**Is Ensemble Learning Always the Best Approach? A Comparative Study on Insurance Price Prediction**

Ann Sheng Chin1, 2, a), Pey Yun Goh1, 2, b), Chin Poo Lee3, c) and Lee Ying Chong 1, 2, d)

*1Faculty of Information Science & Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450, Bukit Beruang, Melaka.*

*2Centre for Advanced Analytics, CoE for Artificial Intelligence, Multimedia University, Jalan Ayer Keroh Lama, 75450, Bukit Beruang, Melaka.*

*3University of Nottingham Ningbo China, Ningbo, Zhejiang, China, 315104*

*b) Corresponding author: pygoh@mmu.edu.my*

*a) anson280602@gmail.com*

*c) Chin-Poo.Lee@nottingham.edu.cn*

*d) lychong@mmu.edu.my*

**Abstract.** Insurance price forecasting plays a vital role in enabling fair and data-driven pricing strategies in the insurance industry. Although ensemble methods have shown promising results in prior studies, inconsistencies remain, and the limited number of regression models compared in those studies may not be sufficient to conclude that ensemble methods are superior to individual models. Thus, this study aims to examine whether ensemble models consistently outperform individual regression models in insurance price prediction. Two public datasets, i.e. medical insurance price and medical premium are used with pre-processing steps including data cleaning, encoding, and feature selection. A total of 24 regression models, including 14 individual models and 10 ensembles such as bagging, boosting and stacking were evaluated. The performance was assessed using R², RMSE, and MAE metrics. Experimental results show that while ensemble models often perform well, several individual models achieved comparable or even superior performance in certain cases. Benchmarking with prior studies shows significant performance improvement, possibly due to hyperparameter tuning. A concluding remark is provided at the end of this study.

**INTRODUCTION**

Insurance price forecasting plays a critical role in shaping pricing strategies, enhancing risk assessment, and improving customer satisfaction within the insurance industry [1]. Accurately predicting insurance premiums allows insurers to maximize profits while providing personalized rates to policyholders. However, traditional approaches often have problems in modeling complex and nonlinear relationships among demographic, lifestyle, and regional factors [2].

Recent progress in machine learning (ML) has led to the development of advanced tools for predictive analytics, particularly in the domain of insurance pricing. Commonly used regression techniques include Linear Regression, Decision Trees (DT), and Support Vector Machines (SVM) [3]. While these models are effective in many cases, they often face challenges such as overfitting, underfitting, sensitivity to noisy or incomplete data, and limited ability to generalize beyond the training data [4].

To overcome these restrictions, ensemble learning approaches like Random Forest (RF), Gradient Boosting (GB), and XGBoost (XGB)—have grown in popularity. These approaches incorporate the results of multiple base models to improve overall predictive accuracy [5]. Although ensemble approaches have been demonstrated to outperform individual models in insurance pricing, their superiority is not always guaranteed. In fact, their effectiveness can vary depending on the dataset and problem context. Moreover, ensemble models often request significant computational resources and can be highly sensitive to hyperparameter tuning and data quality issues [6].

Although ensemble techniques such as bagging and boosting have been commonly applied, there is also a lack of studies that carefully evaluate the effectiveness of stacking ensemble regression specifically for insurance datasets [7]. Stacking, which integrates predictions from multiple base models through a meta-learner, offers potential for improved accuracy and generalization. However, there is still limited study about stacking in the insurance domain.

This study aims to fill these gaps by conducting a comprehensive comparative analysis of 24 regression models – including individual learners and ensemble techniques such as bagging, boosting, and stacking across two real-world insurance datasets, i.e. Medical Insurance Price (MedIP) [8] and medical premium (MedPr) [9] from Kaggle. However, only one dataset is considered in benchmarking with other related works due to the availability of benchmarking results from prior research. While only one of the datasets has published benchmarking results, the second dataset introduces valuable variation in feature distribution and complexity, allowing for cross-contextual validation of model robustness.

In contrast to prior work that typically evaluates a limited subset of models or focuses on a single dataset, our research makes the following contributions: 1) a broad spectrum of 24 regression models are evaluated, providing a comprehensive understanding of model performance in insurance price prediction; 2) we assess model behavior across two distinct datasets to evaluate generalizability; 3) we examine the role and effectiveness of stacking ensembles by comparing them to bagging, boosting and individual models; and 4) through detailed performance comparisons using RMSE, R², and MAE, we find that certain individual models can achieve performance comparable to complex ensembles, suggesting that ensemble methods are not always the superior choice.

