IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141249

Optimized Ensemble Regression with Explainable AI for Interpretable Healthcare Cost Prediction

Md. Shahidur Rahman Saklain[1], Antar Sarker[2], Md. Sadiq Iqbal[3]

Student, Department of Computer Science and Engineering, Bangladesh University, Dhaka, Bangladesh[1]

Lecturer, Co-Supervisor-1, Department of CSE, Bangladesh University, Dhaka, Bangladesh[2]

Professor, Supervisor, Department of Computer Science and Engineering, Bangladesh University, Dhaka, Bangladesh[3]

Abstract: Accurate prediction of healthcare insurance costs plays a crucial role in improving cost management, policy design, and healthcare planning. This study investigates the effectiveness of various machine learning (ML) algorithms in forecasting healthcare insurance expenditures and identifies the most suitable model for reliable cost estimation. A publicly available dataset containing demographic and lifestyle-related attributes such as age, sex, body mass index (BMI), number of children, smoking status, and region was utilized. Multiple regression-based ML models, including Linear Regression (LR), Support Vector Regression (SVR), Random Forest Regressor (RFR), XGBoost Regressor (XGBR), LightGBM (LGBM), and Gradient Boosted Regression (GBR), were implemented and compared. The evaluation results demonstrate that the GBR model outperformed other approaches by achieving the lowest mean squared error (MSE = 18,153,562.14) and mean absolute error (MAE = 2,270.97), along with the highest coefficient of determination (R^2 = 0.87), peak signal-to-noise ratio (PSNR = 22.97), and signal-to-noise ratio (SNR = 9.97). Cross-validation further confirmed its robustness, with the tenth fold achieving an R^2 of 0.91. To enhance model interpretability, explainable artificial intelligence (XAI) tools such as SHAP and LIME were applied to the final GBR model, revealing that “region” and “smoker” were the most influential factors affecting insurance costs. The findings confirm that GBR, combined with explainable AI techniques, offers a robust, transparent, and reliable solution for predicting healthcare insurance costs. Future work will focus on integrating more advanced explainable frameworks and real-world healthcare datasets to further improve reliability and applicability.

Keywords: Healthcare insurance cost prediction; machine learning; explainable artificial intelligence (XAI); regression models; gradient boosting

I. INTRODUCTION

The accurate prediction of healthcare insurance costs has become a critical component in advancing financial planning, resource allocation, and policy design within modern health systems. As healthcare expenditures continue to rise globally, insurers, providers, and policymakers face increasing pressure to anticipate cost burdens and mitigate financial risk. For individual policy-holders, precise cost estimations enable more informed choices in plan selection and contribute to improved outcomes and affordability.

Healthcare insurance premiums reflect a multitude of interdependent factors: demographic attributes (such as age and gender), health-related metrics (such as body mass index and comorbidities), behavioral indicators (such as smoking status), and regional or socio-economic variations. Traditional statistical approaches, for instance, multiple linear regression or generalized linear models, offer interpretability but often fall short of capturing complex non-linear relationships and interactions among these features [1][2]. With the rapid proliferation of data from electronic health records, claims databases, and wearable devices, the adoption of machine learning (ML) methods presents a compelling opportunity to improve prediction accuracy in this domain [3].

Supervised machine learning models, including ensemble algorithms such as Random Forest, Gradient Boosting, and XGBoost, have demonstrated superior performance in forecasting individual medical expenditures when compared to conventional models [4][5]. However, the “black-box” nature of many high-performing models remains a concern, especially in regulated and sensitive environments like healthcare. To build trust and facilitate adoption, Explainable Artificial Intelligence (XAI) techniques (for example, SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME)) are increasingly used to interpret model outputs and elucidate key driving features [6].

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ∺ Peer-reviewed & Refereed journal ∺ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141249

This study aims to evaluate a suite of supervised machine learning algorithms for forecasting healthcare insurance costs and to identify the most effective approach in terms of predictive performance and interpretability. Specifically, contributions of this work are threefold:

1. It investigates the effectiveness of computational intelligence techniques for predicting healthcare insurance costs using demographic, behavioural and regional data.
2. It provides a comparative assessment of multiple regression-based models, including linear, support vector, and ensemble architectures.
3. It integrates explainable AI methods to enhance transparency of prediction results and to highlight the most influential cost-drivers.

