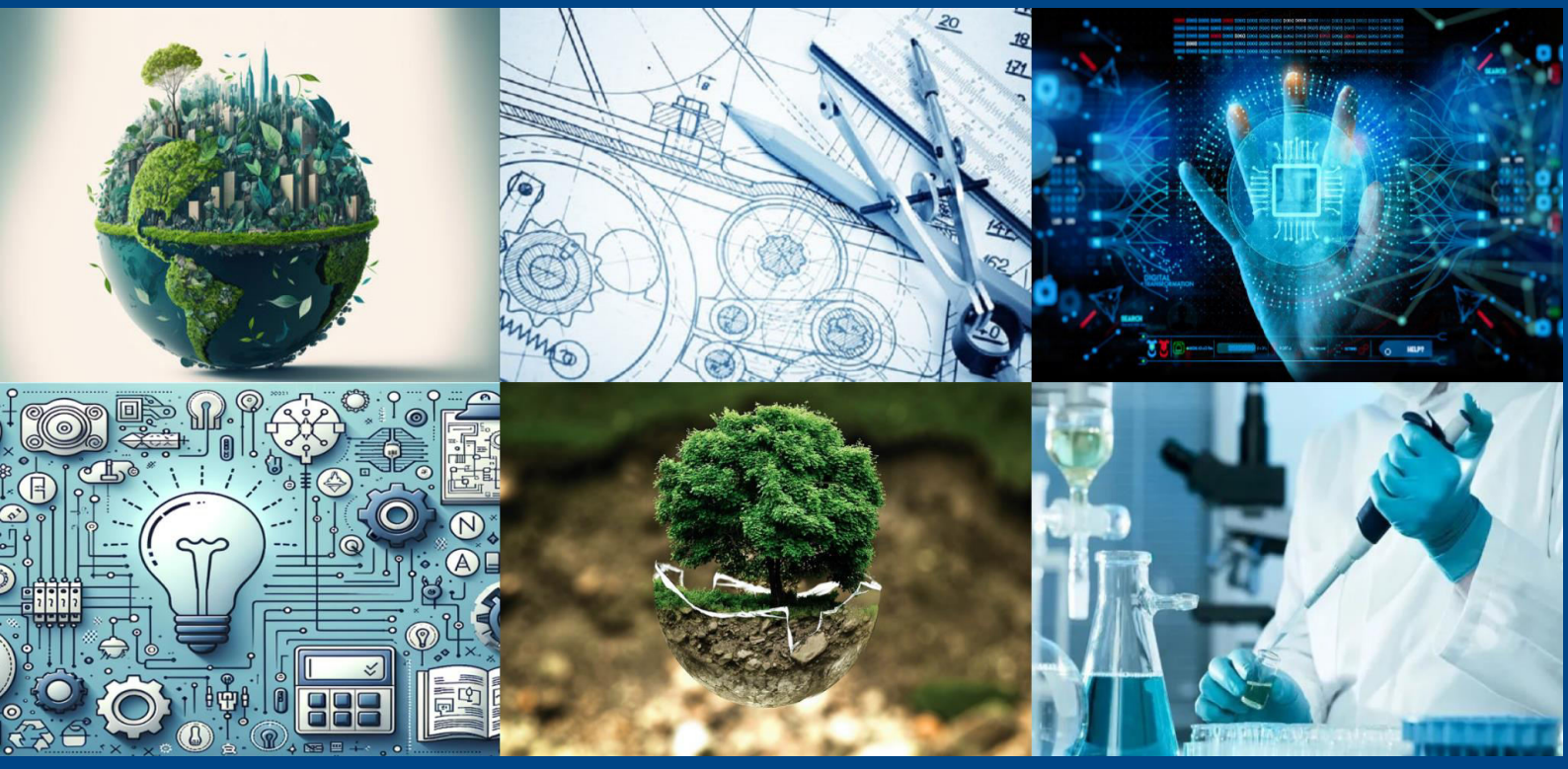
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 3, March 2025

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803020

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Diabetes Prediction

Mrs.Y.Prasanthi[1], P. Hara Om[2], T.Vamsi[3], P.Srikanth[4], M.Manohar[5], O.Naga Bhaskar Sai[6]

Assistant Professor, Department of CSE-Data Science, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India [1]

