated with the five senses of the human body, and they form the basis of Ayurvedic principles. Additionally, they are linked to the three Doshas: Vaatha, Pittha, and Kapha. The balance of these elements and Doshas is crucial for maintaining good health, according to Ayurveda [1].

In Ayurveda, the human body is viewed as a complex system, and maintaining balance among the Doshas is essential for good health. Ayurvedic techniques like Nadi Pariksha, which involves pulse diagnosis, help in identifying imbalances in the Doshas. Each Dosha - Vaatha, Pittha, and Kapha - has its own characteristics and can lead to specific health issues when imbalanced [1]. Nadi Pariksha, also known as Pulse Diagnosis, is an ancient Ayurvedic technique used for non-invasive disease diagnosis, encompassing physical, mental, and

emotional imbalances [2]. Pulse rates and positions are used to determine the predominant Dosha in an individual, enabling personalized treatment and lifestyle adjustments to restore balance and promote well-being. An ayurveda expert places three fingers on the wrist radial artery for checking the three doshas respectively as shown in Fig.1. [3]. Sensor used for collecting the wrist pulse is placed in the position shown in the Fig.1 with the help of expert. The pulse signals, representing Vaatha, Pittha, and Kapha, are examined to diagnose various diseases. The Ayurvedic doctors examined the patient's wrist with their knowledge to detect any dosha imbalance. The process involve pulse examination, questionnaire responses, and analysis of cough, eyes, and face to generate a prescription for the diagnosed disease. [4]. Understanding the characteristics of each Dosha - Vaatha, Pittha, and Kapha - is essential in Ayurveda. Vaatha governs bodily functions involving movement. Pittha governs bodily functions concerned with heat, metabolism and digestion, and Kapha governs the structural aspects of the body and its fluids. Each Dosha has its own set of attributes and potential health issues when imbalanced. By recognizing one's predominant Dosha and making appropriate lifestyle choices, individuals can maintain balance, prevent illness, and promote overall health and well-being. [5]

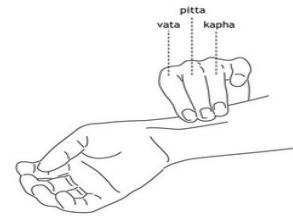


Fig. 1. Dosha positions in Radial Artery

Machine learning algorithms play a crucial role in analyzing pulse characteristics and identifying imbalances in the Tridosha, offering early notification of potential health issues. Various signal processing and machine learning methods are employed to examine vaatha, pittha, and kapha signals, and able to determine and diagnose the imbalance of Dosha. [6]

The research conducted by Vishu.M and Anjali.G explored various methodologies for machine learning algorithms, focusing on precision and accuracy rate. They used machine learning techniques such as K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), Artificial Neural Network (ANN). Classification evaluation metrics such as accuracy, F-score, precision, recall, and root mean square error (RMSE). After examining the results, the authors suggest that CatBoost, which modifies with hyper-parameters and achieves 0.95 F-score, 0.96 accuracy rate, 0.95 recall, and 0.95 precision [8]. In the study by N.Narendra and V.Yashwant, delve into Nadi-Pariksha detecting whether the person taken food or not. They used Fast Fourier Transform spectrum feature are extracted. Support Vector Machine (SVM) used for classification of person taken pre-meal and post-meal of a 40 members. The experimental results show that 88.8 % for pre-meal signal and 81.48% for post-meal signal is obtained. [23] Based on research, machine learning algorithms such as Logistic Regression, Support Vector Machine, Decision Tree, XGBoost, and KNN are implemented to predict disease diagnosis and calculate the accuracy, precision, and recall for performance.

II. DATASET DESCRIPTION

The dataset contains 1071 patients's wrist pulse signals (Vaatha, Pittha, and Kapha) and their associated 32 symptoms of diseases [24]. The dataset was divided into four classes namely Migraine, Arthritis, Diarrhea and Gastritis. Class A contains 259 samples, Class B contains 259 samples, Class C contains 280 samples, and Class D contains 273 samples.