The remainder of this paper is structured as follows: Section II reviews related work in healthcare cost prediction and interpretable ML. Section III presents the methodology, including dataset description, model development, and explainability framework. Section IV discusses experimental results and interpretability insights. Finally, Section V highlights conclusions and future research directions.

II. LITERATURE REVIEWS

Early studies on healthcare insurance cost prediction primarily relied on traditional statistical methods such as Linear Regression (LR) and Generalized Linear Models (GLM). These models were effective for capturing linear relationships but struggled with complex and nonlinear dependencies among factors such as age, lifestyle, and medical history, which significantly influence healthcare costs [7]. To address these limitations, researchers began adopting ML techniques, including Support Vector Regression (SVR), Decision Trees, Random Forests, Gradient Boosting, and XGBoost, which demonstrated superior predictive accuracy over classical statistical models [8].

Subsequent research explored ensemble and hybrid approaches that combine multiple algorithms to enhance prediction accuracy. For example, Albalawi et al. [9] and Zou et al. [10] reported that hybrid ML frameworks outperformed single-model approaches in forecasting medical costs. With the increasing availability of healthcare data, deep learning (DL) models were also introduced for cost prediction tasks [11]. These models improved accuracy by capturing highly nonlinear patterns; however, their lack of interpretability remains a major challenge in the healthcare domain.

To overcome this limitation, recent studies have incorporated XAI techniques to enhance transparency and trust in model decisions. Orji and Ukwandu [12] and Bhongade et al. [13] employed ensemble learning methods integrated with explainability tools such as SHAP and ICE (Individual Conditional Expectation) to identify key factors influencing insurance costs. Similarly, Goel and Chaudhary [14] proposed interpretable ML approaches for insurance pricing, while Kaushik et al. [15] developed regression-based frameworks emphasizing model transparency and interpretability.

Although these studies demonstrate substantial improvements in prediction accuracy, a noticeable research gap remains. Most prior work prioritizes performance metrics, with comparatively less emphasis on explainability and validation using real-world datasets. Addressing this gap, the present study investigates multiple supervised regression models and integrates XAI tools (SHAP and LIME) to achieve both high accuracy and model interpretability, ensuring reliable and transparent healthcare insurance cost predictions.

III. METHODOLOGY

This study aims to minimize prediction errors in estimating healthcare insurance costs through the application of ML techniques. The overall research framework includes data collection, preprocessing, exploratory data analysis, model construction, evaluation, and explainability integration. The methodology followed in this research is illustrated in Fig. 1.

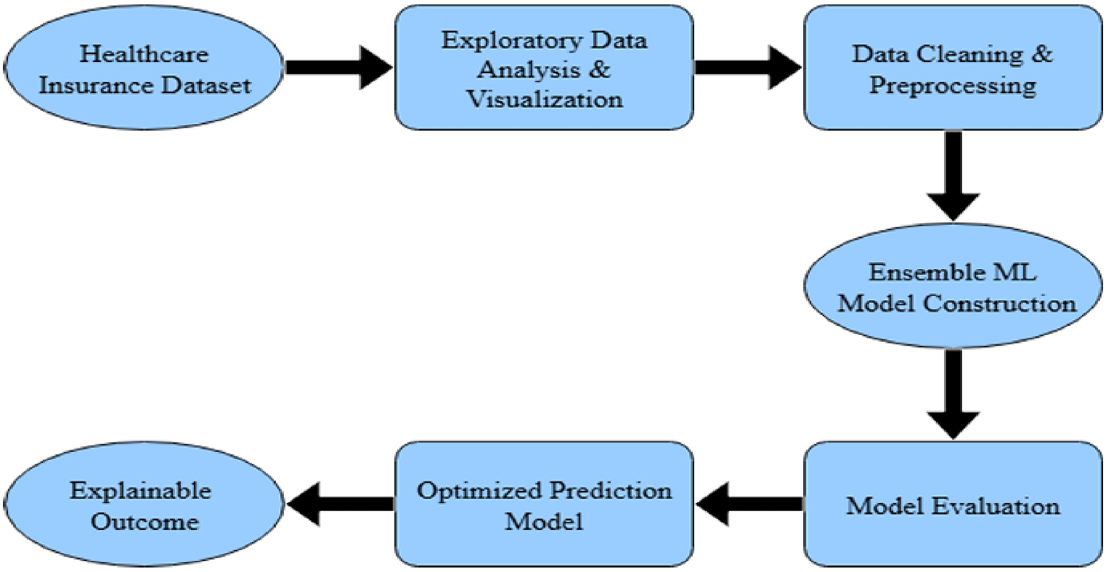

Fig. 1. Architecture of the research.

