International Journal of Multidisciplinary Research and Growth Evaluation.

Adaptive Health Insurance Modeling for Global Crises: Forecasting Impact of COVID-19 through Machine Learning Methods

Rajesh Goyal
Glen Mill, Pennsylvania, United States

* Corresponding Author: **Rajesh Goyal**

Article Info

ISSN (online): 2582-7138
Volume: 06
Issue: 03
May-June 2025
Received: 10-03-2025
Accepted: 08-04-2025
Page No: 820-827

Abstract
The insurance industry is affected by the COVID-19 pandemic because it contractually covers mortality and health risks. There are many different impacts, some of which balance out the bad ones. Over the past few decades, the health insurance market has been crucial to the overall growth of the Indian insurance sector. This study suggests using COVID-19 patient data to estimate risk using a machine learning-based method. The dataset undergoes a thorough preparation process that involves normalizing it with the Min-Max scaler, handling missing values, and identifying outliers. The RF and LR classifiers are evaluated with respect to measures such as ROC-AUC, F1score, re-call, accuracy, and precision. With an AUC of 0.921 and an accuracy of 89.58%, LR beats RF in the experiments, whereas RF only manages an AUC of 0.504 and an accuracy of 88.19%. The suggested LR model outperforms methods like XGBoost, K-Nearest Neighbors, and Linear SVM in terms of accuracy, according to a comparative study with other models. These results demonstrate how well machine learning approaches may improve risk assessment for life insurance underwriting.

DOI: https://doi.org/10.54660/.IJMRGE.2025.6.3.820-827

Keywords: Global Health Crisis, healthcare, COVID-19, Pandemic Impact, Health Insurance, Machine Learning, Covid-19 patients Dataset

1. Introduction

Healthcare is a fundamental pillar of human well-being, ensuring access to medical services, disease prevention, and treatment. Over the years, advancements in healthcare technologies have significantly improved patient care, early disease detection, and treatment outcomes [1]. Worldwide health emergencies reveal fundamental weaknesses in healthcare provisions because they fully expose both insufficient medical infrastructure and unpreparedness, together with unstable funding sources [2]. The COVID-19 pandemic, together with other new pandemics, brought attention to the necessary development of healthcare systems that can adjust to unanticipated medical together with economic conditions [3]. COVID-19 emerged as the most disruptive worldwide health disaster of its kind in contemporary times [4]. The pandemic spread across multiple regions [5] while at the same time, it created excessive strain on medical facilities and economies, remained inactive, and people faced immense psychological distress. Governments enforced rigorous controlling procedures consisting of lockdowns as well as social distancing rules and travel restrictions for virus containment. Strategies meant to protect public health caused major negative economic results that harmed various sectors single businesses and individual citizens. Healthcare specifically dealt with overwhelming demands for medical services and intensive care units (ICUs) and ventilators, as well as healthcare professionals, resulting in resource inadequacy and financial pressure [5].

The pandemic showed how health insurance became the essential defense against medical costs [6]. Healthy people began looking for health coverage due to their need to protect their financial assets in case of hospitalization or long-term treatments and critical care [7]. Health insurance in India made substantial contributions to national insurance operations by producing approximately 29% of all premium payments. Market expansion in the insurance industry resulted from regulatory mechanisms along with public enlightenment about health insurance solutions and the introduction of diversification in coverage options.

The unpredictable nature of COVID-19 revealed weaknesses in established risk assessment models, so researchers started developing flexible insurance systems that include worldwide health emergencies [8].

The healthcare industry uses ML and DL techniques as revolutionary instruments to transform modeling within healthcare and insurance. The methods exploit massive health records from electronic health records (EHRs) [9] combined with insurance claims as well as clinical data for health outcome predictions, insurance premium optimization, and financial risk evaluation. Health insurers gain insights about patient statistics and hospitalization patterns through AI models, which help them create selectable policies that react to changing threats. Using ML algorithms enables predictions of pandemic healthcare costs and risk assessment for individuals to help insurance underwriting processes through improved decision-making [10, 11]. The adoption of ML and DL methods allows insurers to create sustainable policies that use data to deliver efficient management strategies for future international health emergencies.