III. METHODOLOGY

The Dataset collected has been classified based on the disease diagnosed, and later classified based on the dosha, for the respective disease. Rectifying missing values, eliminating duplicate records, and mitigating data inconsistencies. The Dataset consists of object feature data type, is transformed into numerical feature data types [12]. The methods used for data transformation are Replace technique and Label encoding technique. It converts object data types to numerical ones so that machine learning models that only accept numerical input can fit them. [13]

A. Data Analysis

After conducting a thorough analysis of the dataset, the mean value was calculated.

B. Machine Learning Models

- Supported vector machines
- Logistic Regression
- Decision Tree Classification
- XGBoost (eXtreme Gradient Boosting)
- K Nearest Neighbors

Supported vector machines (SVM) are supervised machine learning algorithms that aid in solving regression or classification issues. It seeks to identify the best cut-off point between the potential results. Depending on the chosen kernel function,

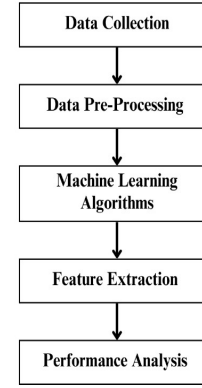


Fig. 2.

SVM performs intricate data transformations. Then, using the results of those transformations, it attempts to maximise the boundaries separating your data points based on the labels. The goal is to identify a hyperplane in a n-dimensional space that maximises the separation of the data points to their prospective classes in order to generalise. [14]

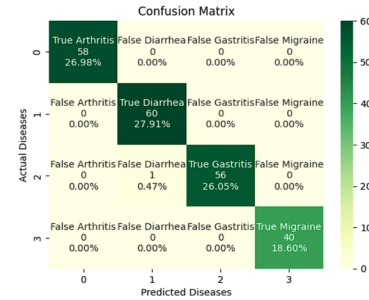


Fig. 3. SVM - Full Features

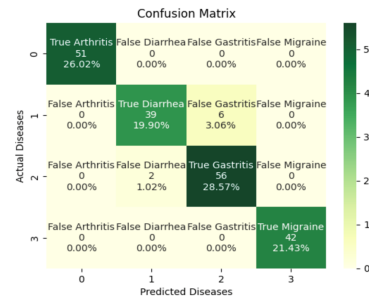


Fig. 4. SVM - Reduced Features

Logistic regression is a statistical technique used to investigate the relationship between two data components. Three forms of logistic regression may be distinguished based on the categories:

- 1) Binary: Binary logistic regression was mentioned earlier in the case of classifying an object as an animal or not

an animal—it's an either/or solution.

- 2) Multinomial: In multinomial logistic regression, the dependent variable, such as "snake", "hen", or "fish," may be one or more of three potential unordered varieties.
- 3) Ordinal: Three or more ordered sorts of dependent variables, such as "small", "average", or "big", are conceivable in ordinal logistic regression. [16]

As we have four classes in our dataset we use Ordinal Logistic Regression. The confusion matrix of logistic regression can be seen below.

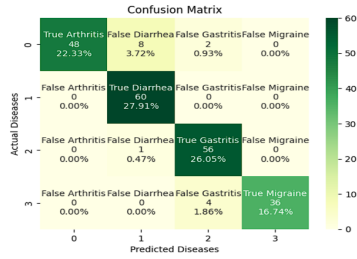


Fig. 5. Logistic Regression - Full Features

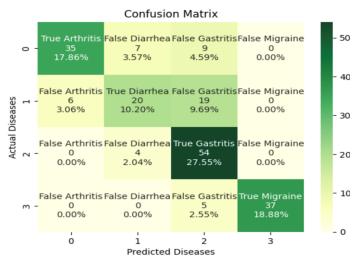


Fig. 6. Logistic Regression - Reduced Features

A tree structure that resembles a flowchart is a decision tree, in which the leaf nodes represent algorithmic results (classes), internal nodes represent features, and branches represent rules. This algorithm for supervised machine learning is adaptable and can be used for both regression and classification problems. The classification rules define the pathways that lead from the roots to the leaves. The process of creating a decision tree is selecting the optimum feature at each node to split the data in order to maximize information gain. [17]

