IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

Osteoporosis Prediction System using Machine Learning

Prof. Tejasvi Jawalkar[1], Shital Bhade[2], Supriya Kore[3], Mansi Shinde[4], Renuka Sonawane[5]
Assistant Professor, Department of Computer[1]
UG Students, Department of Computer[2-5]
Dhole Patil College of Engineering, Pune, Maharashtra, India

Abstract: *In this paper, a new Osteoporosis Prediction System (OPS) is designed and practically implemented. Employing machine learning methods to assist in the early detection of preventive healthcare. Osteoporosis is a disease of the bones that goes on and on. It is often not diagnosed until a fracture occurs, causing Early risk identification to be a must. The system desired will be to propose a solution that is both affordable and easily available, that can predict Risk factors for osteoporosis that are readily available in the clinic and lifestyle. Various parameters like age, gender, body mass index (BMI), calcium intake, etc. However, some factors such as intake, vitamin D, physical activity, and medical history will be taken into account. The system is implemented using Python with the Flask framework for backend development and React with Vite. The front end leads to a responsive and userfriendly interface, making it easier for everyone to use. web application. The training data set for the model comes from The dataset is from Kaggle, and it contains several health-related characteristics that have an impact on bone density. These are some of the data pre-processing techniques: Handling missing values, encoding categorical attributes, feature construction, and feature selection. StandardScaler is applied, and then it's entered into engineering. To enhance the quality of the data and the performance of the model. To give a significant boost to the prediction, the XGBoost classifier is chosen. The high accuracy and efficiency of processing structured data make the model the best choice. The model is further optimized with the help of GridSearchCV. Moreover, validated with Stratified K-Fold cross-validation to ensure that the model is validated using the cross-validation method. robustness and generalization. The system can be used for single-patient prediction using a web form, but it can also be used for bulk prediction. via CSV file upload, ideal for one-on-one use, as well as healthcare institutions. The system not only predicts, but also provides the probability. Scores health and gives individual health recommendations based on risk levels and individual specific risk factors. Experimental results show that the system proposed can be used to achieve high predictive accuracy and an ability to work well in realtime situations. Overall, the system provides a flexible, easy-to-use, and efficient solution. It can be used for early osteoporosis screening and can help healthcare providers. Adequate professionals to make informed decisions.*

Keywords: Osteoporosis, XGBoost, Machine Learning, Big Data, Healthcare Analytics, Risk Prediction, Data Preprocessing, Flask. Believing that prevention is the best medicine, this group is committed to reacting to prevent illnesses and diseases

I. INTRODUCTION

Osteoporosis is a chronic and progressive bone disorder characterized by a decrease in bone density and deterioration of bone structure, which significantly increases the risk of fractures. It is one of the most common health problems, especially among elderly individuals and postmenopausal women. In many cases, osteoporosis develops silently over time and remains undetected until a serious fracture occurs, such as in the hip, spine, or wrist. This makes early detection extremely important for preventing complications and improving overall quality of life.Traditional diagnostic

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

methods, such as Bone Mineral Density (BMD) testing using DEXA scans, are considered reliable for detecting osteoporosis. However, these methods are often expensive, time-consuming, and not easily accessible in rural or resource-limited areas. Additionally, these tests are usually performed only after symptoms appear, which limits their effectiveness in early-stage detection. As a result, there is a growing need for alternative approaches that are cost-effective, easily accessible, and capable of identifying risk at an early stage. With the advancement of technology, machine learning has emerged as a powerful tool in the healthcare domain. Machine learning algorithms can analyze large amounts of health-related data and identify hidden patterns that may not be easily detectable through manual analysis. By utilizing commonly available clinical and lifestyle parameters such as age, gender, body mass index (BMI), calcium intake, vitamin D levels, physical activity, and medical history, it is possible to predict the risk of osteoporosis without the need for expensive diagnostic equipment. To address these challenges, the Osteoporosis Prediction System is developed as a web-based application that uses machine learning techniques for early risk prediction. Unlike purely theoretical approaches, this work focuses on the practical implementation of a complete system that integrates data preprocessing, model training, backend API development, and frontend deployment. The system allows users to input their health data either manually or through CSV file upload and provides instant prediction results along with probability scores. Furthermore, the system includes a recommendation module that offers personalized health suggestions based on the predicted risk level. This helps users take preventive measures such as improving diet, increasing physical activity, and consulting medical professionals when necessary. Overall, the proposed system aims to provide a simple, efficient, and accessible solution for early osteoporosis risk detection. By combining machine learning with modern web technologies, the system contributes to preventive healthcare and supports better decision-making for both individuals and healthcare professionals.