A. Healthcare Insurance Dataset

The study utilized a publicly available healthcare insurance dataset obtained from Kaggle [16]. The dataset contains 1,338 records with six independent features and one target variable (charges). The input features include Age, Sex, Body Mass Index (BMI), Number of Children, Smoker status, and Region. The output variable represents the medical charges covered by healthcare insurance. A detailed description of the dataset features is presented in Table I.

TABLE I. DESCRIPTIVE FEATURES OF THE DATASET.

Feature	Description	Variable Type
Age	Age of person	Numerical
Sex	Gender of person	Nominal
BMI	Body Mass Index	Numerical
Children	Number of children	Numerical
Smoker	Indicator of whether the person smokes or not	Nominal
Region	Area of residence	Nominal
Charges	Medical costs paid by healthcare insurance	Numerical

B. Exploratory Data Analysis (EDA)

In this research, several visualization techniques, such as histograms, and correlation heat maps [17-19], were applied during the preprocessing stage of the dataset. These visual tools help in understanding the structure of the data, identifying patterns, and guiding the development of accurate models. Histograms are especially helpful because they show how frequently different values appear in the dataset. By observing the overall shape and spread of the data, it becomes easier to notice trends or irregularities. In this study, histograms were used to visualize the distribution of key features such as age, sex, BMI, number of children, smoker status, region, and medical charges. These visualizations help identify the most common value ranges, the presence of variability, and any unusual or extreme points. Understanding these patterns supports better decision-making when building prediction models and analyzing health-related factors. Fig. 2 presents these distributions clearly, allowing easier interpretation of how each feature behaves within the dataset.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁂ Peer-reviewed & Refereed journal ⁂ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141249