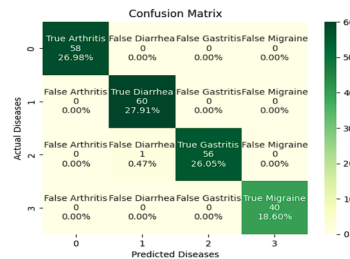


Fig. 7. Decision Tree - Full Features

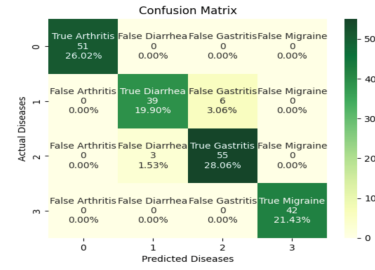


Fig. 8. Decision Tree - Reduced Features

Particularly in the case of decision trees, XGBoost (eXtreme Gradient Boosting) has become a potent and well-liked approach for boosting ensemble models. It is based on the ideas of gradient boosting, which adds weak learners (usually decision trees) to an ensemble one after the other in a sequential fashion, with each new model learning from the mistakes of the ones that came before it. Regression and classification problems are two areas in which XGBoost excels because of its effective implementation and optimisation methods. Its capacity to handle intricate datasets with plenty of features is one of its main advantages. This is made possible by sophisticated regularisation techniques like L1 regularization (controlled by the alpha term) encourages sparsity - so it encourages weights to go to 0 and L2 (controlled by the lambda term) encourages the weights to be small, which reduce overfitting. Moreover, XGBoost uses parallel processing and tree pruning strategies to improve model training speed and accuracy. [18]

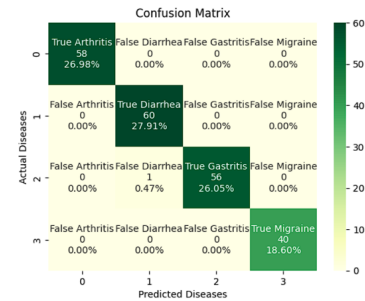


Fig. 9. XGBoost - Full Features

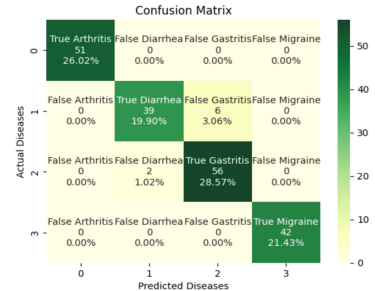


Fig. 10. XGBoost - Reduced Features

The field of sickness prediction has made extensive use of it. Using traits and labels from the training set, KNN is a supervised algorithm that forecasts how unlabeled data will be categorized [18] [20]. The Euclidean distance to metric of similarity based on the distance measure is used, it is evident that several data points with the same class label are in close proximity to one another in numerous local locations. While there are other ways to choose a k-value, one simple way is to run the algorithm iteratively with different k-values and choose the one that works best.

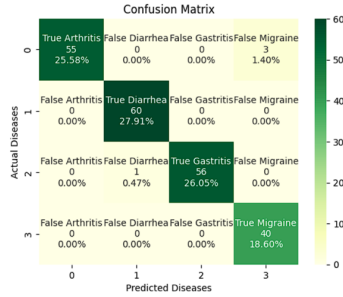


Fig. 11. KNN - Full Features

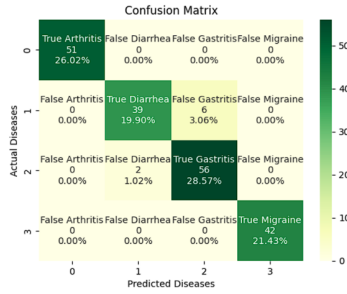


Fig. 12. KNN - Reduced Features

IV. FEATURE EXTRACTION METHODS

A. Principal component analysis

A statistical method called Principal Component Analysis (PCA) is used to minimize a dataset's dimensionality while preserving the possible amount of information. The principal components of the newly created variables are linear combinations of the original variables, and they are created by identifying the most significant features or variables within the dataset.