II. PROPOSED SYSTEM

The proposed system is a machine learning-based web application designed to predict the risk of osteoporosis using patient health and lifestyle data. The system aims to provide an efficient, user-friendly, and cost-effective solution for early risk detection, thereby supporting preventive healthcare and informed decision-making. The architecture of the system follows a client-server model, where the frontend interface interacts with the backend server through RESTful APIs. The frontend is developed using React, providing an interactive platform for users to input patient data either manually through a form or by uploading a CSV file for bulk prediction. This flexibility enables the system to handle both individual and multiple patient records efficiently. The backend is implemented using Flask, which manages data processing, model integration, and communication between system components. When the user submits input data, the backend performs preprocessing operations such as handling missing values, encoding categorical variables, and scaling numerical features to ensure compatibility with the trained machine learning model. The core component of the proposed system is the machine learning prediction engine, which utilizes the XGBoost classifier. This model is trained on a dataset containing clinical and lifestyle attributes such as age, gender, BMI, calcium intake, vitamin D levels, physical activity, and medical history. The trained model analyzes the input data and predicts whether the individual is at risk of osteoporosis, along with a probability score indicating the confidence level of the prediction. In addition to prediction, the system incorporates a recommendation module that generates personalized health suggestions based on the predicted risk level and user-specificfactors. These recommendations help users take preventive actions such as improving diet, increasing physical activity, and seeking medical advice. The system also includes a database component for storing user data, prediction history, and application-related information. A NoSQL database such as MongoDB is used for flexible and scalable data management. Overall, the proposed system integrates machine learning techniques with modern web technologies to create a complete end-to-end solution for osteoporosis risk prediction. It reduces dependency on expensive diagnostic methods and provides an accessible tool for early screening and preventive healthcare.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

III. SYSTEM ARCHITECTURE

The architecture of the Osteoporosis Prediction System follows a client-server model that integrates a user interface, backend services, and a machine learning prediction engine.