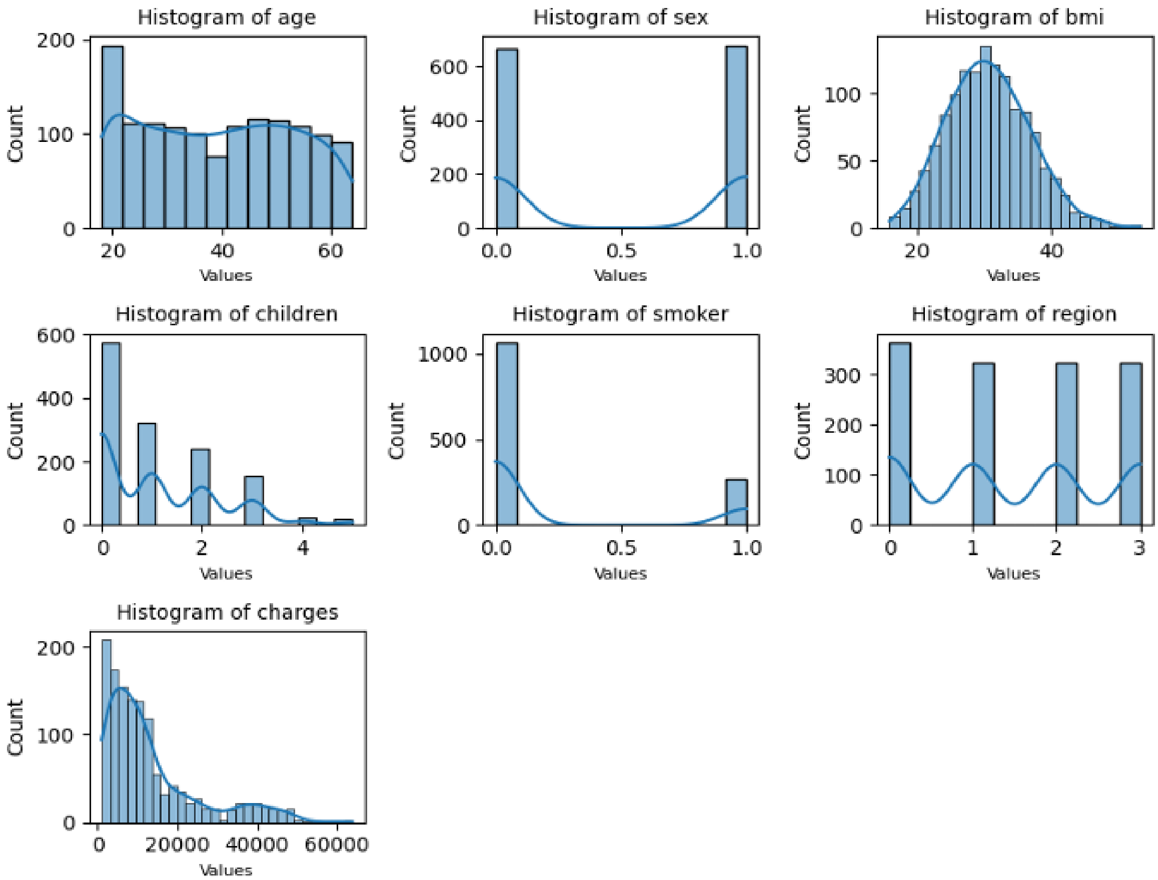

Fig. 2. Histogram illustrating the distribution of different features.

A heatmap is a visual tool that uses color gradients to represent the strength of relationships between different features. Darker red shades indicate a strong positive correlation, while darker blue shades represent a strong negative correlation. This type of visualization helps in quickly identifying which variables are closely connected. Fig. 3 presents a correlation matrix heatmap used to examine the relationships among the dataset's features. From this heatmap, it is evident that smokers and charges have a strong positive correlation, suggesting that smokers generally incur higher medical charges. Overall, heatmaps make it easier to observe such patterns, supporting more informed analysis, decision-making, and predictive modeling.

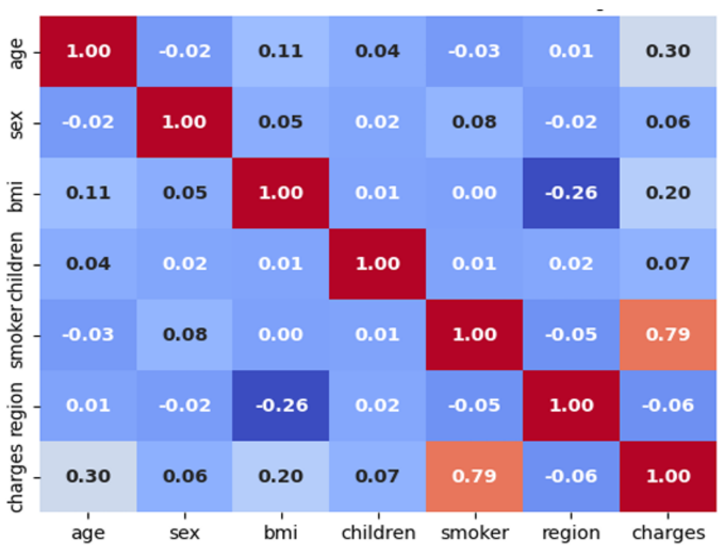

Fig. 3. Correlation matrix with a heat map.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ∺ Peer-reviewed & Refereed journal ∺ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141249

C. Data Cleaning and Preprocessing
The dataset was examined for missing or inconsistent values; none were found. Consequently, no imputation techniques were required. The data was then normalized and encoded, where necessary, to make it suitable for regression-based ML models.

D. Ensemble ML Model Construction
In this study, several well-known ensemble regression techniques, such as SVR [20], RFR [21], XGBOOSTR [22], LGBM, and GBR [23], were applied to predict healthcare insurance costs. Among these methods, the evaluation results showed that GBR performed the best.
GBR is an ensemble learning approach that builds a series of decision trees, where each new tree is trained to correct the errors made by the previous ones. Instead of modeling the target variable directly, GBR fits each new model to the residuals, which represent the difference between the actual and predicted values from earlier iterations. This stage-wise process continues until the chosen loss function is minimized, allowing the model to improve its accuracy by gradually adding weak learners.

