# Prediction of Osteoporosis Risk Level Using Machine Learning Techniques

*Rohan Vanmali* [1], *Tejas Kashid* [2]
*William Rodrigues* [3], *Asher Rodrigues* [4], *and*
*Sonali Suryawanshi* [5]

## Abstract

Low bone density and bone tissue degeneration are prominent symptoms of osteoporosis, which increases the risk of fractures. To avoid long term consequences and enhance patient outcomes, osteoporosis fractures must be identified early and prevented. In this research, we offer a Machine Learning based method for calculating the risk of osteoporosis fractures based on a variety of inputs, including age, gender, weight, height, smoking, alcohol use, diabetes, arthritis, parental fractures, and *T*-score. We use the Logistic Regression, K-Nearest Neighbour, and Support Vector Machine Machine Learning models to predict the risk level of osteoporosis fractures, which may be high risk, medium risk, or low risk. We evaluate the performance of these models based on a number of factors, such as accuracy, precision, recall, and *F*1-score. The estimated risk level of osteoporosis fractures is stored in a database together with other input data for future use.

***Keywords :*** Bone Mass Density, Fracture, Osteoporosis

## I. INTRODUCTION

Low bone mass and structural bone tissue degeneration, which increase the risk of fractures, are symptoms of osteoporosis, a chronic and progressive illness. Osteoporotic fractures can cause significant morbidity, death, and healthcare expenditures, and they are a significant public health problem, particularly in aging populations. To avoid long-term consequences and enhance patient outcomes, osteoporosis fractures must be identified early and prevented.

The precision and effectiveness of osteoporosis fracture analysis have new prospects, thanks to recent developments in Machine Learning and Artificial Intelligence. In order to detect risk variables and forecast the likelihood of osteoporosis fractures, machine learning algorithms can analyze vast databases of clinical and genetic data. These algorithms can also help doctors make better educated choices about the diagnosis, care, and monitoring of osteoporosis patients.

In light of this, the aim of the project is to create a Machine Learning based method for analyzing

Paper Submission Date : April 27, 2023 ; Paper sent back for Revision : May 7, 2023 ; Paper Acceptance Date : May 10, 2023 ; Paper Published Online : June 5, 2023

[1] R. Vanmali, *Student,* Department of Information Technology, Email : vanmalirohan58@gmail.com ;
ORCID iD : https://orcid.org/0009-0001-4043-8424
[2] T. Kashid, *Student,* Department of Information Technology, Email : tejaskashid2901@gmail.com ;
ORCID iD : https://orcid.org/0009-0005-5443-5020
[3] W. Rodrigues, *Student,* Department of Information Technology, Email : williamrathod12@gmail.com ;
ORCID iD : https://orcid.org/0009-0008-6431-3408
[4] A. Rodrigues, *Student*, Email : asherrodrigues21@gmail.com ; ORCID iD : https://orcid.org/0009-0005-3389-2855
[5] S. Suryawanshi, *Assistant Professor*, Email : skudwe@gmail.com ; ORCID iD : https://orcid.org/0000-0002-3695-563X

[1,2,3,4,5] St. Francis Institute of Technology, Sardar Vallabhbhai Patel Road, Near Bhagwati Hospital, Mount Poinsur, Borivali West - 400 103, Mumbai, Maharashtra.

**DOI : https://doi.org/10.17010/ijcs/2023/v8/i3/172863**

osteoporosis fracture risk that takes into account a variety of input factors including age, gender, weight, height, smoking, alcohol use, diabetes, arthritis, parental fractures, and *T*-score. This method will be used to predict the risk level of osteoporosis fractures. We employ three distinct machine learning models to predict the risk level of osteoporosis fractures, which might be high risk, medium risk, or normal using Logistic Regression, K-Nearest Neighbour, and Support Vector Machine algorithms. Along with other input information, the projected risk level of osteoporosis fractures is stored in a database for future use.

By giving precise and reliable forecasts of the risk level of osteoporosis fractures, the proposed approach has the potential to help clinicians diagnose osteoporosis fractures. The approach can also aid in identifying people at high risk for fractures from osteoporosis who need more frequent monitoring and management techniques. These project results may lead to better patient outcomes and lower healthcare expenses for osteoporosis fractures.

