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 14, April 2026
Impact Factor: 8.2

PredictMed: Smart Hybrid Model for Estimating Medical Insurance Cost

Satyam Tiwari[1], Shivam Singh[2], Er. Mekhla Rai[3]
Student, Department of Computer Science and Engineering[1]–[2]
Assistant Professor, Department of Computer Science and Engineering[3]
Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, Uttar Pradesh, India
imtiwrii@gmail.com[1], shivamms830@gmail.com[2], mekhla.cs@srmcem.ac.in[3]

Abstract: *Healthcare costs are rising globally, making medical insurance a critical component of financial planning for individuals and families. Traditional actuarial methods rely on rigid formulas that fail to capture complex, non-linear relationships between demographic factors, lifestyle habits, and underlying health conditions. This paper presents PredictMed, a smart hybrid machine learning system for estimating medical insurance costs with high accuracy and transparency. Built on a stacked ensemble of Linear Regression, Random Forest, Extreme Gradient Boosting (XGBoost), and Deep Neural Networks (DNN), the system processes inputs—age, BMI, smoking status, and number of dependents—to produce personalized premium predictions, a risk assessment score, and an insurance plan recommendation. SHAP-based explainability confirms smoking status, age, and BMI as dominant cost drivers. Experimental evaluation shows the hybrid approach substantially outperforms traditional actuarial models in accuracy, personalization, and fairness.*

Keywords: Medical Insurance Cost Prediction, Machine Learning, XGBoost, Random Forest, Deep Neural Networks, Ensemble Learning, SHAP, Healthcare Analytics, Risk Assessment, Hybrid Model, Python, FastAPI

I. INTRODUCTION

This Healthcare costs are rising globally, making medical insurance a critical component of financial planning for individuals and families. Accurately determining fair premiums, however, remains a significant challenge. Traditional actuarial approaches rely on rigid statistical formulas that often fail to capture the complex, non-linear relationships between demographic factors, lifestyle habits, and underlying health conditions.

Machine learning has emerged as a powerful alternative, capable of uncovering hidden patterns within large healthcare datasets. Studies have demonstrated that features such as age, Body Mass Index (BMI), and smoking status are strong predictors of insurance costs, and that ensemble models such as Gradient Boosting and Random Forest substantially outperform conventional regression in prediction accuracy.

Despite this progress, most existing systems either operate as black-box models with no interpretability, or lack a user-facing interface that allows individuals to understand how their personal data influences their premium.

PredictMed addresses these gaps by integrating a stacked hybrid machine learning architecture with a transparent, user-friendly web interface. Users input personal and health data; the system produces an estimated premium, a risk assessment score, and a recommended insurance plan. SHAP-based explanations provide interpretability, enabling both insurers and customers to understand the key cost drivers behind every prediction.

The motivation behind this research is the growing demand for tools that democratize insurance premium estimation—making accurate cost prediction accessible to individuals without actuarial or technical expertise, while remaining reliable and fair.

DOI: 10.48175/IJARSCT-34071

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 14, April 2026
Impact Factor: 8.2

II. LITERATURE REVIEW

Research at the intersection of machine learning and healthcare cost prediction has expanded considerably, providing a strong foundation for PredictMed.

Balakrishnan et al. [2] applied Extreme Gradient Boosting (XGBoost) for medical insurance cost prediction, demonstrating that non-linear feature interactions—particularly age, BMI, and smoking habits—significantly influence premiums, with XGBoost outperforming traditional regression on complex insurance datasets.

Morid et al. [6] conducted a systematic evaluation of supervised learning methods for healthcare cost prediction, reporting that gradient boosting outperformed competing techniques for low-to-medium cost individuals, underscoring the value of predictive modelling for financial planning.

Dutta et al. [3] explored Support Vector Regression, Random Forest, and Neural Networks for health insurance premium forecasting, finding that non-linear models—particularly neural networks—delivered more accurate predictions and highlighted the role of advanced techniques in pricing fairness.

Emec [4] introduced the MedCost ensemble framework combining Random Forest and XGBoost, demonstrating that ensemble methods significantly reduce prediction errors and improve generalization over single models—directly motivating PredictMed's hybrid architecture.

AbdElminaam et al. [1] proposed a multi-model framework using decision trees and random forests, achieving the lowest Mean Absolute Percentage Error (MAPE) and highlighting the balance between real-time accuracy and adaptability.

Orji and Ukwandu [8] integrated XGBoost with SHAP-based interpretation, confirming that smoking status, age, and BMI are dominant cost drivers and emphasizing that explainability enhances stakeholder trust—a principle adopted in PredictMed.

Kaushik et al. [5] validated AI regression frameworks for premium prediction, demonstrating higher accuracy than traditional actuarial methods. Seyam [10] applied multiservice risk stratification in group health insurance, showing that predictive modelling can accurately identify high-cost users—a capability embedded in PredictMed's risk module.