E. Model Evaluation Metrics
To assess model performance, five statistical metrics were applied:
- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- Coefficient of Determination (R^2)
- Peak Signal-to-Noise Ratio (PSNR)
- Signal-to-Noise Ratio (SNR)

Lower MSE and MAE values, combined with higher R^2, PSNR, and SNR values, indicate better model performance [24].

F. Ensemble ML Model Construction
All implemented models were compared based on the above metrics. The GBR demonstrated the lowest MSE and MAE, and the highest R^2, PSNR, and SNR values, confirming its superiority. Consequently, the GBR model was selected as the final optimized model for predicting healthcare insurance costs.

G. Explainable Outcome
To ensure model transparency, XAI techniques were integrated. Two widely adopted methods, SHAP and LIME, were used to interpret model behavior. The combination of SHAP and LIME enhanced interpretability, enabling better insight into the influence of key features such as Region, Smoker status, and BMI on insurance cost predictions [25-26].

IV. RESULTS AND DISCUSSION

Table II presents the overall performance of all evaluated models. The results indicate that the GBR model achieves the lowest MSE, MAE, and R^2 values of 18,153,562.14, 2,270.97, and 0.87, respectively. Additionally, GBR attains the highest PSNR and SNR values, recorded at 22.97 and 9.97. These outcomes demonstrate that GBR is particularly well-suited for healthcare insurance cost prediction, offering superior accuracy and robustness compared to the other models.

TABLE II MODEL PERFORMANCE METRICS.

Model	MSE	MAE	R^2	PSNR	SNR
SVR	166559544.67	8281.98	0.115	13.35	0.35
RFR	21921469.12	2600.28	0.85	22.16	9.15
XGBR	25842087.81	2988.23	0.83	21.44	8.44
LGBM	20541778.25	2675.68	0.86	22.44	9.43
GBR	18153562.14	2270.97	0.87	22.97	9.97

Table III summarizes the model's performance across the different cross-validation folds. Among these, Fold 10 demonstrates the strongest results, producing the highest R^2 value of 0.91, indicating that the model explains most of the variance in the data. It also yields low error values, with an MSE of 18,175,561.53 and an MAE of 2,512.68, along with high PSNR (23.33) and SNR (9.88). These results suggest that Fold 10 provides the most accurate and reliable predictions. In contrast, Fold 9 exhibits the weakest performance. It records the lowest R^2 value of 0.82 and the highest error values, with an MSE of 27,175,039.92 and an MAE of 2,670.29, as well as the lowest PSNR (19.19) and SNR (8.07). This indicates that Fold 9's predictions are less accurate compared to the other folds. Overall, this fold-wise evaluation highlights how model accuracy can vary across different cross-validation splits. Such analysis provides

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ∺ Peer-reviewed & Refereed journal ∺ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141249

valuable insights into the model's consistency, helping to identify both strengths and limitations and supporting the development of a more stable and trustworthy prediction system.

TABLE III 10-FOLD CROSS-VALIDATION RESULTS (GBR)

Fold	MSE	MAE	R^2	PSNR	SNR
Fold 1	21686913.98	2511.16	0.86	22.20	9.98
Fold 2	13686906.15	2013.65	0.90	22.54	10.34
Fold 3	22045265.07	2527.93	0.83	20.75	8.78
Fold 4	24142731.44	2667.35	0.84	21.00	8.79
Fold 5	24142731.44	2121.46	0.87	21.69	9.31
Fold 6	22188381.95	2663.76	0.84	20.05	9.38
Fold 7	22466326.72	2669.75	0.84	22.57	9.25
Fold 8	18390394.33	2385.53	0.86	21.14	10.03
Fold 9	27175039.92	2670.29	0.82	19.19	8.07
Fold 10	18175561.53	2512.68	0.91	23.33	9.88

Fig. 4 shows the relationship between the actual and predicted values generated by the healthcare insurance cost prediction model. The blue points represent the model's predicted outputs, while the solid red line indicates the ideal case where predictions perfectly align with the actual values. The close clustering of the blue points around the red line demonstrates that the GBR model achieves strong predictive accuracy.

Fig. 5 presents the SHAP summary plot for the final GBR model used in predicting healthcare insurance costs. The plot illustrates the contribution of each feature to the model's output, with each point representing an individual prediction. The results indicate that region is the most influential feature, demonstrating a broad range of impact on the model's predictions. The smoker attribute also shows a strong positive SHAP value for higher feature values, suggesting that smoking status substantially increases predicted insurance costs. Other features, including sex, children, and BMI, exhibit smaller yet noticeable effects. This visualization provides clear insights into how key variables influence model outcomes, thereby enhancing the interpretability and overall understanding of the model's behavior.