## II. LITERATURE REVIEW

Intelligent Predictive Osteoporosis System: Using decision tree techniques like ID3, C4.5, and Random Forest as well as simulation, the model for the early identification of osteoporosis was created. The outcomes demonstrated that the random forest decision trees algorithm outperformed other decision trees algorithms with a 99.9% accuracy rate [1].

In [2], a non-linear model was built to examine the BMD relationship, food, and choices of lifestyle for 305 samples of postmenopausal women. According to the findings, increasing calcium intake, getting enough sun exposure, maintaining a healthy weight, engaging in regular exercise, and consuming enough calories were the key ways to prevent post-menopausal women from losing bone mass [2].

In [3], Genomic and Phenotypic data of 5130 older men were compared. Each participant's 1,103 linked SNPs were used to create a genetic risk score (GRS). The development of prediction models for major osteoporotic fractures individually using GRS, bone density, and other risk factors as predictors involved the use of random forest, gradient boosting, neural network, and logistic regression models. More accurate than the logistic regression model were the random forest, gradient boosting, and neural networks [3].

In order to predict osteoporosis, in [4] the system compared various classification algorithms employing wrapper-based feature selection. Multilayer feed-forward neural network (MFNN), Naive Bayes, and Logistic Regression were used as classification techniques. A subset of significant SNPs was also found using a wrapper-based feature selection technique [4].

In [5], Weka system is used to analyze the experimental study and determine the best classifiers, which were then enhanced and utilized to better forecast whether a woman will continue to have osteopenia or turn osteoporotic. Despite the modest sample size of the study, the classification accuracy was unsatisfactory. For increased accuracy, a bigger sample size and more sophisticated techniques are required. Another drawback is that classification algorithms need accurate parameter sets, which is a difficult and unavoidable problem [5].

In [6], using identified risk factors, the project created the best bone disease prediction model. Following the determination of the initial risk variables for predicting the start of bone illnesses, pre-training and fine tuning were used. The Deep Belief Network (DBN) was used to create the model, and the study's findings showed that by incorporating pertinent variables, the predictive model's performance was enhanced [6].

In [7], the study examines the various feature selection methods offered by WEKA and evaluates the advantages of utilizing the most promising features in terms of execution speed and accuracy. Instances successfully identified, processing speed, Kappa statistics, and mean absolute value were compared between studies using and excluding the feature selection technique. IBk training set testing without feature selection produced the best results. However, IBK, J48, SMO, JRip, and bagging methods with feature selection produced the greatest outcomes [7]. The illustration of a single approach with the combination of two or more MLMs by applying layered form approaches rather than hybrid approaches will efficiently lead to fast detection [8].

## III. METHODOLOGY

### *A. Problem Definition*

The aim of this study is to develop Machine Learning model which correctly predicts the probability of osteoporotic fractures based on a number of risk factors. The model should be rapid, easy to use, and understandable so that doctors can quickly identify people at a huge risk of osteoporotic fractures. Additionally, it must be suitable for use in a clinical setting for professionals.

### B. Scope

An enhanced risk of fractures, morbidity, and death are associated with osteoporosis, a chronic condition. Even though long-term treatment for osteoporosis may be necessary, nothing is known about the long-term hazards associated with these medications. The project is state of the art and can be improved more on the basis of big datasets, different training models etc. It can also be used for Research and Development purpose to predict outcomes easily. Basically, this project has a wide scope as prediction of Osteoporosis should be done at an early stage. Many patients stay unaware regarding Osteoporosis even when they suffer from it. Hence, this project can also be used where a person can check the chances of getting osteoporosis based on the parameters given in the system.

### C. Proposed Methodology

**1) *Data preprocessing*:** The dataset must first be preprocessed to eliminate any errors or missing information, normalize the data, and put it in a format that the machine learning algorithms can use. The dataset must be cleaned, missing values must be handled, the data must be normalized, and categorical features must be encoded.

**2) *Feature Selection*:** The following stage is to choose the most pertinent input features for determining the degree of osteoporosis fracture risk. Various feature selection methods, including correlation analysis and feature importance are used in this step.