The remainder of this paper is structured as follows. Section 2 reviews related works in insurance price prediction and ensemble modeling. Section 3 presents the proposed methodology, including data preprocessing, model development, and stacking ensemble construction. Section 4 reports the experimental setup, evaluation metrics, and results analysis. Finally, Section 5 concludes the paper and outlines potential directions for future research.

**LITERATURE REVIEW**

A growing number of studies have used machine learning approaches to estimate health and medical insurance costs based on demographic information and lifestyle-related factors. This section reviews recent research, with a special emphasis on regression algorithms, ensemble learning approaches, and data preprocessing methodologies applicable to insurance pricing.

Ensemble learning methods have gained popularity because of their ability to improve prediction accuracy and model robustness. For example, Wang et al. [5] introduced a robust ensemble framework combining models such as Linear Regression, Decision Trees (DT), Random Forests (RF), and Gradient Boosting (GB) within a stacked architecture. Their study also looked into model interpretability using SHAP and PDP visualizations, while noting limitations in computational cost and the need for high-quality labeled data. Similarly, Lee et al. [6] proposed a stacked ensemble that combined Artificial Neural Networks with XGB, RF, and DT. Although their model improved predictive performance, it suffered with generalizability due to a lack of dataset diversity.

Shakhovska et al. [7] examined the effectiveness of stacking, boosting, and bagging by integrating k-nearest neighbor (KNN) and Stochastic Gradient Boosting with a RF meta-learner. Their stacking ensemble outperformed Support Vector Regression (SVR). Nonetheless, questions about computing complexity and real-time deployment persisted. Chittilappilly et al. [10] utilized the Tree-based Pipeline Optimization Tool (TPOT), an AutoML framework, to automate model selection and hyperparameter tuning. Their TPOT-generated ensemble achieved high accuracy. However, this approach has longer training times and the interpretability is reduced.

In terms of model comparison, Sharma and Jeya [11] evaluated tree-based regression models including XGB, GB, and Lasso Regression, focusing on encoding strategies and data splits. Panda et al. [12] introduced machine learning health insurance prediction system and tested various regression algorithms on a Kaggle dataset. Bharti and Malik [13] compared five regression techniques using 1,338 records. Thejeshwar et al. [14] compared Linear Regression, SVR, and RF, highlighting their scalability and computational efficiency. Cenita et al. [15] evaluated GB, SVR, and Linear Regression, reporting the highest R² value of 0.892 for GB.

Ramachandran et al. [16] proposed a RF-based model that incorporated lifestyle and health history features to improve prediction accuracy, while also addressing interpretability and ethical considerations. Vijayalakshmi et al. [17] used a large dataset with 25,000 records and 24 features, including unique attributes such as glucose levels and exercise habits, showing how feature diversity can improve regression outcomes. Islam et al. [18] tested multiple regressors and found GB performed best with 92% accuracy, highlighting the importance of preprocessing and variable interaction.

Overall, majority studies [5-7] demonstrated that ensemble learning methods such as stacking and boosting often outperform standalone regression models in insurance price prediction tasks. However, whether ensemble models consistently deliver superior results across diverse datasets and problem contexts remains an open question. In addition, common challenges remain, including the complexity of hyperparameter tuning, difficulties in modeling nonlinear relationships and categorical variables, and sensitivity to incomplete or noisy data [5][6][7]. Another limitation in many studies [6-7][10-15] is the evaluation of limited subset of regression models. Although several studies have compared a few regression algorithms for insurance cost prediction, their conclusions are inconsistent—some report better results with polynomial models, while others highlight linear or tree-based regressors [12][13][15].

Motivated by these research gaps, this study poses the critical research question: is ensemble learning always the best approach? A comprehensive evaluation of 14 individual models and 10 ensemble models (including bagging, boosting and stacking) are performed under consistent preprocessing and tuning conditions. Through this extensive comparison, we aim to provide a more robust conclusion about the reliability of ensemble models in insurance cost prediction tasks.

**METHODOLOGY AND ENSEMBLE MODELS**

This study adopts a structured methodology that includes data acquisition, preprocessing, exploratory analysis, model selection, and performance evaluation. Two publicly available datasets were used: Medical Insurance Price (MedIP) dataset with 1,338 records and 7 attributes [8]. This dataset includes demographic and health-related features such as age, BMI, gender, smoking status, income, and family health history. It provides a rich foundation for modeling insurance cost and premium structures. Another medical premium (MedPr) dataset contains 986 samples and 11 attributes [9]. Most individuals in this dataset are aged between 25 and 55, with varying health profiles. The dataset also considers major surgeries, chronic diseases, and family history of cancer, which significantly contribute to determining premium pricing.