A. Motivation and Contribution of Paper

The COVID-19 pandemic created substantial changes in the life insurance business, which demonstrates why strong predictive modeling, together with risk assessment approaches, are essential. Life insurance providers need reliable assessment methods for policyholder risk levels because current health effects from the virus present an uncertain long-term health situation. Life insurance underwriting decisions will benefit from predictive accuracy improvements through the application of ML models in this research. The systematic approach of this research to identify patterns in COVID-19 patient data through data preprocessing and model evaluation helps improve risk assessment and fair policy pricing. The main aims and contributions of this study are these:

Increase the precision of life insurance risk assessment for COVID-19-affected persons by utilizing ML models.

Implements data cleaning, outlier removal, and normalization (Min-Max scaling) to enhance model performance.

Implementation of LR and RF in predicting risk factors related to life insurance.

Important metrics to look at when assessing models for reliable decision-making include the F1-score, recall, accuracy, and precision.

B. Novelty and Justification

This study's originality is for comparing many ML models, such as RF and LR, to assess health insurance risk during the COVID-19 pandemic. Unlike previous studies that primarily focused on DL approaches, their research highlights the superior performance of traditional ML models in structured data classification. The findings justify the use of LR due to its highest accuracy and RF for its robust recall, demonstrating their effectiveness in predicting patient outcomes. This study provides valuable insights for insurers by recommending efficient and computationally feasible models for COVID-19-related risk assessment.

C. Structure of the paper

The study is structured as follows: Relevant study on COVID-19 pandemic health insurance is presented in Section II. The methods, materials, and processes are described in depth in Section III. In Section IV, the proposed system's analysis, experimental findings, and discussion are provided. Section V contains the conclusion and more work.

2. Literature Review

This section reviews the literature on applying machine learning to forecast COVID-19 health insurance. Table I shows the literature review summary of the COVID-19 pandemic for health insurance, different papers, methods, datasets used, their key findings and their limitations and future work.

Thejeshwar *et al*. (2023) aim to increase public information about insurance so that people may purchase it at an accurate and reasonable cost. To encourage the adoption of health insurance, it is important to consider factors impacting both the supply-side and demand-side perspectives on insurance, as well as decision-making procedures, attitudes toward buying insurance, and other relevant elements. Furthermore, new and renewal enrolment in health insurance plans may be significantly impacted by awareness levels. The model was tested and validated by comparing the projected quantity with the actual data. Compared to other methods that have been studied, it is faster with a maximum accuracy rate of 87%. This is because it takes less computer time to attain the performance measure [12].

Azam *et al*. (2023) primary contribution is the innovative strategy it presents for using a heterogeneous ensemble learning methodology to predict COVID-19 results. The approach combines many ML models that have been trained on different data segments in an attempt to produce a prediction model that is more accurate and dependable. The study then employs a variety of classifiers, such as NNN, ANN, RF, KNN, LR, DT, and SVM. These models' human accuracy ranges from 74.58 to 82.29 percent. The overall accuracy of 82.29% is achieved by using the effective majority voting method of heterogeneous ensemble learning. The results indicate that this technique performs better than standalone models or other ensemble techniques in the prediction of COVID-19 outcomes, and therefore, this technique can be valuable for healthcare decision-makers [13].

Panda *et al*. (2022) in order to help insurance businesses determine premium rates for market efficiency and health expenditure reduction, the researcher use ML methods to develop the MLHIPS, a real-time insurance premium prediction tool. The suggested model employs a variety of regression approaches to predict insurance premiums and evaluate their efficacy, including Multiple Linear Regression, Polynomial Regression, Simple Linear Regression, Ridge Regression, and Lasso Regression, among others. In comparison to other models, the suggested model's Polynomial Regression model fared the best, with an R-squared value of 0.80 and an RMSE of 5100.53 [14].

Bhatia *et al*. (2022) using the US medical cost personal dataset, which has 1338 entries, is accessible on Kaggle.

Utilizing the dataset's attributes, such as age, gender, BMI, smoking status, number of children, etc., to forecast insurance costs also investigated how price and these characteristics were related using linear regression. They achieved 81.3% accuracy by training the algorithm with a 70-30 split [15].