**3) *Data Splitting*:** After feature selection and preprocessing, the dataset is split into two subsets: a training set and a testing set. The training set is used to develop the machine learning models, whereas the testing set is used to evaluate how well they function.

**4) *Model Training*:** The three Machine Learning methods used in this study are K-Nearest Neighbour, Logistic Regression, and Support Vector Machine. On the training dataset, these algorithms are trained using a supervised learning technique. To calculate the risk level of osteoporosis fractures, the model's parameters are trained using training data.

**5) *Model Evaluation*:** Using the testing dataset, the models' performance is evaluated once they have been trained. A number of factors, including the $F1$-score, recall, accuracy, and precision are used to evaluate the performance. Best performing model is selected for deployment after each model's performance has been assessed.

**6) *Deployment*:** To help clinicians forecast the likelihood of osteoporosis fractures, the chosen model is put to use in a clinical context. The model can be designed as a standalone application that can be used on a desktop or mobile device or it can be integrated into Electronic Health Records (EHRs).

**7) *Performance Evaluation*:** Finally, in clinical settings, the performance of the deployed model is assessed by contrasting its projections with the actual patient outcomes. This step is crucial to verifying the proposed system's precision, safety, and efficiency.

### D. Proposed Algorithm

**1) *Support Vector Machine (SVM)*:** The SVM is a powerful technique that is commonly used in classification tasks. It works by finding the hyperplane that categorizes the data and maximizing the distance between the hyperplane and the closest data points. The computationally effective SVM performs well for both linear and nonlinear classification problems. It also has resistance to outliers and noisy data.

**2) *K-Nearest Neighbor (K-NN)*:** To begin, the KNN algorithm trains on a dataset of labeled samples. The algorithm determines the separation between the new data point and each training example in the feature space during the testing phase. Then, depending on their distances, the $k$ closest neighbours to the new data point are determined. An essential KNN algorithm parameter is the value of $k$. The model's performance and accuracy can be impacted by the choice of $k$.

**3) *Logistic Regression*:** Logistic regression is a straightforward but effective method that is frequently used in classification applications. It is a linear model that converts the input data into a probability score using a logistic function. It can handle problems involving binary and multiple classes in classification and is computationally effective. Additionally, interpretable logistic regression allows clinicians to quickly comprehend and evaluate the model's parameters.

## IV. SYSTEM DESIGN

System Design is shown in Fig. 1 and System Architecture is shown in Fig. 2.

## V. RESULTS AND ANALYSIS

Based on the examination of the created Machine Learning model to predict the risk level of osteoporotic fractures, the following conclusions were made:

According to the input features, Logistic Regression is the technique that can most reliably classify individuals as having a high, medium, or normal risk of osteoporotic fractures, with an accuracy rate of 83%.(according to Fig. 3).