B. Tech, Department of CSE-Data Science, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India[2-6]

ABSTRACT: Diabetes has turned into a worldwide health emergency and arose to become a leading global health hazard affecting mortality. Early diagnosis and prevention are the keys to cure this long-term disorder. This project provides a machine learning-based diabetes prediction system. The System uses different patient medical records and health parameters like Pregnancies, Glucose levels, BMI, Blood pressure, Insulin, Diabetes Pedigree Function. The system proposed here predicts whether any user can get diabetes (or) not. Here we employ different machine learning algorithms. In addition to prediction, the site also offers personalized diet such as what nutrients they should take in order to prevent getting diabetes and whether they get assists them not to raise the values and exercise guidance to encourage healthy living. The approach is all inclusive in terms of offering a full solution to managing diabetes through enhanced diagnosis and prevention measures.

KEYWORDS: Machine Learning ,KNN, SVM, Random Forest, LightGBM, Health Care, Medical Parameters

I. INTRODUCTION

Diabetes is a disease that disables the body's mechanism to use glucose efficiently, and as a result, high levels of sugar are found in the blood. Diabetes has two major forms: Type 1, in which the pancreas fails to make enough insulin, and Type 2, in which the body stops responding to insulin. With rising rates of unhealthy diets and lifestyles, Type 2 diabetes is spreading fast, even to the younger population. And another is Gestational Diabetes develops during pregnancy and normally disappear after child birth. It occurs when the body fails to make enough insulin to keep up with the increased demand of Pregnancy and resulting in excess sugar. The early diagnosis of diabetes is essential to avoid its complications like cardiovascular disease, kidney failure, and vision loss.