Correlation heatmaps and distribution plots were used to discover influential features and detect anomalies. Preprocessing involved label encoding for categorical variables (e.g., region, gender), handling missing values via mean imputation, and applying feature selection using mutual information scores to retain the most relevant predictors. Mutual information was selected over other techniques such as PCA, Lasso, or Recursive Feature Elimination (RFE) due to its ability to capture non-linear dependencies between features and the target variable, while maintaining interpretability. It is also computationally efficient and model-agnostic, making it suitable for different regression tasks without depending on assumptions like linearity or feature independence. Both datasets were split into training and testing sets using an 80:20 ratio, which is in line with previous research practices [18], ensuring unbiased evaluation. The benchmarked works are [6] and [11], where the former employed a 10-fold cross-validation, while the latter used 80:20 train-test split. In this study, we adopt similar experimental settings to enable fair and meaningful comparison with these prior works. For all the model training process, fine-tuning is performed using grid search, and standard preprocessing techniques are applied to mitigate outliers and noise.

In constructing the stacking ensemble, we first trained 24 regression models on each dataset individually. The models were primarily selected based on their R-squared (R²) scores, which reflect explanatory power in regression tasks. In cases where R² values were comparable, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were also considered to support a balanced selection. Finally, we selected the top 5 models, i.e. DT, RF, KNN, Extra Trees and GB. This fixed set was then consistently used as the base learners for stacking across both datasets. This approach ensures that only the most accurate and diverse models contribute to the meta-learner, enhancing the ensemble’s predictive strength and generalization capability. While computational cost and training time are important considerations for large-scale deployments, they were not emphasized in this study due to the relatively small dataset size (approximately 1300 samples). All models, including the stacking ensembles, were trained within seconds on standard hardware, and therefore computational overhead was minimal and not a limiting factor in this context.

Three performance metrics: RMSE, MAE, and R² are used. Two ensemble techniques were explored: soft voting and stacking. In the voting ensemble, predictions from regressors were aggregated through a weighted average, where weights were assigned based on each model’s prior accuracy. In contrast, the stacking ensemble employed a two-level structure. The base models generated predictions at level one, which were then passed to a meta-learner (Linear Regression) at level two. This strategy enabled the meta-model to learn optimal combinations of base predictions, enhancing accuracy and generalization.

**EXPERIMENTAL RESULTS**

This section presents the experimental setup and performance analysis of ensemble regression models used for predicting medical insurance charges and premiums. The comparative analysis of 24 regression models are reported in Table 1 and 2.

As shown in Table 1 and 2, while many ensemble models perform well overall, some individual models also demonstrate strong predictive performance. In the MedIP dataset, the DT model achieves an R² of 0.9645, closely approaching the R² of XGBoost (0.9655). Moreover, DT records the lowest MAE at 496.12. Although its RMSE is not the lowest, it is the second-best at 2310.44 — just behind XGB (2301.05). Individual models achieved an average R², MAE and RMSE of 0.7778, 3482.79 and 5640.47, respectively. In contrast, ensemble models performed better, with average results of 0.9391 for R², 1464.73 for MAE and 2969.45 for RMSE.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TABLE 1.** Comparative analysis of individual regression models | | | | | | | | |
| **No.** | **Individual Models** | **Dataset 1: MedIP** | | |  | **Dataset 2: MedPr** | | |
| **R²** | **MAE** | **RMSE** |  | **R²** | **MAE** | **RMSE** |
| 1 | Linear Regression | 0.7397 | 4163.33 | 6320.89 |  | 0.9801 | 659.22 | 920.62 |
| 2 | DT | **0.9645** | **496.12** | **2331.85** |  | 0.9625 | 730.64 | 1264.87 |
| 3 | Bayesian Ridge | 0.7397 | 4161.80 | 6319.74 |  | 0.9801 | 659.47 | 920.32 |
| 4 | Stochastic Gradient Descent | 0.7395 | 4169.13 | 6322.87 |  | 0.9789 | 692.99 | 946.79 |
| 5 | Kernel Ridge | 0.8496 | 2720.73 | 4803.90 |  | 0.9913 | 377.18 | 609.67 |
| 6 | ElasticNet | 0.7393 | 4175.63 | 6325.57 |  | 0.9803 | 667.87 | 917.14 |
| 7 | Lasso | 0.7396 | 4161.83 | 6322.02 |  | 0.9801 | 662.48 | 921.19 |
| 8 | Gaussian Process | 0.8682 | 2352.40 | 4497.56 |  | 0.9903 | 410.32 | 643.85 |
| 9 | KNN | **0.9579** | **668.85** | **2540.26** |  | 0.9929 | 112.13 | 550.20 |
| 10 | Huber | 0.7363 | 4111.68 | 6362.32 |  | 0.9799 | 650.85 | 924.19 |
| 11 | Passive Aggressive | 0.6054 | 5663.86 | 7782.46 |  | 0.9747 | 649.03 | 1039.03 |
| 12 | Orthogonal Matching Pursuit | 0.7386 | 4170.97 | 6333.96 |  | 0.9708 | 850.72 | 1115.45 |
| 13 | RANSAC | 0.6945 | 3590.6 | 6847.39 |  | 0.9801 | 659.22 | 920.62 |
| 14 | Tweedie | 0.7766 | 4152.11 | 5855.8 |  | 0.9799 | 652.92 | 924.91 |