The accuracy of K-Nearest Neighbour is 62%, which

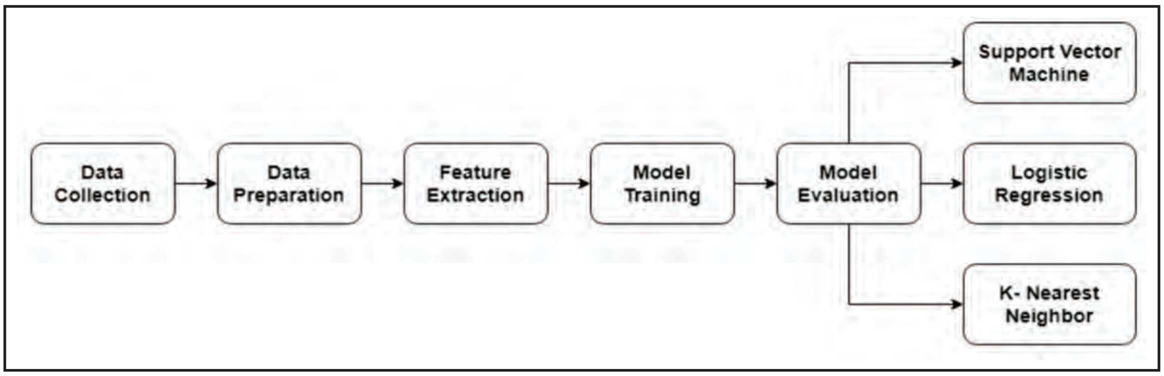


**Fig. 1. High Level Design**

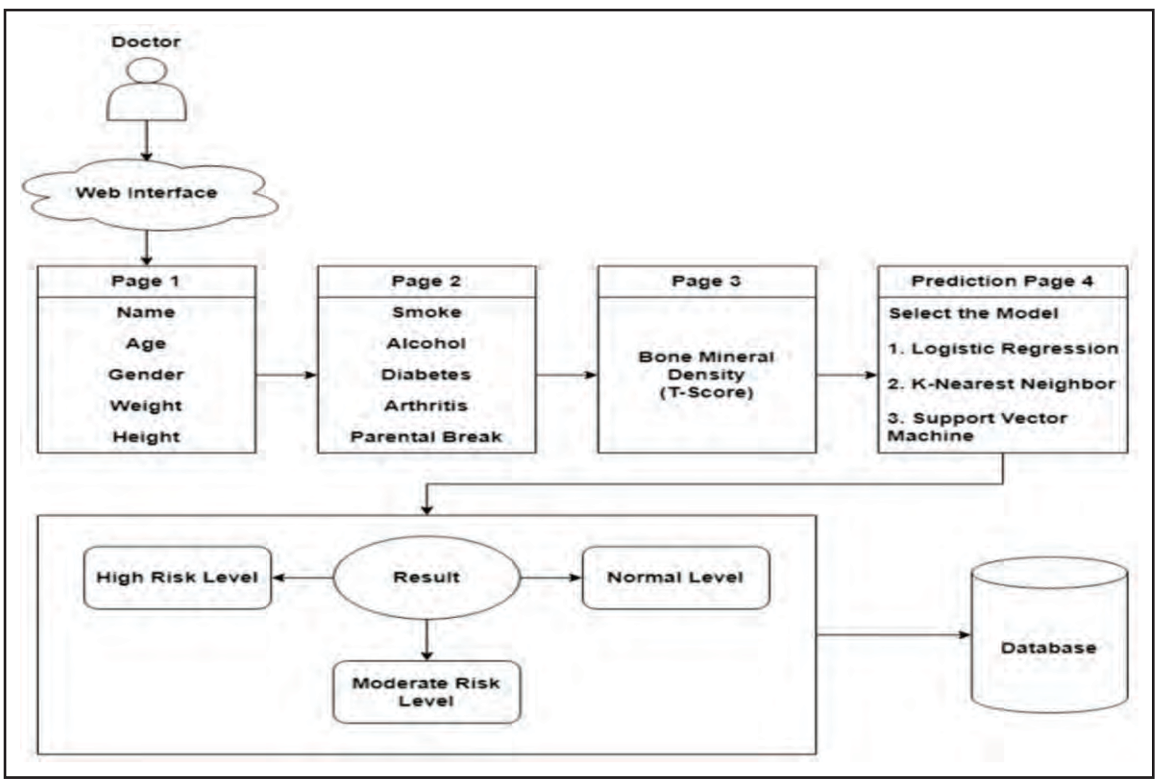


**Fig. 2. System Architecture**

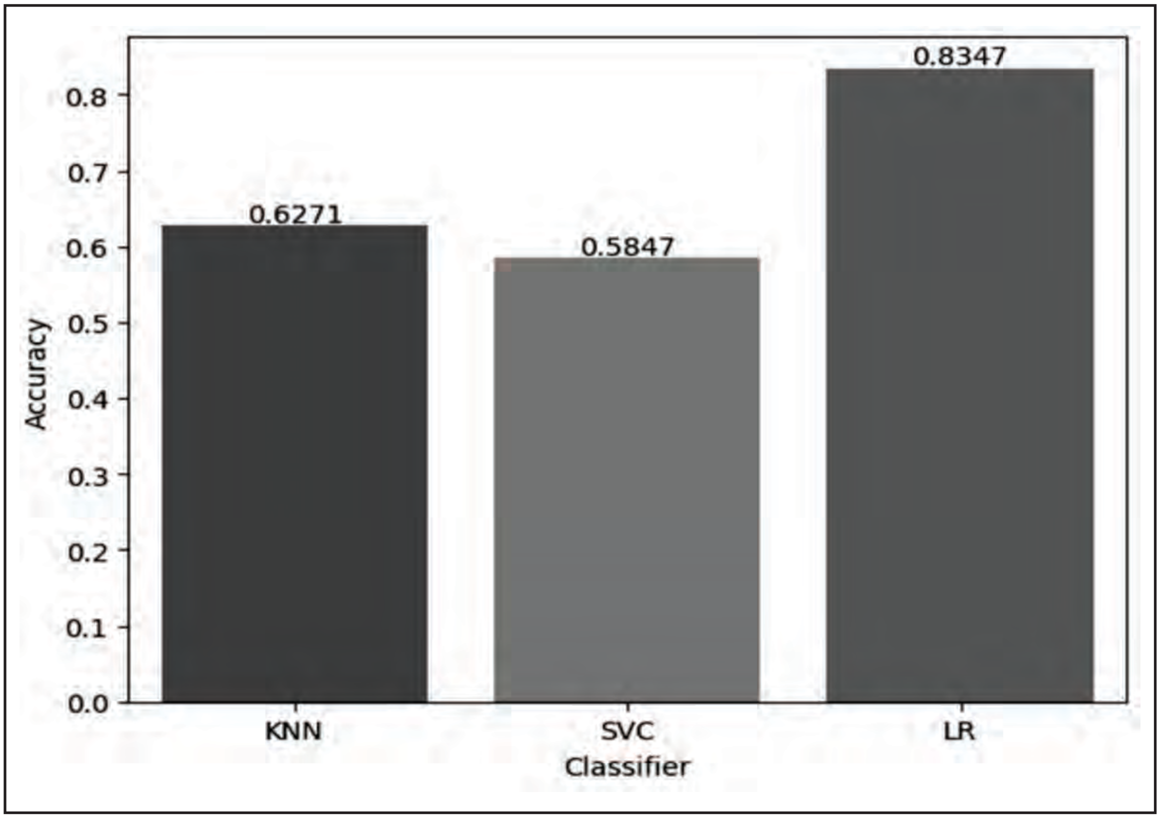


**Fig. 3. Accuracy of Each Algorithm**

**TABLE I.**

**CLASSIFICATION REPORT FOR LOGISTIC REGRESSION**

| | *Precision* | *Recall* | *F1-score* | *Support* |
|---|---|---|---|---|
| ***High*** | 0.81 | 0.81 | 0.81 | 48 |
| ***Low*** | 0.89 | 0.92 | 0.90 | 107 |
| ***Moderate*** | 0.77 | 0.74 | 0.75 | 81 |
| ***Accuracy*** | | | 0.83 | 236 |
| ***Macro average*** | 0.82 | 0.82 | 0.82 | 236 |
| ***Weighted average*** | 0.83 | 0.83 | 0.83 | 236 |

**TABLE II.**

**CLASSIFICATION REPORT FOR K-NEAREST NEIGHBOR**

| | *Precision* | *Recall* | *F1-score* | *Support* |
|---|---|---|---|---|
| ***High*** | 0.56 | 0.52 | 0.54 | 48 |
| ***Low*** | 0.73 | 0.79 | 0.76 | 107 |
| ***Moderate*** | 0.51 | 0.48 | 0.50 | 81 |
| ***Accuracy*** | | | 0.63 | 236 |
| ***Macro average*** | 0.60 | 0.60 | 0.60 | 236 |
| ***Weighted average*** | 0.62 | 0.63 | 0.62 | 236 |

TABLE III.