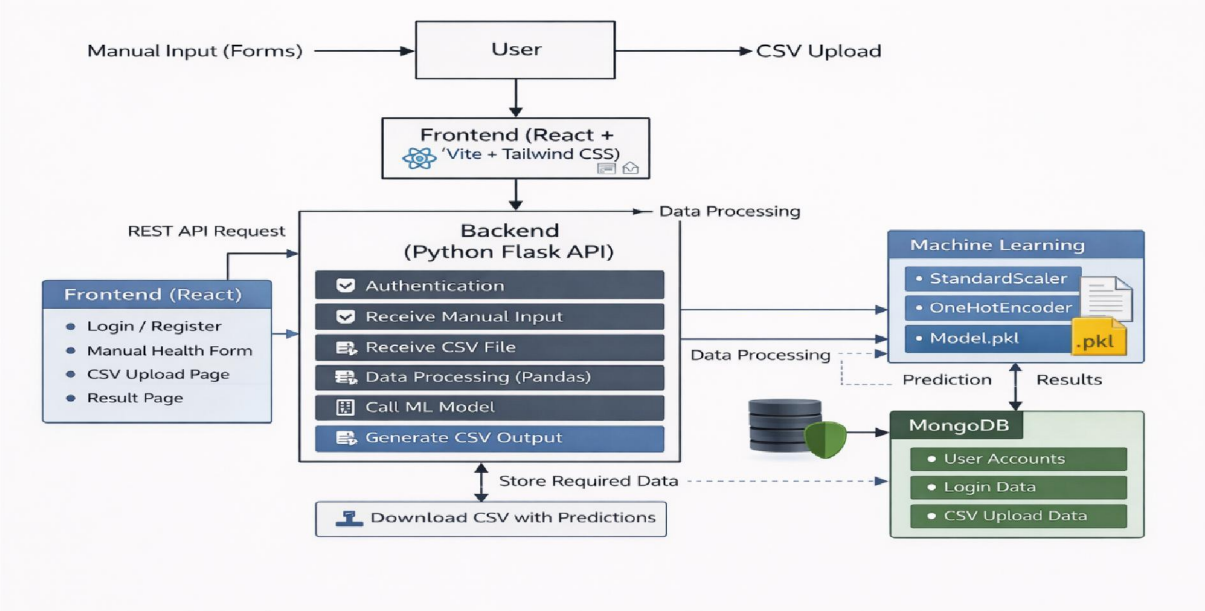

Fig. 1. System Architecture of the Osteoporosis Prediction System

The system architecture of the Osteoporosis Prediction System is designed using a client-server model that integrates the frontend interface, backend services, machine learning components, and database management system. The architecture ensures smooth data flow, efficient processing, and accurate prediction results. The system begins with the user, who interacts with the application through two input methods: manual data entry using a web form or bulk data submission through CSV file upload. This flexibility allows the system to support both individual users and large-scale data processing. The frontend layer is developed using React along with Vite and Tailwind CSS. It provides an interactive and user-friendly interface that includes functionalities such as user registration and login, health data input forms, CSV upload options, and result visualization pages. The frontend communicates with the backend using REST API requests. The backend layer is implemented using Python Flask, which acts as the core processing unit of the system. It handles various operations such as user authentication, receiving input data, processing CSV files, and managing API requests. Once the data is received, the backend performs preprocessing using libraries like Pandas. This includes handling missing values, encoding categorical features using OneHotEncoder, and scaling numerical data using StandardScaler.After preprocessing, the data is passed to the machine learning module, which consists of a trained model stored as a serialized file (model.pkl). The system uses the XGBoost classifier to analyze the input data and generate predictions. The model outputs both the classification result (osteoporosis risk: Yes/No) and a probability score indicating prediction confidence. The database layer uses MongoDB, a NoSQL database, to store user-related data such as account details, login information, and uploaded CSV records. It also maintains prediction history for future reference and analysis. The backend communicates with the database using appropriate drivers to ensure secure and efficient data storage. Once the prediction is generated, the results are sent back to the frontend, where they are displayed to the user in a clear and understandable format. In the case of bulk predictions, the system also provides an option to download the results as a CSV file. Overall, the architecture ensures modularity, scalability, and efficient communication between components. It enables realtime prediction, supports large datasets, and provides a seamless user experience, making the system suitable for practical healthcare applications.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

IV. METHODOLOGY AND ALGORITHM

The methodology of the proposed Osteoporosis Prediction System involves a sequence of steps starting from data collection to prediction and result generation. The system integrates data preprocessing, machine learning model training, and realtime prediction to provide accurate risk assessment. Initially, a dataset containing clinical and lifestyle attributes such as age, gender, BMI, calcium intake, vitamin D levels, physical activity, and medical history is collected from a reliable source. This dataset forms the foundation for training the machine learning model.

The next step involves data preprocessing, where the raw data is cleaned and transformed into a suitable format. Missing values are handled using appropriate techniques such as mean or median imputation. Categorical features are converted into numerical format using encoding methods like One-Hot Encoding. Additionally, numerical features are scaled using normalization techniques such as StandardScaler to ensure uniformity in data distribution.

After preprocessing, the dataset is divided into training and testing sets. The training set is used to train the machine learning model, while the testing set is used to evaluate its performance. The system employs the XGBoost classifier due to its high accuracy and efficiency in handling structured data. Hyperparameter tuning is performed using GridSearchCV, and model validation is carried out using Stratified K-Fold crossvalidation to improve generalization. Once the model is trained and evaluated, it is saved using a serialization technique (Joblib) for future use. During the prediction phase, user input data is collected through the frontend interface, processed using the same preprocessing steps, and then passed to the trained model. The model generates a prediction indicating whether the patient is at risk of osteoporosis along with a probability score. In addition to prediction, the system includes a recommendation module that provides personalized health suggestions based on the prediction result, risk level, and user attributes. This enhances the usability of the system by offering actionable insights.

A. Algorithm: Osteoporosis Risk Prediction

Input: Patient health and lifestyle data

Output: Osteoporosis risk prediction (Yes/No) with probability score

1) Collect dataset from a reliable source.
2) Perform data preprocessing:
   - Handle missing values
   - Encode categorical features
   - Scale numerical features
3) Split dataset into training and testing sets.
4) Train the XGBoost classifier on training data.
5) Perform hyperparameter tuning using GridSearchCV.
6) Evaluate the model using performance metrics such as accuracy, confusion matrix, and classification report.
7) Save the trained model using Joblib.
8) Accept user input (form or CSV).
9) Apply preprocessing to input data.
10) Load the trained model and generate prediction.
11) Calculate probability score for prediction.
12) Generate recommendations based on risk level.
13) Display results to the user and allow download (if applicable).