In unsupervised learning, PCA is a method used to examine the connections among a set of variables. A statistical method called PCA uses an orthogonal transformation to convert a set of correlated variables into a set of uncorrelated variables. Without any prior knowledge of the target variables, the main goal of PCA is to reduce the number of dimensions in a dataset while maintaining the most important patterns or connections between the variables. The way PCA works is by calculating each characteristic's variance. This is due to the fact that characteristics with high variance typically show a distinct

difference between classes, which makes dimensionality reduction easier.

Correlation indicates the degree of association between two variables. The correlation coefficient varies between -1 and +1. In this context, the value of -1 implies that the variables are inversely proportional to each other, whereas a value of +1 shows that the variables are directly proportional to each other. Eigenvectors are non-zero vectors that produce a scalar multiple of themselves when multiplied by a square matrix M.

The following features are selected by Machine Learning Algorithms.

'acidity', 'indigestion', 'blurred and distorted vision', 'excessive hunger', 'muscle weakness', 'stiff neck', 'swelling joints', 'depression', 'abdominal pain', 'vomiting', 'blood in mucus', 'fatigue', 'loss of appetite', 'heartburn'.

V. MAJOR FEATURES FOR PARTICULAR DISEASE

Disease Classification: Migraine
Eigenvalues with Feature Names:
acidity: 1.7051099876856755
indigestion: 1.4147489881651332
excessive_hunger: 1.2862876417199418
muscle_weakness: 1.2176137627168986
stiff_neck: 1.183309691433393
swelling_joints: 1.0339168058317265
movement_stiffness: 0.9413497627044188
depression: 0.8777427372642257
irritability: 0.8204718487711866
painful_walking: 0.7434280782637187
visual_disturbances: 0.7298775223731016
blurred_and_distorted_vision: 0.5816886975143226
headache: 0.5148420724554859

Fig. 13. Migraine - Eigen Vectors and Eigen Values

Disease Classification: Arthritis
Eigenvalues with Feature Names:
acidity: 1.519202496063712
indigestion: 1.4881579938250697
headache: 1.3347632047729583
blurred_and_distorted_vision: 1.1587712315970358
painful_walking: 1.063660778765785
visual_disturbances: 1.0328320453153055
irritability: 0.9788245002816933
depression: 0.8502525634595641
movement_stiffness: 0.8353522982027515
swelling_joints: 0.79668486435446
stiff_neck: 0.7207605549760562
muscle_weakness: 0.654027959511897
excessive_hunger: 0.6170971057732865

Fig. 14. Arthritis - Eigen Values with features

Disease Classification: Diarrhea
Eigenvalues with Feature Names:
acidity: 2.484475104343585
indigestion: 1.5447468357721619
headache: 1.3181020318477552
excessive_hunger: 1.1411218939305978
movement_stiffness: 0.9066564950508967
irritability: 0.852258719572392
depression: 0.822595437796122
swelling_joints: 0.6906912626147953
stiff_neck: 0.5427790252415509
muscle_weakness: 0.4527061853367394
blurred_and_distorted_vision: 0.28329353179090316

Fig. 15. Diarrhea - Eigen Values with features

Disease Classification: Gastritis
Eigenvalues with Feature Names:
acidity: 4.310173763957892
indigestion: 2.0967204572731872
headache: 1.2169609700655846
blurred_and_distorted_vision: 1.0892933016208766
excessive_hunger: 0.8146103626271541
stiff_neck: 0.6998834279367435
swelling_joints: 0.6694554906425111
visual_disturbances: 0.5763103191528527
abdominal_pain: 0.5285124164822119
painful_walking: 0.5090198692281234
irritability: 0.43856141247831937
depression: 0.40449445363252434
movement_stiffness: 0.3973068367730637
muscle_weakness: 0.3001675063642552

Fig. 16. Gastritis - Eigen Values with features

VI. RESULTS AND DISCUSSIONS

Machine Learning Algorithm(s)	Accuracy	Precision	Recall	F1 - Score
Decision Tree	99.5 %	100 %	100 %	100 %
XG Boost	99.7 %	100 %	100 %	100 %
Supported Vectored Machine	96 %	96 %	96 %	96 %
Logistic Regression	94 %	94 %	93 %	93 %
KNN	99.4 %	99.5 %	99.5 %	99 %

TABLE I
MACHINE LEARNING RESULTS WITH ALL FEATURES

The TABLE I presents the performance metrics of various machine learning algorithms using all features. XGBoost shows the highest accuracy at 99.7%, closely followed by Decision Tree at 99.5%, both achieving perfect precision, recall, and F1 scores. These two algorithms are the most effective for this dataset, while logistic regression exhibits the lowest performance among the listed algorithms.