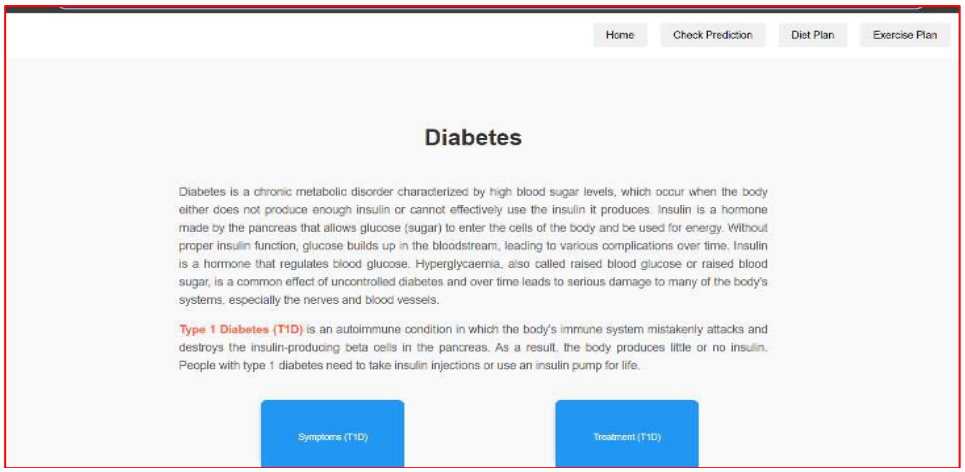


Fig1.Home Page

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803020

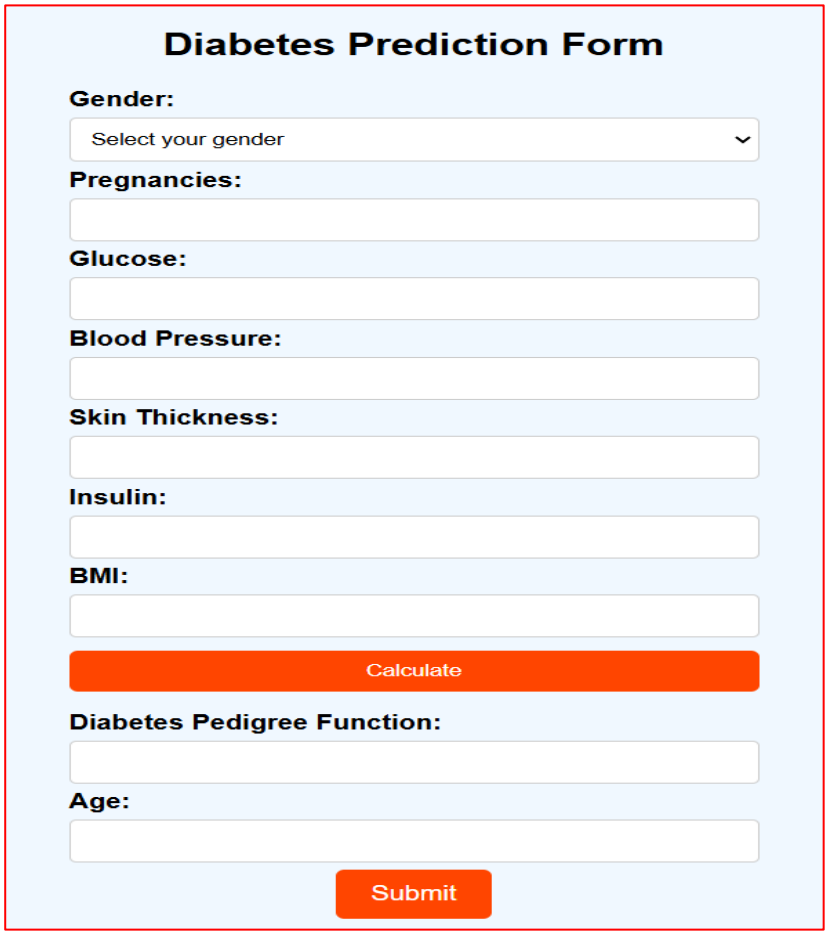
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Machine learning (ML) has, in the last few years, proven to be an effective tool for disease prediction using clinical data. ML models utilize large datasets and predictive algorithms to identify patterns and risk factors that cannot be identified with conventional diagnostic procedures. The current paper addresses applying machine learning algorithms to predict diabetes from clinical features and integrating the predictions into an easy-to-use web interface with personalized diet and exercise plans.

Fig 2.Prediction Form



II. RESEARCHAND METHODOLOGY

The current research brings machine learning together with a web platform to offer users' real-time predictions of their risk of developing diabetes along with actionable tips for managing their health. The following methodology was utilized:

2.1 Data Collection

We used the Pima Indian Diabetes Dataset, a well-known dataset in diabetes literature. The dataset contains 768 patient records and eight important features: pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, diabetes pedigree function, and age. These were selected because of their high correlation with the onset of diabetes.