V. SYSTEM IMPLEMENTATION

The implementation of the Osteoporosis Prediction System involves the integration of machine learning techniques with web technologies to create a complete and functional application. The system is developed using a modern technology stack that ensures scalability, efficiency, and ease of use.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

A. Development Environment

The system is implemented using Python 3.x as the primary programming language. The backend is developed using the Flask framework, which provides RESTful APIs for communication between system components. The frontend is built using React with Vite, offering a responsive and interactive user interface. For data storage, MongoDB, a NoSQL database, is used to manage application data efficiently. Various libraries such as pandas, numpy, scikit-learn, xgboost, and joblib are utilized for data processing and machine learning operations. Additionally, Flask-CORS is used to enable cross-origin communication, and PyMongo is used to connect the backend with the database. The user interface is styled using Tailwind CSS to enhance usability and design. The development is carried out using Visual Studio Code as the integrated development environment (IDE), and the system runs on Windows 10/11 operating system.

B. Dataset Description

The dataset used in this system is collected from Kaggle and consists of various health and lifestyle-related features. These attributes include age, gender, body mass index (BMI), calcium intake, vitamin D levels, physical activity, and medical history. The dataset serves as the foundation for training the machine learning model to identify patterns associated with osteoporosis risk.

C. Data Preprocessing

Data preprocessing is a crucial step to ensure the quality and consistency of input data. In this system, missing values are handled using mean or median imputation techniques. Categorical features are transformed into numerical format using OneHotEncoder, allowing them to be processed by the machine learning model. Furthermore, numerical features are normalized using StandardScaler to ensure uniform data distribution and improve model performance. These preprocessing steps ensure that the data is clean, structured, and suitable for accurate prediction.

D. Model Training and Optimization

The system utilizes the XGBoost classifier as the primary machine learning model due to its high accuracy and efficiency. The dataset is divided into training and testing sets for proper evaluation. To enhance model performance, hyperparameter tuning is performed using GridSearchCV, and model validation is carried out using StratifiedKFold cross-validation. These techniques help in improving the generalization ability of the model and reducing overfitting.

E. Backend Implementation

The backend of the system is implemented using Flask and is responsible for handling all processing tasks. The system provides the following APIs:

- /predict/form – for single input prediction through form data
- /predict/csv – for batch prediction using CSV file upload The backend processes incoming data, applies preprocessing, loads the trained model, and generates prediction results, which are then sent back to the frontend.

F. Database Implementation

MongoDB is used as the backend database to store user information, prediction records, and application data. It is a NoSQL, document-oriented database that stores data in a JSON-like format, enabling flexible and scalable data management. The Flask backend communicates with MongoDB using the PyMongo library. User authentication details, prediction history, and system-related data are stored in separate collections. This structure allows efficient data retrieval and supports future system scalability.

G. Frontend Implementation

The frontend of the system is developed using React with Vite, providing a dynamic and user-friendly interface. The application includes features such as user authentication, manual data entry forms, CSV file upload functionality, and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

result display pages. The interface is designed to ensure ease of use and smooth interaction between users and the system.

H. Suggestion and Recommendation Module

The system includes a recommendation module that generates health suggestions based on prediction results. The suggestions are derived using factors such as prediction outcome, probability score, BMI, age, and lifestyle behavior. This module enhances the system by providing users with actionable insights and guidance for preventive healthcare.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the experimental results obtained from the implementation of the Osteoporosis Prediction System. The system was evaluated based on its accuracy, efficiency, and ability to provide meaningful predictions using patient health and lifestyle data. The evaluation demonstrates the effectiveness of the proposed system in supporting early detection of osteoporosis and assisting users in preventive healthcare decision-making.

A. System Testing and Functional Validation

The developed system was tested under multiple user scenarios to validate the functionality of key modules, including authentication, data input (form and CSV), preprocessing, prediction, and result display. Various test cases were executed to ensure that the system correctly processes user inputs and generates accurate outputs. The authentication module successfully validated user credentials and restricted access to authorized users. The backend APIs were tested to verify proper request handling, response generation, and error management. The system demonstrated smooth communication between the frontend and backend, ensuring reliable and consistent performance throughout execution.