Mohebbi *et al*. (2022) in order to diagnose COVID-19, In a database of 1354 records, the most popular ML models—K-NN, SVM, DT, RF, NB, NN, and XGBoost—were employed to examine laboratory and clinical information from individuals with and without COVID-19. Given the significance and utilization of clinical and laboratory data in virus identification, this decision was made. XGBoost and K-NN, which were evaluated using the Accuracy, Precision, Re-call, and F1Score criteria, were found to have 97% and 96% accuracy, respectively, in detecting COVID-19 illness [16].

Brilliandy *et al*. (2022) estimate tourism rates using COVID-19 data from many nations using five different regression models. SVM, linear regression, polynomial regression, RF regression, and KNN regression are examples of regression models. They utilize two datasets: the COVID-19 data, which shows the number of cases, and the Indonesian tourism data, which shows the number of foreign visitors to Indonesia each month. The dataset will be processed in the countries with the most tourist arrivals. The preprocessed dataset is split in half (8:2) so that the models may be trained and tested. Random forest regression offers the maximum accuracy, according to the evaluation's results, with an R2 score of 0.9. The number of datasets utilized in their study is limited since other factors may not be taken into account [17].

Table 1: Summary of Studies on COVID-19 in health insurance using ML approaches

Author	Methodology	Dataset	Key Findings	Limitations & Future Work
Thejeshwar *et al*. (2023) [12]	Machine Learning for health insurance pricing	Patient data	Achieved 87% accuracy in insurance price prediction with minimal computation time	Further exploration of additional variables influencing insurance adoption is needed.
Azam *et al*. (2023) [13]	RF, KNN, ANN, SVM, DT, LR, and Random Forest are examples of heterogeneous group learning.	COVID-19 patient data	Achieved 82.29% accuracy using majority voting ensemble method	Future work could involve integrating deep learning techniques for improved accuracy
Panda *et al*. (2022) [14]	Insurance cost estimation is done using ML algorithms as Ridge, Lasso, Simple Linear, Multiple Linear, and Polynomial Regression.	Health insurance cost data	Polynomial Regression performed best with RMSE = 5100.53 and $R^2 = 0.80$	Needs validation on larger datasets and real-world applications
Bhatia *et al*. (2022) [15]	Linear regression for insurance cost prediction	Kaggle Medical Cost Personal dataset (1338 entries)	Achieved 81.3% accuracy in estimating insurance premiums using variables such as smoking, age, and BMI	Dataset size is relatively small; future work could include additional features for improved predictions
Mohebbi *et al*. (2022) [16]	To diagnosis COVID-19, ML methods such as KNN, SVM, DT, RF, NB, NN, and XGBoost	COVID-19 clinical and laboratory dataset (1354 records)	XGBoost and KNN achieved the highest accuracy (97% and 96%)	Further research on real-world implementation and additional patient data is needed
Brilliandy *et al*. (2022) [17]	Regression models (Linear, Polynomial, KNN, Random Forest, SVR) for forecasting tourist	COVID-19 cases and statistics on tourism in Indonesia	Using R2 = 0.9, Random Forest regression performed the best	Limited dataset; additional variables could be considered for more comprehensive analysis

3. Methodology

The proposed methodology for covid-19 pandemic for life insurance is illustrated in Figure 1. The following methodology begins with acquiring the COVID-19 patient's Dataset, which contains labeled transaction data. Initially, a dataset containing COVID-19 patient information is acquired and undergoes data analysis to identify inconsistencies. Data pre-processing is performed, including handling null values, detecting and removing outliers, data cleaning, and eliminating inconsistency and normalization with the min-max scaler method. To allow for model evaluation, after processing, the dataset is divided into training (80%) and testing (20%) groups. RF and LR are trained using the dataset of the implementation. A variety of metrics are used to assess the models' performance, including F1-score, recall, accuracy, and precision.