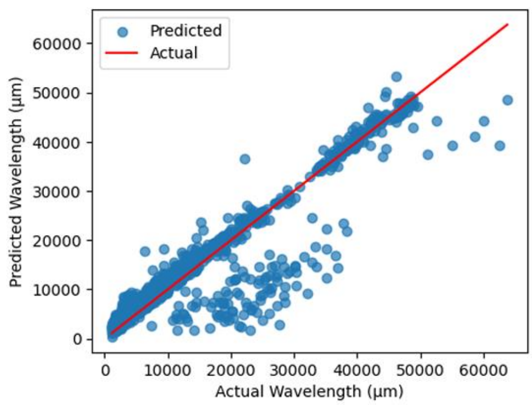

Fig. 4. Predicted value vs Actual value.

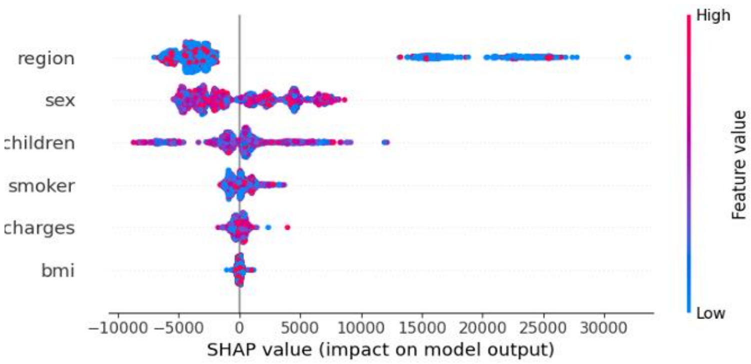

Fig. 5. The impact of various features on the model's output.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141249

Fig. 6 presents the LIME visualization for the final GBR model. The plot indicates that all features exhibit negative contributions to the predicted healthcare insurance costs, suggesting that higher values of these attributes are associated with lower cost estimates. This representation enhances model interpretability by clarifying both the direction and magnitude of each feature's influence. As a result, it supports greater transparency and offers a clearer understanding of how individual variables affect the model's cost predictions.

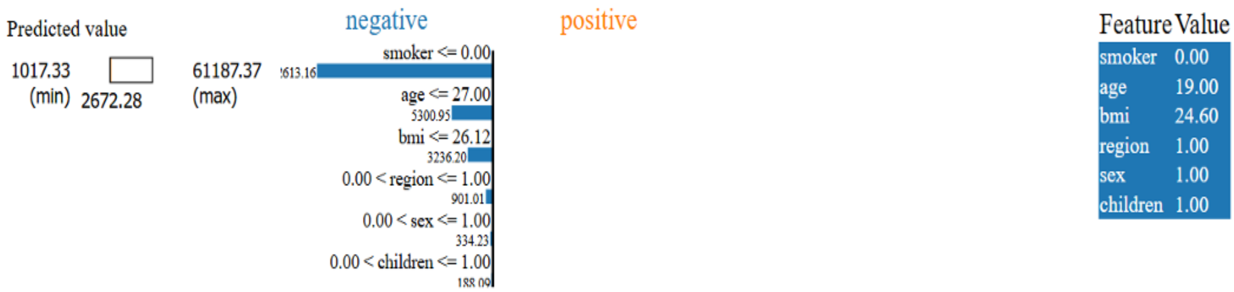

Fig. 6. LIME analysis with key features.

Accurate prediction of healthcare insurance costs is crucial for effective cost management and planning. In this study, we employed the GBR to predict future healthcare expenses. The results demonstrate that GBR outperforms other models across all evaluated performance metrics. This superior performance is attributed to two key strengths of GBR. First, it can automatically identify and leverage the most important features without requiring manual selection. Second, it effectively captures complex, non-linear interactions among variables, which simpler models often fail to address. Overall, GBR provides more accurate and reliable predictions, making it a robust tool for healthcare insurance cost forecasting and supporting informed decision-making in healthcare systems.