CLASSIFICATION REPORT FOR SUPPORT VECTOR MACHINE

| | *Precision* | *Recall* | *F1-score* | *Support* |
|---|---|---|---|---|
| ***High*** | 0.00 | 0.00 | 0.00 | 48 |
| ***Low*** | 0.70 | 0.84 | 0.77 | 107 |
| ***Moderate*** | 0.44 | 0.59 | 0.51 | 81 |
| ***Accuracy*** | | | 0.58 | 236 |
| ***Macro average*** | 0.38 | 0.48 | 0.42 | 236 |
| ***Weighted average*** | 0.47 | 0.58 | 0.52 | 236 |

is lower than that of Logistic Regression. This suggests that when categorizing individuals based on the input features, the KNN method may not be as effective as Logistic Regression.

The accuracy of the Support Vector Machine method, which was tested, is the lowest of the three, at 58%. This suggests that depending on the input features, SVM may not be as accurate in estimating the degree of osteoporotic fracture risk.

Among the three algorithms tried for this project, the Logistic Regression algorithm has generally proven to be the most effective. A potentially helpful tool in clinical settings for determining a patient's risk of osteoporotic fractures is the model created using Logistic Regression, which can predict the level of osteoporotic fracture risk with an accuracy of 83% and the results are stored in a database.

## VI. CONCLUSION

To conclude, the Machine Learning model that was created to estimate the degree of risk for osteoporotic fractures has yielded encouraging findings. Age, gender, weight, height, use of tobacco and alcohol, diabetes, arthritis, parental fracture, and *T*-Score were some of the input characteristics used to train the model. The model was examined using three different algorithms: Support Vector Machine, K-Nearest Neighbour, and Logistic Regression.

According to the evaluation's findings, Support Vector Machine had a 58% accuracy rating, K-Nearest Neighbour had a 62% accuracy rating, and Logistic Regression had an accuracy rating of 83%. These findings suggest that the most effective technique for estimating the risk of osteoporotic fractures based on the input features is Logistic Regression.

The created model may be utilized in clinical settings to evaluate patient risk for osteoporotic fractures. It can assist medical practitioners in making well-informed choices regarding the management and treatment plans for individuals with osteoporosis.

## ACKNOWLEDGMENT

The authors extend their sincere gratitude to everyone who helped with this initiative by offering advice, resources, and support. First and foremost, they would like to thank their project mentors, Sonali Suryawanshi and Amrita Mathur for their constant guidance, feedback, and support. Their insightful opinions and helpful recommendations played a crucial role in determining the project's course and results.

Finally, they would like to thank all of the research team members who put in countless hours to gather and analyze the data. Their commitment and diligence have been important in making this project a success.

## AUTHORS' CONTRIBUTION

Rohan Vanmali played a crucial role in developing and fine tuning the machine learning algorithms and models. His dedication was instrumental in achieving accurate predictions and identifying potential indicators of Osteoporosis at an early stage. His involvement in the research phase demonstrated his commitment to understanding the underlying principles of Osteoporosis and exploring the most effective approaches to early detection. Tejas Kashid delved deep into scientific literature studying various methodologies and existing research findings related to Osteoporosis detection. He analyzed the latest advancements in machine learning techniques and adapted them to suit the specific requirements of the project. The contribution of William Rodrigues to the project was highly valuable as he skillfully designed an intuitive and visually appealing user interface. His attention to detail and understanding of user experience greatly enhanced the usability and accessibility of the application. The involvement of Asher Rodrigues in the project as a tester was invaluable. He meticulously conducted extensive testing, ensuring the reliability and robustness of the application. As the college project guide, Sonali Suryawanshi provided guidance, mentorship, and oversight throughout the development process. Her extensive knowledge and experience in the field of data mining and Machine Learning helped shape the project's direction and ensured its adherence to academic standards.

## CONFLICT OF INTEREST

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in the manuscript.

## FUNDING ACKNOWLEDGEMENT

The authors have not received any financial support for the research, authorship, and/or for the publication of the article.

## REFERENCES

[1] W. Moudani, A. Shahin, F. Chakik, and D. Rajab, "Intelligent Predictive Osteoporosis System," *Int. J. Comput. Appl.*, vol. 32, no. 5, pp. 28–37, Oct. 2011, doi: 10.5120/3901-5468.

[2] J.A. Kanis, C. Cooper, R. Rizzoli, J. –Y. Reginster, "European guidance for the diagnosis and management of Osteoporosis in postmenopausal women," *Osteoporosis Int.*, vol. 30, pp. 399–428, Oct. 2015, doi: 10.1007/s00198-018-4704-5.