For the MedPr dataset, the results are even more noteworthy. Several models achieve an R² above 0.99, including individual models such as Kernel Ridge, Gaussian Process, and KNN, alongside all ensemble models (except AdaBoost, with an R² of 0.9829). Bagging achieves the highest R² (0.9974), RF achieves the lowest MAE (85.45), and Extra Trees achieves the lowest RMSE (353.94). For the MedPr dataset, both individual and ensemble models performed comparably well. Individual models achieved average values of 0.9801 for R², 602.50 for MAE and 901.35 for RMSE, while ensemble models achieved slightly better averages of 0.9953 (R²), 192.77 (MAE) and 419.12 (RMSE).

We expected stacking ensembles to enhance performance, as the top five performing models were selected for combination. On the MedIP dataset, the stacking model achieved an R² of 0.9613, MAE of 868.99, and RMSE of 2437.54—comparable to the best-performing individual (DT) and ensemble (XGBoost) models. For the MedPr dataset, stacking reached an even higher R² of 0.9971, with MAE of 132.83 and RMSE of 357.12, outperforming all individual and compatible with other ensemble methods. These results suggest that stacking can offer marginal improvements in highly predictable datasets, but its added complexity may not always justify the gains, especially when simpler models already perform strongly. Thus, stacking should be considered selectively, particularly in scenarios where performance gains are critical and computational resources are not a constraint.

While ensemble models generally outperform individual models in terms of average metrics, our findings suggest that their superiority is not universal. On the MedIP dataset, which includes nonlinear and categorical features such as smoking status and family health history, ensemble models like XGBoost and Extra Trees consistently achieved higher R² and lower MAE/RMSE. However, individual models such as DT and KNN also performed competitively, indicating that simpler models can still capture essential patterns when the dataset is moderately sized and structured.

In contrast, for the MedPr dataset—where relationships may be more linear or the signal-to-noise ratio is higher—even individual models achieved excellent R² scores (up to 0.98 on average), with ensemble models offering only marginal improvements. These results suggest that model selection should be context-sensitive: in smaller datasets (may be fewer than 1,500 records), time-sensitive applications, or environments with limited computational resources, well-tuned individual models can provide a strong balance between performance and efficiency.

Although the primary focus of this study is to evaluate the comparative performance of individual and ensemble models, we also include benchmarking against prior studies for reference. Two relevant studies are [6] and [11], both using the MedIP dataset. The former employed 10-fold cross validation and the latter used 80:20 train-test split which the experimental setup is similar to this study. The benchmarked results and the experiment results of this study are summarized in Table 3 and 4.