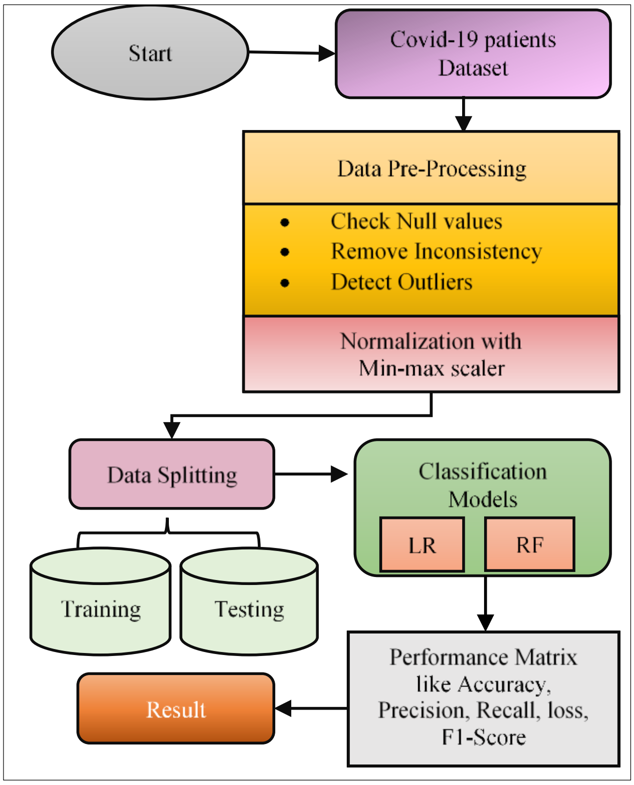

Fig 1: Flowchart for of covid-19 pandemic for health insurance

The following flowchart's steps are simply described below:

A. Data Collection and Analysis

The dataset was gathered from the data repository, which contains 10,000 COVID-19 patient records for both male and female patients. Each patient has 112 qualities, such as physiological and demographic information. After pre-processing, dimensionality reduction is done on this dataset to preserve just 18 key characteristics.

B. Data Preprocessing

Data processing is a fundamental step in building reliable detection models, especially when comparing them. Data cleaning is essential to eliminate noisy or inconsistent data in order to obtain the correct dataset; it may be necessary to do this if the dataset has many unneeded values, outliers, or inconsistencies. A list of the pre-processing procedures is provided below:

- **Check Null values:** Null values in a dataset pose challenges to data analysis and model accuracy, necessitating appropriate handling methods like imputation, removal, or substitution to maintain data quality and ensure robust predictive performance.
- **Remove Inconsistency:** In this step, the inconsistent data in the dataset is eliminated to clean it up.
- **Detect Outliers:** Further study is required to detect contextual and collective outliers, as the detection of global outliers has been the main emphasis of outlier detection.

Normalization with Mini-Max Scaler

The data is converted into a range of 0–1 using the Min-Max normalization. The mathematical description of the Min-Max normalization is given by Equation (1):

$$x_{scaledi} = \frac{x_i - x_{miin}}{x_{max} - x_{min}} \quad (1)$$

Where

$x_{scaledi}$ = the scaled value of a feature;
x_i = the feature's initial value is denoted by this symbol;
x_{min} = dataset's minimum feature value;
x_{max} = dataset's maximum feature value;

D. Data splitting

Datasets are divided into two categories: testing (20%) and training (80%). Using the training data, the models are built and trained, and the test data is used to evaluate how well they perform.

E. Classification of Models

The RF and LR models were used to categorize COVID-19 patient outcomes according to health-related characteristics. The explanation of these models as follows:

1. Logistic Regression (LR)

It is a statistical model-based supervised machine-learning technique that produces the likelihood of a certain class as an output. Its logistic function is used to compute the probability [18]. It determines the probability as Equation (2):

$$P(X) = \frac{1}{1+e^{-(a+bX)}} \quad (2)$$

Where,
X is the input variable
e is the base of natural logarithm$a$and
b are the weights of the Logistic Regression Model
According to Equation 2, $P(X)$ methods 1 as X approaches ∞, while P(X) methods 0 as X approaches -∞. The outcome of applying the logistic function ranges from 0 to 1, including both.

$$Y = \begin{cases} 1, if\ (P(X) > 0.5) \\ 0, otherwise \end{cases} \quad (3)$$

The threshold, which should be the lowest value required to be classified in class 1, is 0.5, as indicated in Equation (3), in order to forecast class from $P(X)$.

2. Random Forest (RF)

Several DT are utilized in the RF ensemble learning approach to generate predictions collectively [19]. Each DT in RF generates its own forecasts, which are then integrated to provide the ultimate forecast. Let X stand for the input characteristics, Y for the target variable, and RF model. This is how the RF prediction may seem if the forest has N decision trees Equation (4):

$$RF(X) = mode(Tree_1\,(X),\, Tree_2\,(X), \ldots.., Tree_n(X)) \quad (4)$$

where the i-th DT forecast is shown by $Tree_1\,(X)$. n a classification job, $mode$ () yields the class label that appears most frequently in all of the trees' predictions. By averaging the predictions, $mode$ () can be substituted in a regression job. The training data is bootstrapped for each DT, and a randomly selected collection of attributes is used to build the predictions of each node. The RF model's ability to aggregate predictions reduces overfitting and enhances generalization performance [20].