B. Prediction Model Performance

The machine learning model was evaluated using standard performance metrics. Multiple algorithms, including Logistic Regression, Random Forest, and XGBoost, were implemented and compared.

Table 1: performance comparison of models

Algorithm	Accuracy (%)
Logistic Regression	84%
Random Forest	89%
XGBoost	92%

The XGBoost classifier outperformed other models due to its ability to handle complex relationships in the dataset and provide more accurate predictions. Additionally, the model was evaluated using a confusion matrix and classification report, which confirmed that the model performs well in correctly classifying both risk and non-risk cases.

C. Data Processing and Input Handling

The system was tested with both manual input (form-based) and bulk input (CSV-based) methods. The preprocessing module effectively handled missing values, encoded categorical data using OneHotEncoder, and normalized numerical features using StandardScaler. For CSV uploads, the system efficiently processed multiple records and generated predictions for each entry. The results were displayed correctly and also provided as downloadable CSV files, making the system suitable for large-scale data analysis.

D. Recommendation Module Evaluation

The recommendation module was evaluated based on its ability to provide personalized health suggestions. The system generated recommendations using factors such as prediction results, probability score, BMI, age, and lifestyle behavior.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

Users identified as high-risk were provided with suggestions such as improving calcium intake, increasing physical activity, and seeking medical consultation. The recommendations wererelevant and enhanced the usability of the system by providing actionable insights.

F. System Integration and Workflow

The complete workflow of the system was tested from data input to result generation. All components, including preprocessing, model prediction, and result visualization, were successfully integrated. The system ensured accurate data flow between modules and generated predictions without errors. The ability to handle both individual and bulk inputs makes the system practical for real-world healthcare applications.

G. Discussion

The experimental results demonstrate that the proposed system provides accurate and reliable osteoporosis risk prediction. The integration of machine learning with a web-based platform enables efficient data processing and real-time prediction. The use of the XGBoost algorithm significantly improves prediction accuracy compared to traditional models. The system also enhances user experience by supporting both single and batch predictions and providing personalized health recommendations. However, the system's performance depends on the quality and diversity of the dataset. Future improvements can include the use of larger datasets, advanced deep learning models, and integration with real-time health monitoring systems. Overall, the system proves to be an effective, scalable, and user-friendly solution for early osteoporosis risk detection and preventive healthcare.