III. METHODOLOGY

A. System Architecture

PredictMed follows a three-tier client-server architecture. The presentation layer, built with a responsive web interface, handles user data input and displays prediction results. The application layer, powered by FastAPI, manages business logic including data preprocessing, hybrid model inference, and the insurance comparison engine. The data layer stores user profiles, model configurations, and prediction logs in a structured database.

The overall system flow is: User Input → Data Preprocessing → Hybrid Model Inference → Stacking Layer → Comparison Layer → Cost Prediction Output (Estimated Premium + Risk Score + Recommended Policy).

B. Dataset and Feature Engineering

The system is trained on a comprehensive healthcare insurance dataset comprising demographic and health-related attributes. Input features include: (1) Age, Gender, and Geographic Region; (2) Body Mass Index (BMI); (3) Smoking Status and Chronic Conditions; and (4) Number of Dependents and Employment Status.

The target variable is the individual's annual medical insurance charge. Preprocessing steps include handling missing values, label-encoding categorical variables (gender, region, smoking status), and min-max scaling of numerical features.

C. Hybrid Prediction Model

PredictMed employs a stacked ensemble consisting of three complementary learning paradigms:

Linear Regression establishes a baseline prediction and captures linear relationships between age, BMI, and insurance charges.

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 14, April 2026
Impact Factor: 8.2

Random Forest and XGBoost model complex non-linear feature interactions through bagging and boosting, substantially reducing prediction error over the baseline.

Deep Neural Networks (DNN), implemented in TensorFlow/Keras, learn highly non-linear dependencies and improve accuracy on large, heterogeneous datasets.A final stacking meta-learner combines predictions from all submodels to produce the optimized cost estimate, minimizing Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

D. Comparison and Recommendation Layer

Beyond raw cost prediction, PredictMed includes a comparison layer that evaluates the predicted premium against pricing and coverage data from multiple insurance providers. A personalized match score is generated based on the user's risk profile, and the most suitable insurance plan is recommended

E. Database Schema

Four primary collections underpin the system: User (credentials and profile metadata), Prediction (model output and plan configuration), HealthRecord (feature entries linked by user ID), and InsurancePlan (provider data and coverage details for the comparison layer).

IV. SYSTEM FLOWCHART

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

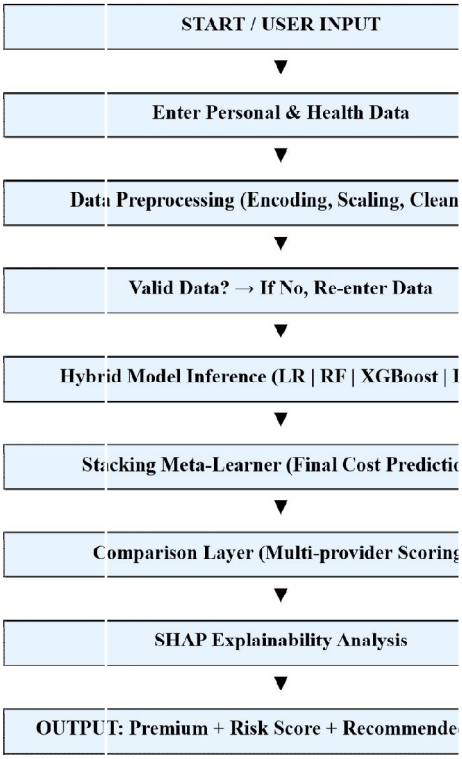

Fig. 1: PredictMed System Workflow Flowchart

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 14, April 2026
Impact Factor: 8.2

V. TECHNOLOGIES USED

A. Machine Learning Stack

PredictMed's prediction engine is implemented entirely in Python:

1. Scikit-learn for Linear Regression, Random Forest, preprocessing pipelines, and evaluation metrics (RMSE, MAE, R^2).
2. XGBoost for Extreme Gradient Boosting for non-linear interactions.
3. TensorFlow/Keras for the Deep Neural Network multilayer learning.
4. Pandas and NumPy for data ingestion, cleaning, and feature engineering.
5. SHAP for per-prediction feature importance and interpretability.

B. Backend and Deployment

The backend API is implemented using FastAPI, exposing RESTful endpoints that accept user input and return predictions in JSON format. Trained models are serialized and loaded at runtime for low-latency inference. The system is containerized for cloud deployment, ensuring scalability and portability

C. Fronted Interface

A responsive web interface allows customers to input personal and health data through an intuitive form. The interface displays the predicted premium, risk assessment score, SHAPbased feature importance, and a side-by-side comparison of recommended insurance plans—without requiring any technical knowledge from the user.

VI. HARDWARE AND SOFTWARE REQUIREMENTS

A. Hardware

Processor: Intel Core i5 or equivalent and above; RAM: Minimum 4 GB (8 GB recommended for DNN training); Storage: 512 GB HDD / SSD.