**TABLE 2.** Comparative analysis of ensemble regression models

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Ensemble Models** | **R²** | **MAE** | **RMSE** |  | **R²** | **MAE** | **RMSE** |
| 1 | RF | 0.9423 | 1428.68 | 2974.97 |  | **0.9973** | **85.45** | **338.96** |
| 2 | CatBoost | 0.9463 | 1459.58 | 2871.17 |  | **0.9970** | **166.62** | **356.77** |
| 3 | LGBM | 0.9340 | 1886.43 | 3181.73 |  | 0.9960 | 184.79 | 395.07 |
| 4 | XGB | **0.9655** | **923.18** | **2301.05** |  | 0.9960 | 212.97 | 400.73 |
| 5 | GB | 0.9359 | 1643.08 | 3134.27 |  | **0.9973** | **120.86** | **340.75** |
| 6 | AdaBoost | 0.8393 | 3109.06 | 4966.1 |  | 0.9829 | 627.23 | 855.02 |
| 7 | Bagging | 0.9496 | 1314.32 | 2781.01 |  | **0.9974** | **79.09** | **336.08** |
| 8 | Extra Trees | **0.9627** | **623.66** | **2391.92** |  | **0.9971** | **88.08** | **353.94** |
| 9 | Hist GB | **0.9541** | **1390.34** | **2654.73** |  | 0.9951 | 229.77 | 456.75 |
| 10 | Stacking (top 5 individual + ensemble models) | 0.9613 | 868.99 | 2437.54 |  | 0.9971 | 132.83 | 357.12 |

In order to ensure a fair comparison with [6], 10-fold cross-validation was applied to the XGB, DT, and RF models, as these models are common to both studies. The performance in the current study is superior, with higher R² values and lower MAE and RMSE for each model. By averaging the results of the selected models in [6], the reported R² is 0.8480, MAE is 2577.75, and RMSE is 4591.93. In contrast, our study achieved a higher R² of 0.9767, a lower MAE of 654.59, and an RMSE of 1716.07. These results are indeed very encouraging.

All models reported in [11] were included in our benchmarking, i.e. Linear Regression, XGB, Lasso, RF, Ridge, DT, KNN and GB. Results of XGB, RF, Ridge, DT, KNN and GB are well performed except Linear Regression and Lasso. Although Linear Regression and Lasso both score a lower R2 and higher RMSE, interestingly these models score a lower MAE in this study. For the above-mentioned models (i.e. XGB, RF, Ridge, DT, KNN and GB), the results of all models are averaged where the R2 is 0.6763, with a mean MAE of 3950.55 and RMSE of 6459.86. In contrast, our study achieved a significantly higher average R2 of 0.9360, with a lower MAE of 1313.44 and RMSE of 3014.38. A huge improvement can be observed in the comparison with [6] and [11]. This could be due to hyperparameters tuning via grid search. Notably, DT and KNN (individual models) achieved performance comparable to some ensemble models such as RF and GB.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **TABLE 3.** Benchmarked result with [6] | | | | | | | |
| **Models** | Results in [6] | | |  | Results in this Study | | |
| **R²** | **MAE** | **RMSE** |  | **R²** | **MAE** | **RMSE** |
| XGBoost | 0.8480 | 2556.07 | 4601.45 |  | **0.9842** | **625.84** | **1421.38** |
| DT | 0.8412 | 2664.09 | 4686.86 |  | **0.9789** | **258.06** | **1568.80** |
| RF | 0.8549 | 2513.10 | 4487.47 |  | **0.9670** | **1079.88** | **2158.02** |

**TABLE 4.** Benchmarked result with [11]

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Results in [11] | | |  | Results in this Study | | |
| Linear Regression | 0.7447 | 4267.21 | 6191.69 |  | 0.7397 | 4163.33 | 6320.89 |
| XGB | 0.8681 | 2381.56 | 4450.43 |  | **0.9655** | **923.18** | **2301.05** |
| Lasso | 0.7447 | 4267.16 | 6191.73 |  | 0.7396 | 4161.83 | 6322.02 |
| RF | 0.8371 | 2747.46 | 4944.73 |  | **0.9423** | **1428.68** | **2974.97** |
| Ridge | 0.7448 | 4273.45 | 6190.80 |  | **0.8496** | **2720.73** | **4803.9** |
| DT | 0.7003 | 3324.37 | 6708.47 |  | **0.9645** | **496.12** | **2331.85** |
| KNN | 0.0394 | 8592.55 | 12010.89 |  | **0.9579** | **668.85** | **2540.26** |
| GB | 0.8679 | 2383.91 | 4453.83 |  | **0.9359** | **1643.08** | **3134.27** |

**CONCLUSION**

This study explored the effectiveness of ensemble models compared to individual regression models for insurance price prediction. In the experiments using two public insurance datasets, 24 regression models were evaluated, and performance was measured using R², RMSE, and MAE. While ensemble models generally performed well, certain individual models such as DT and KNN also showed strong results, highlighting that ensemble methods are not always superior. This suggests that model selection should consider dataset characteristics rather than relying solely on ensemble approaches. For future work, exploring deep learning models or further studies on model interpretability and real-time prediction can also help improve the practical deployment of such machine learning models.