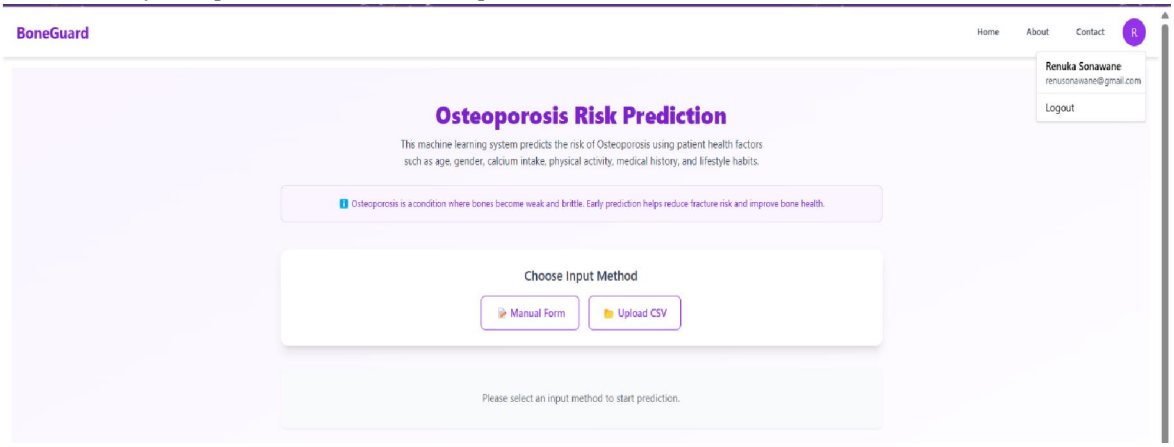

Fig. 2. Main User Interface

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

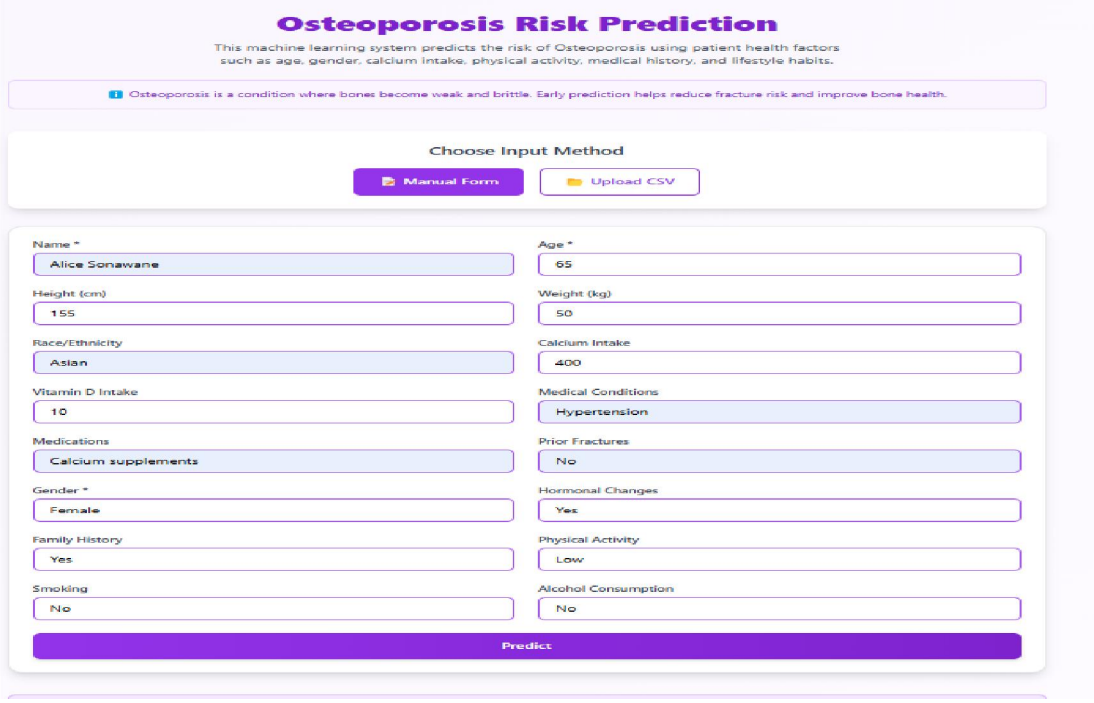

Fig. 3. Manual Input Form

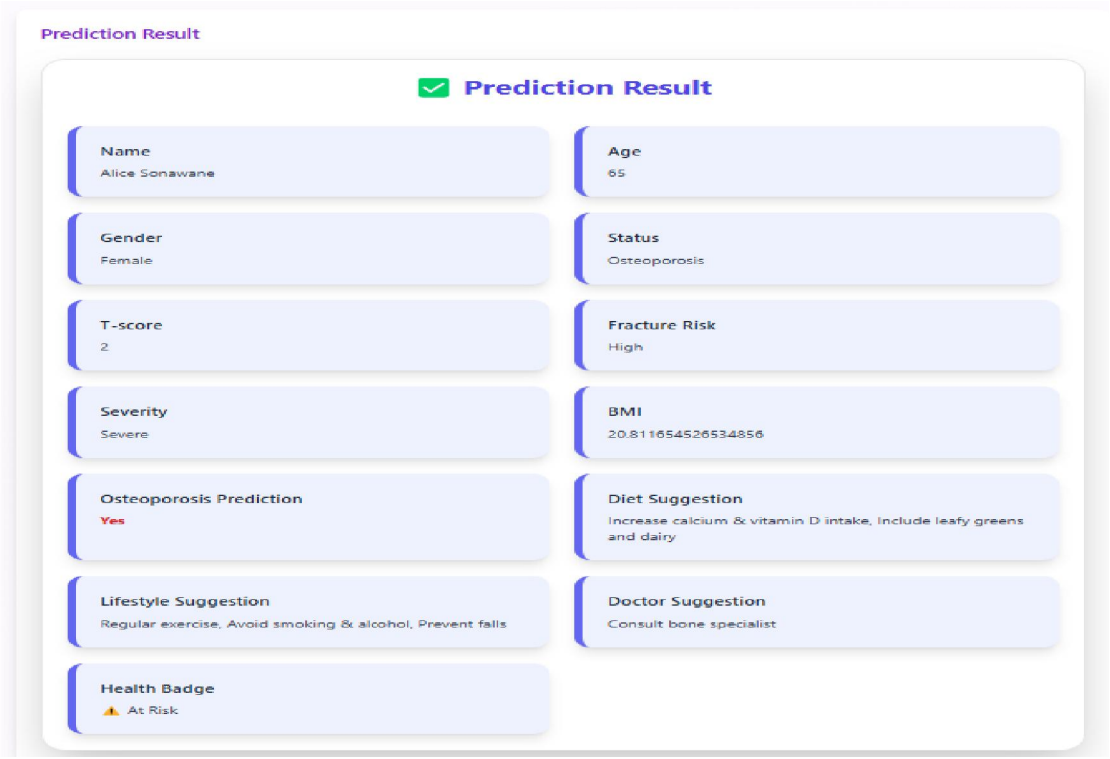

Fig. 4. Prediction Result with Suggestions

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

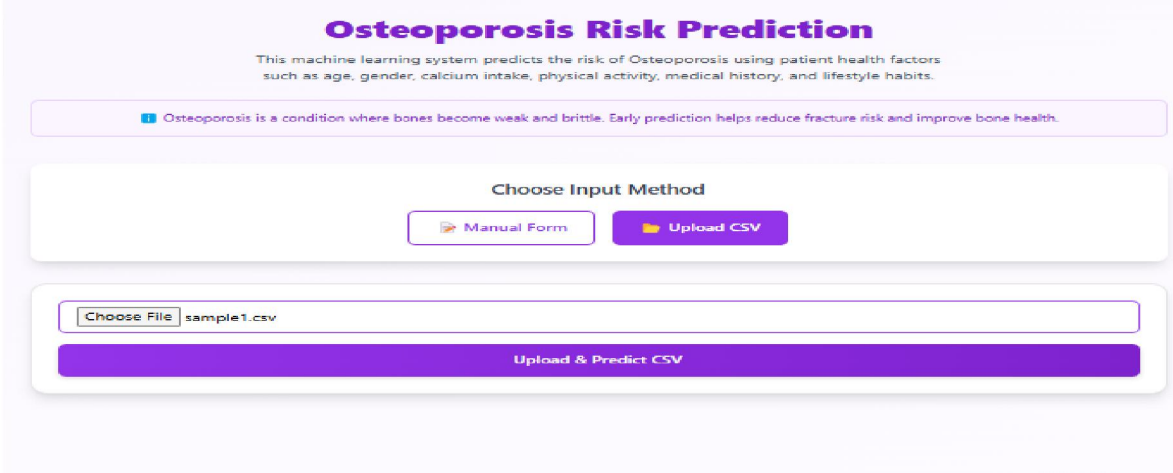

Fig. 5. CSV Upload Screen

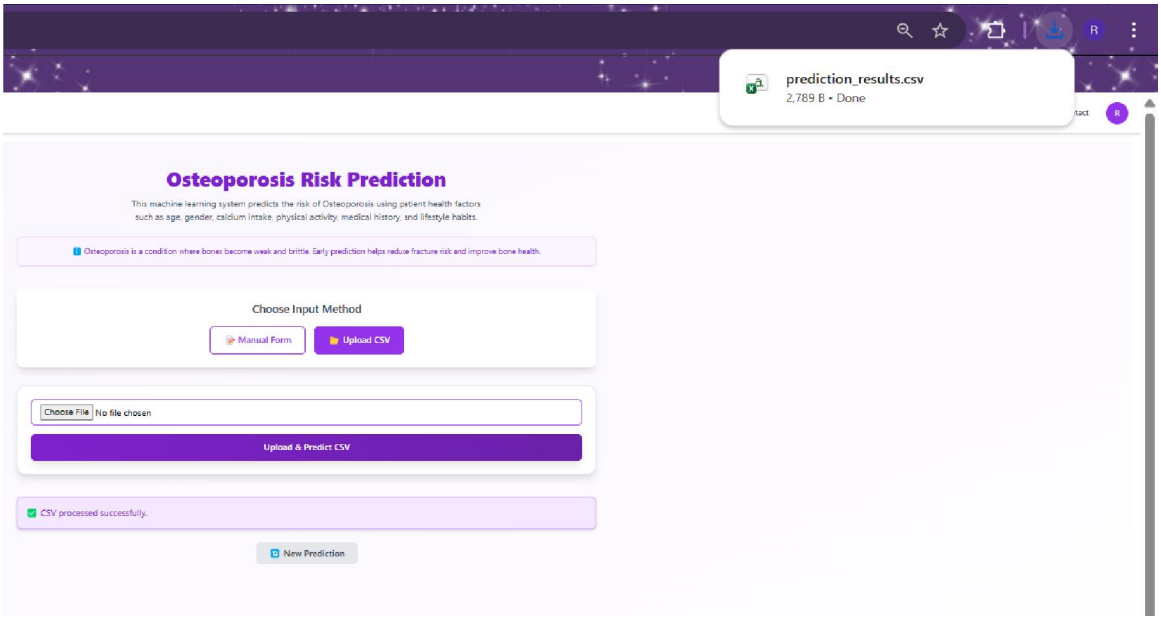

Fig. 7. CSV-Based Prediction Output

VII. CONCLUSION

The proposed Osteoporosis Prediction System demonstrates the effective use of machine learning techniques for early detection of osteoporosis risk. The system successfully integrates data preprocessing, model training, and web-based deployment to provide a complete end-to-end solution for healthcare prediction. By utilizing clinical and lifestyle parameters such as age, gender, BMI, calcium intake, vitamin D levels, and medical history, the system is capable of identifying individuals who may be at risk of developing osteoporosis. The use of the XGBoost classifier resulted in high prediction accuracy compared to other models, making it suitable for reliable risk assessment. The system supports both manual data input and bulk CSVbased prediction, allowing it to handle individual as well as large-scale data efficiently. Additionally, the inclusion of a recommendation module enhances usability by providing personalized health suggestions based on prediction results. Overall, the developed system offers a cost-effective, accessible, and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