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803020

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

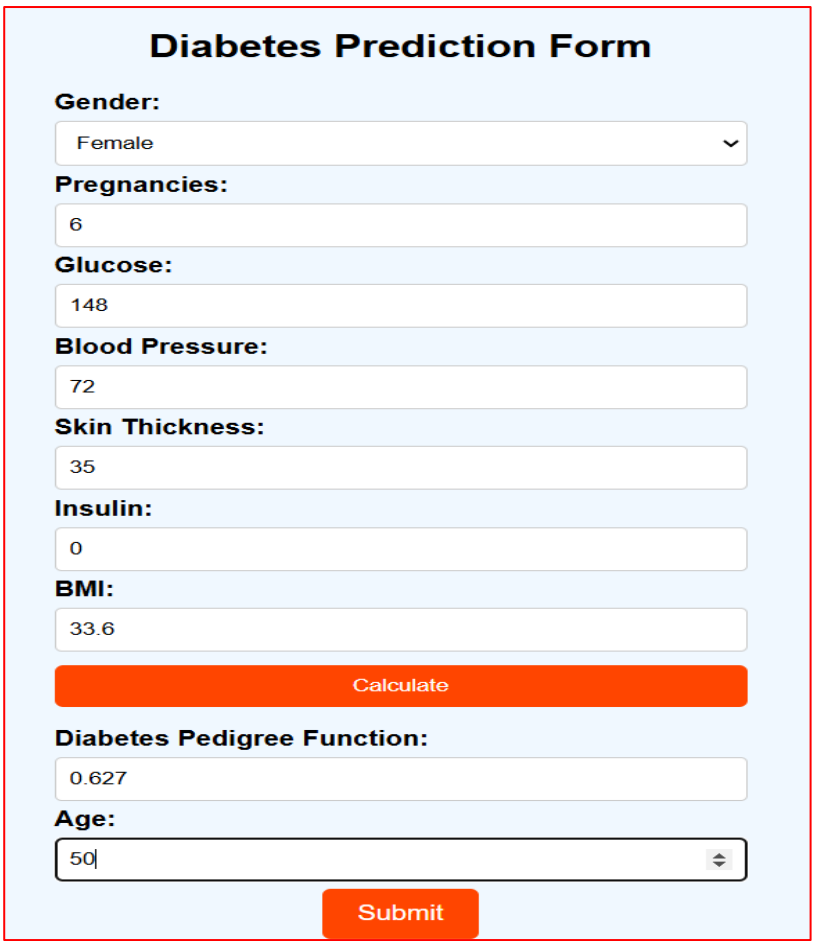


Fig 3. Here we enter the values to predict whether they have diabetes (or) not

2.2 Data Preprocessing Data preprocessing was conducted by dealing with missing values, scaling numeric data through the use of robust scaler, and one-hot encoding of categorical variables. We also binned BMI and insulin measurements into medically accepted intervals for better model prediction. In order to deal with class imbalance in the data, we used the Synthetic Minority Over-sampling Technique (SMOTE).

2.3 Machine Learning Models We implemented and compared four machine learning models:

- Random Forest (RF)
- LightGBM
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

Each model was optimized with grid search to find the best hyperparameters, and cross-validation was done to test the performance of the models. The best models were chosen on the basis of their accuracy, precision, recall, and F1-score.

2.4 Web Application Development In order to make the system available to users, we created a web application based on the Flask framework. The application enables users to enter their health information through an easy-to-use form and offers instant feedback on their risk of diabetes. The application also features diet and exercise suggestion modules, intended to promote healthier lifestyles.

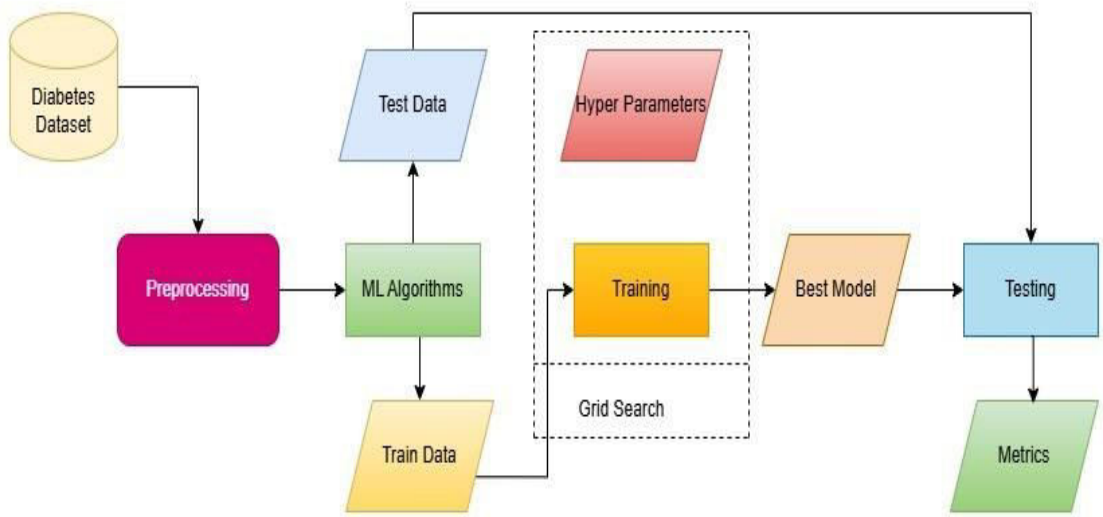
© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803020

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

System Architecture



III. RESULTS AND DISCUSSION

3.1 Model Performance

Each of the models was trained on 80% of the data and evaluated on the remaining 20%. The following table presents the accuracy of each model on the test data:

Model	Accuracy
Random Forest	81.2%
LightGBM	83.4%
SVM	78.5%
KNN	74.6%

The LightGBM model performed better than the other models, with an accuracy of 83.4%. The Random Forest model also performed well, with an accuracy of 81.2%. These models proved to be capable of identifying the non-linear associations between clinical features and diabetes risk.