F. Performance Metrics

This section delves into the performance metrics obtained throughout the assessment. The following information on the performance parameters utilized in this study is given before the debate begins. P and N represent the total number of positive and negative class examples that were tested, whereas all parameters were determined by categorizing the number of test cases that are FN, FP, TP, and TN. TN is used to describe instances that are correctly categorized as negative, whereas true positive is used to describe cases that are classed as positive but are genuinely positive [21]. Classifier performance is evaluated using precision, accuracy, recall, and f1-score.

Accuracy: It is computed by multiplying the outcome by 100 after dividing the proportion of accurate forecasts by all occurrences. The following Equation of accuracy is (5):

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \times 100 \quad (5)$$

Precision: To verify the system's positive predictions, the ratio of real positive forecasts to total positive forecasts is utilized. Precision is defined as Equation (6):

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Recall: The proportion of successfully detected positives for a certain class to all of the test dataset's real-world class activities. It is possible to compute recall using Equation (7).

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

F1-score: The score, sometimes referred to as the F1-measure, is a composite of two individual measurements: the accuracy and recall harmonic mean, which may be written as Equation (8):

$$F1 - Score = 2\frac{(Precision*Recall)}{Precision + Reall} \quad (8)$$

Calculating the AUC, or area under the ROC curve, requires multiplying the TPR and FPR. The calculation of AUC is done using Equation (9). The value of this measure is always between 0 and 1.

$$AUC = \int_0^1 TPRd(FPR) \quad (9)$$

The TPR is defined as the ratio of TP to the sum of (TP + FN). The FPR is calculated by dividing the number of (FP) by the sum of (FP + TN).

Result Analysis and Discussion

This section provides experiment setup and evaluation of proposed models across key performance indicators. The proposed models are run to enhance data interpretation using the Python programming language platform and its numerous modules, including NumPy and pandas. A system that satisfies the subsequent criteria A 256 solid state drive, two cores, four logical processors, 2304 MHz, 2.30 GHz, and 8 GB of RAM are all features of the Intel (R) Core (TM) i3-6100U CPU. The suggested models' experimental outcomes on the COVID-19 patient dataset are shown in Table II. These outcomes are obtained using LR with an F1score of 94.21, accuracy of 89.58%, precision of 96.58%, and re-call of 91.96%. Recall is 96.87%, accuracy is 88.19%, precision is 90.48%, and the RF model's F1-score is 91.45. Based on the provided dataset, these findings demonstrate how well both models predict COVID-19 instances, with each statistic offering information on how well they do in categorization.

Table 2: Experiment Results of Proposed Models on COVID-19 Patient Dataset

Measures	Logistic Regression	Random Forest
Accuracy	89.58	88.19
Precision	96.58	90.48
Recall	91.96	96.87
F1-score	94.21	91.45