Machine Learning Algorithm(s)	Accuracy	Precision	Recall	F1 - Score
Decision Tree	95 %	96 %	96 %	96 %
XG Boost	96 %	96 %	96 %	96 %
Supported Vectored Machine	96 %	96 %	96 %	96 %
Logistic Regression	94 %	98 %	93 %	93 %
KNN	94 %	95 %	94 %	94 %

TABLE II
MACHINE LEARNING RESULTS WITH REDUCED FEATURES

The TABLE II displays the performance metrics of various machine learning algorithms using reduced features. XGBoost, Decision Tree, and Supported Vector Machine all achieve high and equal performance, with accuracies, precision, recall, and F1 scores all at 96%. These three algorithms are the most effective for this dataset with reduced features. Logistic Regression, despite its high precision of 98%, and KNN show marginally lower performance with an accuracy of 94%.

The two tables TABLE I and TABLE II illustrate the performance of various machine learning algorithms on the same dataset using different feature sets. The first table shows results with all available features, demonstrating higher performance

metrics across most algorithms, with accuracies ranging from 94% to 99.7%. In contrast, the second table uses a reduced set of features, resulting in generally lower performance metrics with accuracies between 94% and 96%. Notably, while the Support Vector Machine's performance remains unchanged, other algorithms, such as Decision Tree and XG Boost, exhibit a noticeable decline in accuracy, precision, recall, and F1 score when features are reduced.

VII. FORMULAE

1) *Accuracy*: The percentage of correctly classified instances is provided by accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

2) *Precision*: Precision refers to the degree of precision in optimistic forecasts.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

3) *Recall (Sensitivity or True Positive Rate)*: The percentage of accurately predicted actual good cases in all instances is known as recall.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4) *F1 Score*: The harmonic mean of recall and precision tends to produce the F1 score.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

VIII. CONCLUSION

The conference paper discusses the application of machine learning classifiers for disease diagnosis using wrist pulse signals in the context of Ayurveda. It highlights the importance of personalized diagnosis, adaptability of machine learning models, and the ability to distinguish between normal and abnormal signal patterns for early detection of health conditions.

Furthermore, the paper emphasizes the significance of combining traditional Ayurvedic diagnostic methods with contemporary technologies to enhance diagnostic accuracy. The study showcases the effectiveness of the XG Boost and Supported Vector Machine algorithm in achieving superior accuracy (96%) compared to other machine learning models like Decision Tree, KNN and Logistic Regression.

In conclusion, the research presented in the conference paper demonstrates the potential of machine learning algorithms in leveraging wrist pulse signals for disease diagnosis, paving the way for more effective and tailored healthcare interventions based on individualized patterns and features.