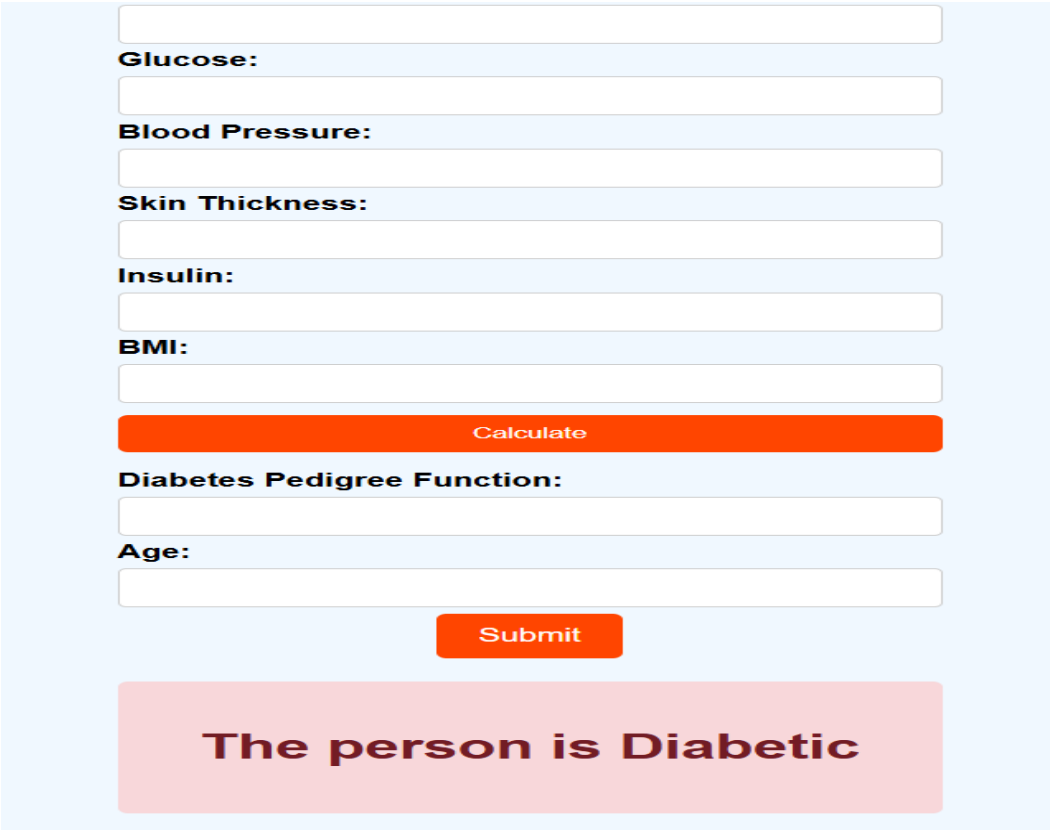
3.2 Explainability In order to improve model transparency, we included SHAP (Shapley Additive explanations) values for explaining predictions. SHAP highlighted glucose, BMI, and age as the most significant features in predicting diabetes risk, in agreement with clinical literature.

3.3 Web-Based System The web-based platform, created with Flask, combines the machine learning models to make real-time diabetes predictions. The users input their health parameters, and the system offers a prediction (whether they have "Diabetes" or not). The platform also offers customized meal plans (vegetarian and non-vegetarian) and exercises as per age groups. The above facility aids in lifestyle modifications towards controlling or preventing diabetes risk.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



IV. THEORY AND CONCLUSIONS

The combination of web-based technology and machine learning provides an end-to-end solution for diabetes risk prediction and management. Using clinical information and advanced ML models, our system accurately predicts diabetes risk and enables users to make lifestyle choices based on informed decisions. The integration of predictive analytics with customized health management plans makes this platform a useful tool for individuals.

Future efforts will be directed at increasing the size of the system to include more varied datasets, enhancing prediction quality, and implementing more sophisticated explainability methods. In addition, investigating the application of deep learning algorithms may unlock even higher performance in prediction.