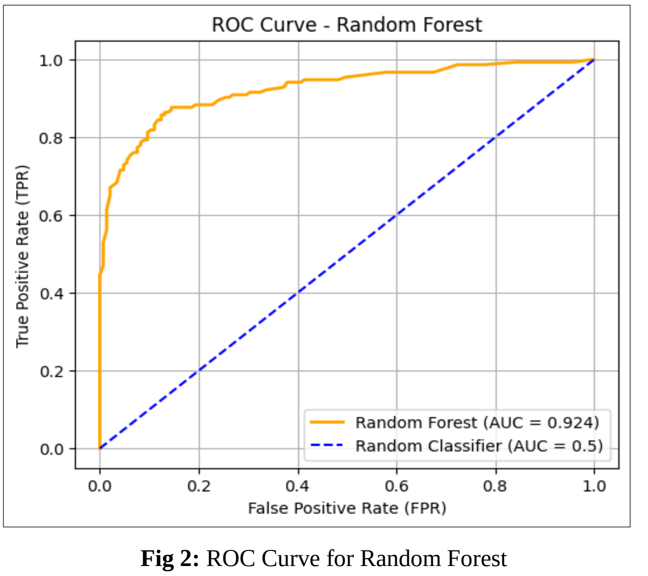

Fig 2: ROC Curve for Random Forest

An RF classifier's ROC curve is displayed in Figure 2. The value of TPR is shown on the y-axis, while the value of FPR is shown on the x-axis. With an AUC value of 0.504 and an orange ROC curve, the RF model performs similarly to a random guess. The random classifier with an AUC of 0.5, which acts as a baseline, is shown by the blue dashed line. The plot suggests that the RF model does not significantly outperform random classification, as its AUC is marginally above 0.5. The figure includes a legend for clarity, gridlines for readability, and a title "ROC Curve - RF" to specify the model under evaluation.

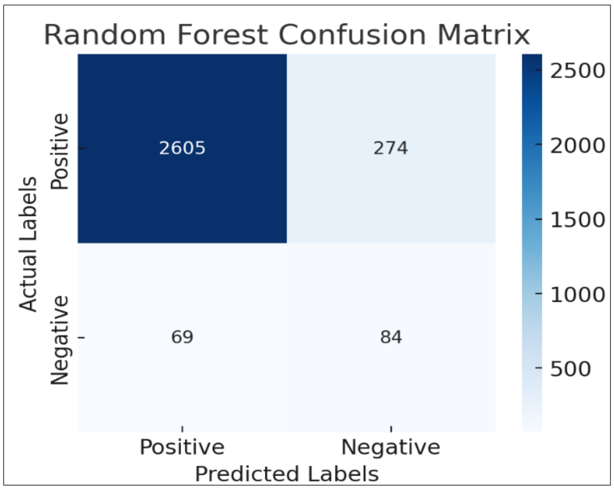

Fig 3: Confusion Matrix of Random Forest

Figure 3 confusion matrix illustrates the RF model's classification performance, which emphasizes actual labels at the rows while predicting labels that appear at the columns. The model properly identified 2605 positive cases (TP) at the same time correctly classifying 84 negative cases (TN). The model wrongly marked 274 positive test results negative and, at the same time, incorrectly identified 69 negative results positive. The figure displays the value data through a visual heatmap that uses darker tones to indicate higher occurrence rates. Evaluation of the model performance is possible through analysis of this matrix which reveals the sensitivity and specificity results.

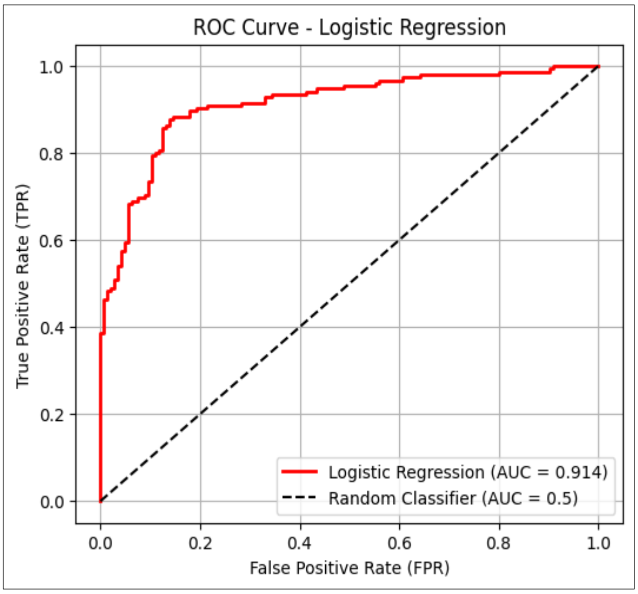

Fig 4: ROC Curve for Logistic Regression

The ROC curve in Figure 4 shows the LR model's performance in Figure 4 through TPR and FPR plots. The model performance appears as a red curve alongside a dashed black line which indicates the AUC value of 0.5 for a random classifier. The discrimination ability of LR is indicated by its AUC value of 0.921. When a model achieves high AUC values, it demonstrates good performance in separating positive from negative classes, thus leading to effective classification results.