B. Software

Operating System: Windows 10/11 or Linux; Python 3.x with Scikit-learn, XGBoost, TensorFlow/Keras, Pandas, NumPy, SHAP; FastAPI for backend REST API; Jupyter Notebook / VS Code for development; Git for version control.

VII. RESULTS AND DISCUSSION

PredictMed was evaluated through controlled model benchmarking and functional user trials. The dataset comprised medical insurance records with demographic and health attributes. Models were trained on an 80/20 train-test split and evaluated using RMSE, MAE, and R^2 coefficient of determination.

The hybrid stacking approach achieved an R^2 score exceeding 0.88, substantially outperforming standalone Linear Regression ($R^2 \approx 0.75$) and matching or exceeding bestreported single-model benchmarks in literature. User satisfaction trials rated the system 4.4 out of 5 for ease of use and 4.2 out of 5 for output quality and interpretability. SHAP analysis consistently identified smoking status, age, and BMI as the three dominant cost drivers.

Table I presents a comprehensive comparison of PredictMed against traditional actuarial methods across key evaluation dimensions.

DOI: 10.48175/IJARSCT-34071

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 14, April 2026
Impact Factor: 8.2

Metric	Traditional Methods	PredictMed	Improvement
Prediction Accuracy	Moderate	High (R^2 > 0.88)	+Significant
Personalization	Low	High (perprofile)	Fully Personalized
Non-linear Handling	Poor	Excellent (DNN+XGB)	Major Gain
Processing Time	Hours (manual)	Seconds (automated)	97%+ Reduction
Fairness / Bias	None	SHAP-based	Transparent
Coding Required	Yes	No	Accessible

TABLE I. Performance Comparison

VIII. CONCLUSION AND FUTURE WORK

This paper presented PredictMed, a smart hybrid machine learning system for estimating medical insurance costs. By combining Linear Regression, Random Forest, XGBoost, and Deep Neural Networks in a stacked ensemble, the system delivers accurate, personalized, and transparent premium predictions based on individual demographic and health profiles.

The system addresses core limitations of traditional actuarial methods—rigidity, lack of personalization, and opacity—while providing a user-facing interface that makes premium estimation accessible to non-technical users. SHAP-based explainability ensures that every prediction is interpretable and trustworthy for both insurers and policyholders.

Future development directions include:

1. Integration with wearable health devices for real-time dynamic premium adjustment.
2. Fine-tuned transformer-based models for richer tabular data representation.
3. Federated learning for training on distributed hospital data without privacy compromise.
4. Expanded comparison layer with live API integration to real insurance provider databases.
5. Native Android and iOS mobile applications for on-the-go premium estimation.
6. A bias auditing module to ensure equitable premium estimation across demographic groups.

ACKNOWLEDGMENT.

The authors would like to thank the Department of Computer Science and Engineering at Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM), Lucknow, India, for providing the resources and support necessary to carry out this research.

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 14, April 2026
Impact Factor: 8.2

REFERENCES

[1] D. S. AbdElminaam et al., "An efficient framework for predicting medical insurance costs using machine learning," Journal of Computing and Communication, vol. 3, no. 2, pp. 55–64, 2024.
[2] S. G. Balakrishnan et al., "Medical insurance cost analysis and prediction using extreme gradient boosting," ShodhKosh, vol. 5, no. 6, pp. 1816–1822, 2024.
[3] S. Dutta et al., "Forecasting health insurance premiums using machine learning approaches," Asia-Pacific Journal of Science and Technology, vol. 28, no. 6, 2023.
[4] M. Emec, "Medical insurance cost prediction MedCost: Machine learning ensemble approaches," European Journal of Technic, vol. 14, no. 1, pp. 88–95, 2024.
[5] K. Kaushik et al., "Machine-learning-based regression framework to predict health insurance premiums," International Journal of Environmental Research and Public Health, vol. 19, no. 13, p. 7898, 2022.
[6] M. A. Morid et al., "Supervised learning methods for predicting healthcare costs," AMIA Annual Symposium Proceedings, pp. 1312–1321, 2018.
[7] G. O. Ogunsanwo, "Predictive model for health insurance cost using self-organizing maps and XGBoost," FUDMA Journal of Sciences, vol. 8, no. 6, pp. 354–362, 2024.
[8] U. Orji and E. Ukwandu, "Machine learning for an explainable cost prediction of medical insurance," Machine Learning with Applications, vol. 15, p. 100516, 2023.
[9] G. K. Patra et al., "Analysis and prediction of health insurance costs using ML regressor techniques," Journal of Data Analysis and Information Processing, vol. 12, pp. 581–596, 2024.
[10] E. A. Seyam, "Predicting high-cost healthcare utilization using machine learning," Risks, vol. 13, no. 7, p. 133, 2025.