REFERENCES

1. Ahamed, K. U., Islam, M., Uddin, A., Akhter, A., Paul, B. K., Yousuf, M. A., . . . Moni, M. A., et al. (2021). A deep learning approach using effective preprocessing techniques to detect covid-19 from chest CT-scan and X-ray images. Computers in Biology and Medicine, 139, Article 105014. 10.1016/j.compbiomed.2021.105014.
2. Albahli, S. (2020). Type 2 machine learning: An effective hybrid prediction model for early type 2 diabetes detection. Journal of Medical Imaging and Health Informatics, 10, 1069–1075.
3. Brownlee, J. (2016a). A gentle introduction to the gradient boosting algorithm for machine learning. https://machinelearningmastery.com/gentle-introductiongradient-boosting-algorithm-machine-learning/ Accessed: 2021-03-20.
4. Brownlee, J. (2016b). K-nearest neighbors for machine learning. https:// machinelearningmastery.com/k-nearest-neighbors-for-machine-learning/ Accessed: 2021-03-20.
5. Brownlee, J. (2016c). Logistic regression for machine learning. https://www. geeksforgeeks.org/understanding-logistic-regression/ Accessed: 2021-03-20.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

6. Choubey, D. K., Paul, S., Kumar, S., & Kumar, S. (2017). Classification of Pima Indian diabetes dataset using naive Bayes with genetic algorithm as an attribute selection. In Proceedings of the international conference on communication and computing system (ICCCS 2016) (pp. 451–455).
7. Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Medical Informatics and Decision Making, 19, 211._
8. Ray, S. (2017). Understanding support vector machine(svm). https://www. analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-examplecode/ Accessed: 2021-03-20. Tigga, N. P., & Garg, S. (2020). Prediction of type 2 diabetes using machine learning classification methods. Procedia Computer Science, 167, 706–716.
9. UCI Machine Learning Repository. (1998). Diabetes data set. https://archive.ics.uci.edu/ml/datasets/diabetes Accessed: 2021-03-20. Vinayakumar, R., Alazab, M., Soman, K., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. IEEE Access : Practical Innovations, Open Solutions, 7, 41525–41550.
10. International Journal on "Wielding Neural Networks to Interpret Facial Emotions in Photographs with Fragmentary Occlusion", on American Scientific Publishing Group (ASPG) Fusion: Practice and Applications(FPA) ,Vol. 17, No. 01, August, 2024, pp. 146-158.
11. International Journal on "Prediction of novel malware using hybrid convolution neural network and long short-term memory approach", on International Journal of Electrical and Computer Engineering (IJECE),Vol. 14, No. 04, August, 2024, pp. 4508-4517.
12. International Journal on "Cross-Platform Malware Classification: Fusion of CNN and GRU Models",on International Journal of Safety and Security Engineering (IIETA),Vol. 14, No. 02, April, 2024, pp. 477-486.
13. International Journal on "Enhanced Malware Family Classification via Image-Based Analysis Utilizing a Balance-Augmented VGG16 Model,onInternational information and Engineering Technology Association (IIETA),Vol. 40, No. 5, October, 2023, pp. 2169-2178.
14. International Journal on "Android Malware Classification Using LSTM Model, International information and Engineering Technology Association (IIETA) Vol. 36, No. 5, (October, 2022), pp. 761 – 767. Android Malware Classification Using LSTM Model | IIETA.
15. International Journal on "Classification of Image spam Using Convolution Neural Network", Traitement du Signal, Vol. 39, No. 1, (February 2022), pp. 363-369 .
16. International Journal on "Medical Image Classification Using Deep Learning Based Hybrid Model with CNN and Encoder", International information and Engineering Technology Association (IIETA), Revue d'IntelligenceArtificiellVol. 34, No. 5, (October, 2020), pp. 645 – 652.
17. International Journal on "Prediction of Hospital Re-admission Using Firefly Based Multi-layer Perception, International information and Engineering Technology Association (IIETA) Vol. 24, No. 4, (sept, 2020), pp. 527 – 533.
18. International Journal on "Energy efficient intrusion detection using deep reinforcement learning approach",Journal of Green Engineering (JGE),Volume-11, Issue-1,January 2021.625-641.
19. International Journal on "Classification of High Dimensional Class Imbalance Data Streams Using Improved Genetic Algorithm Sampling", International Journal of Advanced Science and Technology, Vol. 29, No. 5, (2020), pp. 5717 – 5726.
20. Dr. M. Ayyappa Chakravarthi etal. published Springer paper "Machine Learning-Enhanced Self-Management for Energy-Effective and Secure Statistics Assortment in Unattended WSNs" in Springer Nature (Q1), Vol 6, Feb 4th 2025
21. Dr. M. Ayyappa Chakravarthi etal. published Springer paper "GeoAgriGuard AI-Driven Pest and Disease Management with Remote Sensing for Global Food Security" in Springer Nature (Q1), Jan 20th 2025.
22. Dr. M. Ayyappa Chakravarthi etal. presented and published IEEE paper "Machine Learning Algorithms for Automated Synthesis of Biocompatible Nanomaterials" , ISBN 979-8-3315-3995-5, Jan 2025.
23. Dr. M. Ayyappa Chakravarthi etal. presented and published IEEE paper "Evolutionary Algorithms for Deep Learning in Secure Network Environments" ISBN:979-8-3315-3995-5, Jan 2025.
24. Dr. Ayyappa Chakravarthi M. etal, published Scopus paper "Time Patient Monitoring Through Edge Computing: An IoT-Based Healthcare Architecture" in Frontiers in Health Informatics (FHI), Volume 13, Issue 3, ISSN-Online 2676-7104, 29th Nov 2024.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