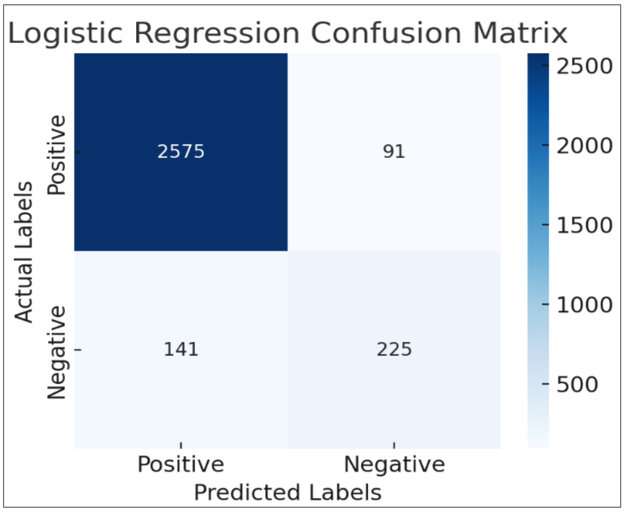

Fig 5: Confusion Matrix of Logistic Regression

The confusion matrix of the LR model, which provides performance information by comparing real label rows with predicted label columns, is displayed in Figure 5. The model produced accurate results in 2575 positive cases and 225 negative cases. The model misdiagnosed both positive cases as negative (FP) and negative cases as positive (FN) 91 times and 141 times, respectively. The pattern of heat values distributed across the heatmap reflects value frequency distribution points through darker color density. The matrix enables the assessment of model effectiveness through sensitivity and specificity evaluations and predictive outcome assessment.

A. Comparative Analysis and Discussion

In this part, the performance of the proposed model and the current model are compared using the same dataset. The numbers in Table III provide details about the subsequent model comparisons.

Table 3: Comparative Analysis based on Propose and existing Models Performance

Models	Accuracy
XGBoost [22]	86.81
DNN [23]	80.97
Linear SVM [24]	87
KNN [25]	87.06
Logistic Regression	89.58
Random Forest	88.19

Table III presents a comparison of proposed models against existing models based on accuracy. The RF model comes in second with an accuracy of 88.19%, while the proposed LR model outperforms all other models with an accuracy of 89.58%. The accuracy rate achieved by KNN stands at 87.06% and Linear SVM shows comparable results at 87%

accuracy. At 80.97%, the DNN model has the lowest accuracy, while the precision of XGBoost is 86.81%, indicating comparable performance. The results of the experiment show that the suggested predictive models outperform earlier predictive techniques in terms of accuracy.

5. Conclusion and Future Direction

The increasing healthcare expenses, together with industry intricacy, demand predictive analysis models that estimate medical insurance costs. This investigation evaluates ML algorithms for medical insurance premium projection during the COVID-19 crisis which generates crucial information for pricing decisions as well as risk management systems. ML serves as a new method to enhance life insurance prediction reliability through analysis of COVID-19-affected patient records. Both proposed models, LR and RF, receive thorough preprocessing combined with evaluation standards, which produce reliable classification results. Results from experiments show that LR performs better than RF since it can discriminate between positive and negative outcomes with 89.58% accuracy and 0.921 AUC metrics. The use of the proposed approach demonstrates higher accuracy than any other models tested in the evaluation process, according to comparative analysis results. The study suggests promising findings; however, it recognizes that DL models provide better performance potential for future model improvements. Additional research ought to examine ensemble systems that integrate several algorithms with data preparation methods after enhance a model's stability. A study adds value to modern decision systems in life insurance underwriting because it produces more precise risk assessments for the post-pandemic world.