V. CONCLUSION

Accurate cost prediction is essential in the health insurance sector because it reduces uncertainty and supports better decision-making. Machine learning models, especially GBR, are highly effective for handling large datasets and capturing complex variable relationships that traditional methods often overlook. In this study, the proposed GBR model was trained to predict healthcare insurance costs. The model was evaluated using several key performance metrics, including MSE, MAE, R2, PSNR, and SNR. GBR consistently outperformed the other baseline models, demonstrating strong predictive accuracy. These results highlight the capability of GBR to deliver reliable and transparent predictions, which can help both insurers and policyholders save time and reduce costs. By automating complex analytical tasks, GBR can aid healthcare providers and insurance companies in resource planning, premium calculation, and risk management. In the future, we aim to integrate additional XAI techniques, further enhance the ML models, and apply them in real-time clinical settings. These improvements can support better cost optimization and contribute to a more efficient, data-driven, and patient-centered healthcare ecosystem.

REFERENCES

[1]. ul Hassan, C., Iqbal, J., Hussain, S., AlSalman, H., Mosleh, M. A. A., & Ullah, S. S. (2021). A Computational Intelligence Approach for Predicting Medical Insurance Cost. Mathematical Problems in Engineering, 2021, Article ID 1162553.

[2]. Orji, U., & Ukwandu, E. (2023). Machine Learning For An Explainable Cost Prediction of Medical Insurance. Machine Learning with Applications, 15, 100516.

[3]. Kshirsagar, R., Hsu, L-Y., Chaturvedi, V., Greenberg, C. H., McClelland, M., Mohan, A., … & Lee, A. (2020). Accurate and Interpretable Machine Learning for Transparent Pricing of Health Insurance Plans. arXiv preprint arXiv:2009.10990.

[4]. Morid, M. A., Liu Sheng, O., Kawamoto, K., Ault, T., Dorius, J., & Abdelrahman, S. (2020). Healthcare Cost Prediction: Leveraging Fine-grain Temporal Patterns. arXiv preprint arXiv:2009.06780.

[5]. Bau, Y-T., & Hanif, S. A. M. (2024). Comparative Analysis of Machine Learning Algorithms for Health Insurance Pricing. International Journal on Informatics Visualization, 8(1).

[6]. Hossain, M. M., Mamun, M., Munir, A., Rahman, M. M., & Chowdhury, S. H. (2025). A Secure Bank Loan Prediction System by Bridging Differential Privacy and Explainable Machine Learning. Electronics, 14(8), 1691.