25. Dr. Ayyappa Chakravarthi M. etal, published Scopus paper "Amalgamate Approaches Can Aid in the Early Detection of Coronary heart Disease" in Journal of Theoretical and Applied Information Technology (JATIT) , Volume 102, Issue 19, ISSN 1992-8645, 2nd Oct 2024.
26. Dr. Ayyappa Chakravarthi M, etal, published Scopus paper "The BioShield Algorithm: Pioneering Real-Time Adaptive Security in IoT Networks through Nature-Inspired Machine Learning" in SSRG (Seventh Sense Research Group) -International Journal of Electrical and Electronics Engineering (IJEEE), Volume 11, Issue 9, ISSN 2348-8379, 28th Sept 2024.
27. Ayyappa Chakravarthi M, Dr M. Thillaikarasi, Dr Bhanu Prakash Battula, published SCI paper "Classification of Image Spam Using Convolution Neural Network" in International Information and Engineering Technology Association (IIETA) - "Traitement du Signal" Volume 39, No. 1.
28. Ayyappa Chakravarthi M, Dr. M. Thillaikarasi, Dr. Bhanu Praksh Battula, published Scopus paper "Classification of Social Media Text Spam Using VAE-CNN and LSTM Model" in International Information and Engineering Technology Association (IIETA) - Ingénierie des Systèmesd'Information (Free Scopus) Volume 25, No. 6.
29. Ayyappa Chakravarthi M, Dr. M. Thillaikarasi, Dr. Bhanu Praksh Battula, published Scopus paper "Social Media Text Data Classification using Enhanced TF_IDF based Feature Classification using Naive Bayesian Classifier" in International Journal of Advanced Science and Technology (IJAST) 2020.
30. Ayyappa Chakravarthi M. presented and published IEEE paper on "The Etymology of Bigdata on Government Processes" with DOI 10.1109/ICICES.2017.8070712 and is Scopus Indexed online in IEEE digital Xplore with Electronic ISBN: 978-1-5090-6135-8, Print on Demand(PoD) ISBN:978-1-5090-6136-5, Feb'2017.
31. Subba Reddy Thumu& Geethanjali Nellore, Optimized Ensemble Support Vector Regression Models for Predicting Stock Prices with Multiple Kernels. Acta Informatica Pragensia, 13(1), x–x. 2024.
32. Subba Reddy Thumu, Prof. N. Geethanjali. (2024). "Improving Cryptocurrency Price Prediction Accuracy with Multi-Kernel Support Vector Regression Approach". International Research Journal of Multidisciplinary Technovation 6 (4):20-31.
33. Dr syamsundararaothalakola et.al. published Scopus paper "An Innovative Secure and Privacy-Preserving Federated Learning Based Hybrid Deep Learning Model for Intrusion Detection in Internet-Enabled Wireless Sensor Networks " in IEEE Transactions on Consumer Electronics 2024.
34. Dr syamsundararaothalakola et.al. published Scopus paper "Securing Digital Records: A Synerigistic Approach with IoT and Blockchain for Enhanced Trust and Transparency " in International Journal of Intelligent Systems and Applications in Engineering 2024.
35. Dr syamsundararaothalakola et.al. published Scopus paper "A Model for Safety Risk Evaluation of Connected Car Network " inReview of Computer Engineering Research2022.
36. Dr syamsundararaothalakola et.al. published Scopus paper "An Efficient Signal Processing Algorithm for Detecting Abnormalities in EEG Signal Using CNN" in Contrast Media and Molecular Imaging 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com