**ACKNOWLEDGMENTS**

The authors would like to express their sincere gratitude to Multimedia University for the support and resources provided throughout the course of this research. This work would not have been possible without the university’s continued encouragement and infrastructure support.

**REFERENCES**

1. D. Banger, “Predictive Insurance Pricing and Risk Analysis,” Medium, (2024).
2. E.W. Frees, R.A. Derrig, and G. Meyers, editors, *Predictive Modeling Applications in Actuarial Science* (Cambridge University Press, New York, 2016).
3. “ML | Underfitting and Overfitting,” GeeksforGeeks, (2017).
4. “1.11. Ensembles: Gradient boosting, random forests, bagging, voting, stacking,” Scikit-Learn, (n.d.).
5. Z. Wang, X. Chen, Y. Wu, L. Jiang, S. Lin, and G. Qiu, “A robust and interpretable ensemble machine learning model for predicting healthcare insurance fraud,” Sci Rep **15**(1), 218 (2025).
6. L. Sijie, F. Sia, R. Alfred, and E.G. Moung, “Ensemble Machine Learning Method for Health Insurance Premium Prediction,” in *2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, (IEEE, Kota Kinabalu, Malaysia, 2024), pp. 642–646.
7. N. Shakhovska, N. Melnykova, V. Chopiyak, and M. Gregus Ml, “An Ensemble Methods for Medical Insurance Costs Prediction Task,” Computers, Materials & Continua **70**(2), 3969–3984 (2022).
8. Kaggle, “Medical insurance price prediction,” (n.d.).
9. Kaggle, “Medical insurance premium prediction,” (4 August 2021). https://www.kaggle.com/datasets/tejashvi14/medical-insurance-premium-prediction
10. R. Mary Chittilappilly, S. Suresh, and S. Shanmugam, “A Comparative Analysis of Optimizing Medical Insurance Prediction Using Genetic Algorithm and Other Machine Learning Algorithms,” in *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, (IEEE, Chennai, India, 2023), pp. 1–6.
11. A. Sharma, and R. Jeya, “Prediction of Insurance Cost through ML Structured Algorithm,” in *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, (IEEE, Greater Noida, India, 2024), pp. 495–500.
12. S. Panda, B. Purkayastha, D. Das, M. Chakraborty, and S.K. Biswas, “Health Insurance Cost Prediction Using Regression Models,” in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, (IEEE, Faridabad, India, 2022), pp. 168–173.
13. A. Bharti and L. Malik, “Regression Analysis and Prediction of Medical Insurance Cost,” Int. J. Create Res. Thoughts (IJCRT) **10**(3), 83-88 (2021).
14. Thejeshwar. T, S.Harsha. T, V.Krishna. V, and Kaladevi. R, “Medical Insurance Cost Analysis and Prediction using Machine Learning,” in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, (IEEE, Uttarakhand, India, 2023), pp. 113–117.
15. Richwell Colleges, Incorporated, Philippines, J.A. Cenita, P.R. Asuncion, Polytechnic College of Botolan, Philippines, J. Victoriano, and Bulacan State University, Philippines, “Performance Evaluation of Regression Models in Predicting the Cost of Medical Insurance,” IJCSR **7**, 2052–2065 (2023).
16. V. Ramachandran, A.R. Kavitha, and R. Pandimeena, “An Accurate Prediction of Medical Insurance Cost Using Forest Regression Algorithms,” in *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, (IEEE, Chennai, India, 2023), pp. 1–4.
17. V. Vijayalakshmi, A. Selvakumar, and K. Panimalar, “Implementation of Medical Insurance Price Prediction System using Regression Algorithms,” in *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, (IEEE, Tirunelveli, India, 2023), pp. 1529–1534.
18. Md.A. Islam, A. Nag, P. Chandra, B.K. Mishra, S.M.F.A. Fahim, and M.M. Hoque, “Healthcare Cost Patterns and Prediction: Investigating Personal Datasets Using Data Analytics,” in *Proceedings of the 3rd International Conference on Signal and Data Processing*, edited by R.K. Shevgaonkar, D. Adhikari, S.K. Nayak, and F.D. Wet, (Springer Nature Singapore, Singapore, 2025), pp. 341–359.