user-friendly solution for early osteoporosis risk detection. It has the potential to assist both individuals and healthcare professionals in making informed decisions and adopting preventive measures to improve bone health.

Future enhancements, such as integration with real-time health monitoring systems, mobile applications, and advanced machine learning models, can further improve the system's performance and practical applicability.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the project supervisor and faculty members for their valuable guidance, continuous support, and constructive suggestions during the development of the Osteoporosis Prediction System. Their insights and encouragement greatly contributed to the successful completion of this research work. The authors also acknowledge the support provided by the institution for offering the necessary academic resources and infrastructure required for this study. Furthermore, the authors appreciate the open-source communities and platforms such as React, Flask, MongoDB, and various machine learning libraries including pandas, numpy, scikit-learn, and xgboost, which enabled the development and implementation of the proposed system.Finally, the authors would like to thank all individuals who directly or indirectly contributed to the successful completion of this project.

REFERENCES

[1] J.-B. Tu et al., "Using machine learning techniques to predict the risk of osteoporosis," Scientific Reports, 2024.
[2] A. M. Sarhan et al., "Knee osteoporosis diagnosis based on deep learning," International Journal of Computational Intelligence Systems, 2024.
[3] Y. Pan et al., "Osteoporosis screening using low-dose CT," European Radiology, 2020.
[4] H. M. Bui et al., "Osteoporosis prediction using ML in Vietnamese women," Scientific Reports, 2022.
[5] S. Tabib et al., "Diagnosis of osteoporosis risk using ML," Endocrine and Metabolic Disorders, 2025.
[6] T. Tanphiriyakun et al., "BMD response prediction using ML," Scientific Reports, 2021.
[7] L. Ji et al., "Osteoporosis and survival in breast cancer using ML," Frontiers in Oncology, 2022.
[8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
[9] World Health Organization, "Assessment of osteoporosis at the primary health care level," WHO Press, 2018.
[10] J. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, 2001.
[11] I. Goodfellow et al., Deep Learning. MIT Press, 2016.
[12] L. Breiman, "Random forests," Machine Learning, 2001.
[13] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, 2011.
[14] Kaggle, "Osteoporosis Dataset." [Online]. Available: https://www.kaggle.com. [Accessed: 2026].
[15] J. Han and M. Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2012.