REFERENCES

- [1] K. Aakant, V. Akant and D. P. Choudhari, "Nadi Signal Analysis: A Review," 2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22), Nagpur, India, 2022, pp. 01-05
- [2] S. Chaudhari and R. Mudhalwadkar, "Nadi pariksha system for health diagnosis," 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, India, 2017, pp. 1-4
- [3] "Nadi Pariksha - Pulse Diagnosis at Sri Sri Ayurveda Hospital," Sri Sri Ayurveda Hospital. <https://srisriayurvedahospital.org/nadi-pariksha/>
- [4] A. Kadarmandalgi and M. Asaithambi, "Pulse diagnosis System for Nadi Pariksha using parametric and statistical analysis," 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India, 2019, pp. 1-7
- [5] V. Madaan and A. Goyal, "Predicting Ayurveda-Based Constituent Balancing in Human Body Using Machine Learning Methods," in IEEE Access, vol. 8, pp. 65060-65070
- [6] "Deep Learning in Data Analytics: How to Extract Complex Patterns from Unstructured Data — Wissen." <https://www.wissen.com/blog/deep-learning-in-data-analytics-how-to-extract-complex-patterns-from-unstructured-data>
- [7] V. Madaan and A. Goyal, "X-Cardio: Fuzzy Inference System to Diagnose Heart Diseases," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2018, pp. 1049-1053
- [8] T. Worth, "Ayurveda: Does it really work?," WebMD, Nov. 23, 2023. <https://www.webmd.com/balance/ayurvedic-treatments>
- [9] "Tridosha in Ayurveda - Vata, Pitta, and Kapha," Sanjeevanam, Jan. 12, 2024. <https://sanjeevanam.com/blog/tridosha/>
- [10] S. Raman, "All about Tri-Dosha: Learn your Tri-Dosha personality!," Mountain Top Clinic, Jul. 26, 2023. <https://mountaintopclinic.com/all-about-tri-dosha-learn-your-tri-dosha-personality/>
- [11] E. Priya, S. Pradeep, K. S. Krishna and S. Saran, "Characterization of Wrist Pulse Using Machine Learning Techniques," 2023 Intelligent Computing and Control for Engineering and Business Systems (IC-CEBS), Chennai, India, 2023, pp. 1-5
- [12] V. Shashkina, "Data Preparation for Machine Learning: A Step-by-Step Guide — ITRex," ITRex, Mar. 20, 2024. <https://itrexgroup.com/blog/data-preparation-for-machine-learning/>
- [13] "Label encoding in Python - JavatPoint," [www.javatpoint.com](https://www.javatpoint.com/label-encoding-in-python). <https://www.javatpoint.com/label-encoding-in-python>
- [14] "Ayurvedic Medicine-based Automated Machine Learning Approach for Determining the Nature of the Human Body", Journal of Pharmaceutical Negative Results, pp. 1161–1165
- [15] A. G. B. Ganesh, A. Ganesh, C. Srinivas, N. Dhanraj, and K. Mensinkal, "Logistic regression technique for prediction of cardiovascular disease," Global Transitions Proceedings, vol. 3, no. 1, pp. 127–130
- [16] "Logistic Regression in Machine Learning - JavatPoint," [www.javatpoint.com](https://www.javatpoint.com/logistic-regression-in-machine-learning). <https://www.javatpoint.com/logistic-regression-in-machine-learning>
- [17] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," Healthcare Analytics, vol. 3, p. 100130
- [18] A. Ogunleye and Q. -G. Wang, "Enhanced XGBoost-Based Automatic Diagnosis System for Chronic Kidney Disease," 2018 IEEE 14th International Conference on Control and Automation (ICCA), Anchorage, AK, USA, 2018, pp. 805-810
- [19] H. Shee, W. Cheruiyot, and S. Kimani, "Application of K-Nearest neighbour classification in medical data mining," ResearchGate, Apr. 2014, [Online]. Available: https://www.researchgate.net/publication/270163293_Application_of_k-Nearest_Neighbour_Classification_in_Medical_Data_Mining
- [20] D. V. Rao, Y. Sucharitha, D. Venkatesh, K. Mahamthy and S. M. Yasin, "Diagnosis of Parkinson's Disease using Principal Component Analysis and Machine Learning algorithms with Vocal Features," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 200-206
- [21] M. Z. I. Chowdhury and T. C. Turin, "Variable selection strategies and its importance in clinical prediction modelling," Family Medicine and Community Health, vol. 8, no. 1, p. e000262
- [22] GeeksforGeeks, "Evaluation metrics in machine learning," GeeksforGeeks, May 16, 2024. <https://www.geeksforgeeks.org/metrics-for-machine-learning-model/>
- [23] N. N. Khaire and Y. V. Joshi, "Diagnosis of Disease Using Wrist Pulse Signal for classification of pre-meal and post-meal samples," 2015 International Conference on Industrial Instrumentation and Control (IIC), Pune, India, 2015, pp. 866-869, doi: 10.1109/IIC.2015.7150864.
- [24] AbdullahShiraz. (n.d.). AYUCARE/Nadun/Disease_Prediction_Model/dataset/data.csv atmain·AbdullahShiraz·AYUCARE.GitHub. https://github.com/AbdullahShiraz/AYUCARE/blob/main/Nadun/Disease_Prediction_M