[3] Q. Wu, F. Nasoz, J. Jung, B. Bhattarai, M. V. Han, "Machine Learning approaches for fracture risk assessment: A comparative analysis of Genomic and Phenotypic data in 5130 older men," *Calcif Tissue Int.*, vol. 107, pp. 353–361, Oct. 2020, doi: 10.1155/2013/850735.

[4] H.-W. Chang, Y.-H. Chiu, H.-Y. Kao, C.-H., Yang, and W.- H. Ho, "Comparison of classification algorithms with wrapper-based feature selection for predicting Osteoporosis outcome based on genetic factors in a Taiwanese women population," *Int. J. Endocrinology*, vol. 1–8, Jan. 2013, doi: 10.1155/2013/850735.

[5] N. Guannoni., R. Sassi., W. Bedhiafi, and M. Elloumi., "A Comparison Between Classification Algorithms for Postmenopausal Osteoporosis Prediction in Tunisian Population," In: Renda, M., Bursa, M., Holzinger, A., Khuri, S. (eds) *Inf. Tech. Bio- Medical Inform*. ITBAM 2016. Lecture Notes in Comput. Sci. (), vol 9832, pp. 234–248, Aug. 2016, Springer, Cham, doi: 10.1007/978-3-319-43949-5_19.

[6] M. Saranya and K. Sarojimi, "An improved and optimal prediction of bone disease based in risk factors," *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 2 , pp. 820–823, 2016. [Online]. Available: https://ijcsit.com/docs/Volume%207/vol7issue2/ijcsit2016070283.pdf

[7] Z. Wang., J. Fu,. R. Yan., J. S. Park, and H. G. Kim, "A comparative analysis of Machine Learning techniques for early detection of Osteoporosis," *Int. J. Pure Appl. Math.*, vol. 119, no. 18, pp. 97–114, 2018.[Online]. Available: https://acadpubl.eu/hub/2018-119-18/1/9.pdf

[8] T. Agarwal, H. Sharma, C. Latha, and S. Gupta, "A review of role of machine learning models in coronary heart disease detection accuracy," *Indian J. of Comput. Sci.*, vol. 7, no. 1, pp. 36–44, 2022. doi:10.17010/ijcs/2022/v7/i1/168955.

## About the Authors

**Rohan Vanmali** is a Full Stack Developer with a proven track record of delivering robust and scalable web applications. He is skilled in both front-end and back-end technologies, proficiently handles database management, and API integrations. He is proficient in programming languages such as JavaScript, Python, and Java, and is adept at using frameworks like React, Node.js, and Django. He is a collaborative team player with excellent problem solving abilities and a passion for staying updated with the latest industry trends. He is committed to delivering high quality solutions that meet clients' needs while adhering to best coding practices.

**Tejas Kashid** is an IT engineer with a deep interest in web development. He is proficient in web development technologies and is dedicated to creating visually appealing and user friendly websites. He is also proficient in crafting innovative and user friendly mobile applications that seamlessly integrate cutting edge features and provide a delightful user experience. He is versatile and keen in constantly expanding knowledge to excel in various development domains.

**William Rodrigues** is a Python developer and a skilled programmer who has specialized in using the Python programming language. He designs, develops, and maintains software applications, writes efficient code, and creates solutions for various industries, leveraging Python's versatility, simplicity, and extensive libraries.

**Asher Rodrigues** is a highly skilled IT engineer specializing in Data Science. With a strong background in computer science and mathematics, he has dedicated his career to unlocking valuable insights from complex data sets. Asher's passion for data driven decision making and his commitment to staying at the forefront of advancements in the field makes him an invaluable asset in the realm of data science.

**Sonali Suryawanshi** is a Computer Engineer and an Assistant Professor in an esteemed engineering college with more than 14 years of experience. Her areas of expertise are Data mining, Machine Learning, data structures. With a genuine passion and love for teaching, she thrives to inspire students and creates an engaging learning environment for them. She is proficient with programming languages like C, Java, Python and has conducted various workshops under skill development for students and faculties.