DOI: 10.17148/IJARCCE.2025.141249

[7]. A. Vimont, H. Leleu, and I. Durand-Zaleski, "Machine learning versus regression modelling in predicting individual healthcare costs from a representative sample of the nationwide claims database in France," Eur. J. Health Econ., pp. 1–13, 2021, doi: 10.1007/s10198-021-01363-4.
[8]. C. A. ul Hassan, J. Iqbal, S. Hussain, H. AlSalman, M. A. A. Mosleh, and S. S. Ullah, "A Computational Intelligence Approach for Predicting Medical Insurance Cost," Math. Probl. Eng., vol. 2021, p. 1162553, 2021, doi: 10.1155/2021/1162553.
[9]. S. Albalawi, L. Alshahrani, N. Albalawi, R. Alharbi, and A. Alhakamy, "Prediction of healthcare insurance costs," Computers and Informatics, vol. 3, no. 1, pp. 9–18, 2023.
[10]. S. Zou, C. Chu, N. Shen, and J. Ren, "Healthcare Cost Prediction Based on Hybrid Machine Learning Algorithms," Mathematics, vol. 11, no. 23, p. 4778, 2023, doi: 10.3390/math11234778.
[11]. X. Zeng, S. Lin, and C. Liu, "Multi-View Deep Learning Framework for Predicting Patient Expenditure in Healthcare," IEEE Open J. Comput. Soc., vol. 2, pp. 62–71, 2021, doi: 10.1109/OJCS.2021.3052518.
[12]. U. Orji and E. Ukwandu, "Machine learning for an explainable cost prediction of medical insurance," *Machine Learning with Applications*, vol. 15, p. 100516, 2024, doi: 10.1016/j.mlwa.2023.100516.
[13]. A. Bhongade, Y. Dubey, P. Palsodkar, and P. Fulzele, "Explainable AI Model for Medical Insurance Premium Prediction using Ensemble Learning with SHAP Analysis," in *Proc. IEEE Pune Section Int. Conf. (PuneCon)*, 2024, pp. 1–6.
[14]. S. Goel and A. Chaudhary, "Prediction of Health Insurance Price using Machine Learning Algorithms," in *Proc. Int. Conf. Comput. Sustain. Global Dev. (INDIACom)*, 2024, pp. 1345–1350.
[15]. K. Kaushik, A. Bhardwaj, A. D. Dwivedi, and R. Singh, "Machine Learning-Based Regression Framework to Predict Health Insurance Premiums," Int. J. Environ. Res. Public Health, vol. 19, no. 13, p. 7898, 2022, doi: 10.3390/ijerph19137898.
[16]. M. Abdel-ghany, "Medical Insurance Cost Dataset," Kaggle, 2025. [Online]. Available: https://doi.org/10.34740/KAGGLE/DSV/12853160.
[17]. Mamun, M., Chowdhury, S. H., Faruq, M. O., Azad, M. M., Biswas, B. R., Hussain, M. I., & Hossain, M. M. (2025). Identification of Maternal Health Risk From Optimal Features Using Explainable Machine Learning. Engineering Reports, 7(11), e70491.
[18]. Das, K., Mamun, M., Safat, Y., Hussain, M. I., Hossain, M. M., & Chowdhury, S. H. (2025, July). Optimized Feature-Driven Dengue Diagnosis Using Explainable Machine Learning Approaches. In 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN) (pp. 1-6). IEEE.
[19]. Mamun, M., Hussain, M. I., Ali, M. S., Chowdhury, M. S. A., Chowdhury, S. H., & Hossain, M. M. (2025, July). An Explainable Ensemble Learning Framework with Feature Optimization for Accurate Maternal Health Risk Prediction. In 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN) (pp. 1-6). IEEE.
[20]. H. M. Alzoubi, N. Sahawneh, A. Q. AlHamad, U. Malik, A. Majid, and A. Atta, "Analysis of Cost Prediction in Medical Insurance Using Modern Regression Models," in Proc. Int. Conf. Cyber Resilience (ICCR), Dubai, UAE, 2022, pp. 1–10, doi: 10.1109/ICCR56254.2022.9995926.
[21]. Chowdhury, S. H., Mamun, M., Hossain, M. M., Hossain, M. I., Iqbal, M. S., & Kashem, M. A. (2024, April). Newborn Weight Prediction And Interpretation Utilizing Explainable Machine Learning. In 2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE) (pp. 1-6). IEEE.
[22]. Hussain, M. I., Munir, A., Mamun, M., Chowdhury, S. H., Uddin, N., & Hossain, M. M. (2025). A Transparent House Price Prediction Framework Using Ensemble Learning, Genetic Algorithm-Based Tuning, and ANOVA-Based Feature Analysis. FinTech, 4(3), 33.
[23]. Mustapha, I. B., Abdulkareem, M., Jassam, T. M., AlAteah, A. H., Al-Sodani, K. A. A., Al-Tholaia, M. M., ... & Ganiyu, A. (2024). Comparative analysis of gradient-boosting ensembles for estimation of compressive strength of quaternary blend concrete. International Journal of Concrete Structures and Materials, 18(1), 20.
[24]. Shaikat, M. T. A., Chowdhury, S. H., Shovon, M., Hossain, M. M., Hussain, M. I., & Mamun, M. (2025, February). Explainability Elevated Obstructive Pulmonary Disease Care: Severity Classification, Quality of Life Prediction, and Treatment Impact Assessment. In 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-6). IEEE.
[25]. Chowdhury, S. H., Mamun, M., Shaikat, M. T. A., Hussain, M. I., & Hossain, M. M. (2025, February). Improving Network Classification Accuracy through Feature Clustering and Ensemble Machine Learning with Explainable AI. In 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-6). IEEE.
[26]. Mamun, M., Chowdhury, S. H., Hossain, M. M., Khatun, M. R., & Iqbal, S. (2025). Explainability enhanced liver disease diagnosis technique using tree selection and stacking ensemble-based random forest model. Informatics and Health, 2(1), 17-40.