6. References

1. Rao D, Sharma DS. Secure and Ethical Innovations: Patenting AI Models for Precision Medicine, Personalized Treatment, and Drug Discovery in Healthcare. Int J Bus Manag Vis. 2023;6(2).
2. Mathur S, Gupta S. Classification and Detection of Automated Facial Mask to COVID-19 based on Deep CNN Model. Proc IEEE 7th Conf Inf Commun Technol. 2023;1-6. https://doi.org/10.1109/CICT59886.2023.10455699
3. Fegert JM, Vitiello B, Plener PL, Clemens V. Challenges and burden of the Coronavirus 2019 (COVID-19) pandemic for child and adolescent mental health: A narrative review to highlight clinical and research needs in the acute phase and the long return to normality. Child Adolesc Psychiatry Ment Health. 2020. https://doi.org/10.1186/s13034-020-00329-3
4. Razu SR, *et al*. Challenges Faced by Healthcare Professionals During the COVID-19 Pandemic: A Qualitative Inquiry From Bangladesh. Front Public Health. 2021. https://doi.org/10.3389/fpubh.2021.647315
5. Haleem A, Javaid M. Medical 4.0 and its role in healthcare during covid-19 pandemic: A review. J Ind Integr Manag. 2020. https://doi.org/10.1142/S2424862220300045
6. Pandya S. A Machine and Deep Learning Framework for Robust Health Insurance Fraud Detection and Prevention. Int J Adv Res Sci Commun Technol. 2023;1332-1342. https://doi.org/10.48175/IJARSCT-14000U
7. Feng A. A Machine Learning Pipeline for Accurate COVID-19 Health Outcome Prediction using Longitudinal Electronic Health Records. AMIA Annu Symp Proc. 2021.
8. Mhlanga D. The Role of Artificial Intelligence and Machine Learning Amid the COVID-19 Pandemic: What Lessons Are We Learning on 4IR and the Sustainable Development Goals. Int J Environ Res Public Health. 2022;19(3):1879. https://doi.org/10.3390/ijerph19031879
9. Pandya S. A Systematic Review of Blockchain Technology Use in Protecting and Maintaining Electronic Health Records. Int J Res Anal Rev. 2021;8(4).
10. Fard SG, Rahimi HM, Motie P, Minabi MAS, Taheri M, Nateghinia S. Application of machine learning in the prediction of COVID-19 daily new cases: A scoping review. Heliyon. 2021. https://doi.org/10.1016/j.heliyon.2021.e08143
11. Neeli SSS. Gen Database Administration: Integrating AI and Technology Advancements. Int J Lead Res Publ. 2022;3(1):10.
12. Thejeshwar T, Harsha TS, Krishna VV, Kaladevi R. Medical Insurance Cost Analysis and Prediction using Machine Learning. Proc Int Conf Innov Data Commun Technol Appl. 2023. https://doi.org/10.1109/ICIDCA56705.2023.10100057
13. Azam M, Yousaf A, Zafar F, Munir MS, Saeed MI. A Real-time Data Analysis Approach for Predicting COVID-19 Outcomes using Heterogeneous Ensemble Learning. Proc Int Conf Eng Emerg Technol. 2023;1-6. https://doi.org/10.1109/ICEET60227.2023.10525802
14. Panda S, Purkayastha B, Das D, Chakraborty M, Biswas SK. Health Insurance Cost Prediction Using Regression Models. Proc Int Conf Mach Learn Big Data Cloud Parallel Comput. 2022. https://doi.org/10.1109/COM-IT-CON54601.2022.9850653
15. Bhatia K, Gill SS, Kamboj N, Kumar M, Bhatia RK. Health Insurance Cost Prediction using Machine Learning. Proc 3rd Int Conf Emerg Technol. 2022. https://doi.org/10.1109/INCET54531.2022.9824201
16. Mohebbi N, Tutunchian M, Alavi M, Kargari M, Kharazmy AB. Supervised Machine Learning Models for Covid-19 Diagnosis using a Combination of Clinical and Laboratory Data. Proc 8th Int Conf Web Res. 2022. https://doi.org/10.1109/ICWR54782.2022.9786248
17. Brilliandy E, Lucky H, Hartanto A, Suhartono D, Nurzaki M. Using Regression to Predict Number of Tourism in Indonesia based of Global COVID-19 Cases. Proc 3rd Int Conf Artif Intell Data Sci. 2022. https://doi.org/10.1109/AiDAS56890.2022.9918731
18. Chalichalamala S, Govindan N, Kasarapu R. Logistic Regression Ensemble Classifier for Intrusion Detection System in Internet of Things. Sensors. 2023. https://doi.org/10.3390/s23239583
19. Tarafdar YH. Finding majority for integer elements. J

Comput Sci Coll. 2018;33(5):187-191.

20. Mahmud I, *et al*. Cardiac Failure Forecasting Based on Clinical Data Using a Lightweight Machine Learning Metamodel. Diagnostics. 2023. https://doi.org/10.3390/diagnostics13152540
21. Khalil MI, *et al*. Deep-COVID: Detection and Analysis of COVID-19 Outcomes Using Deep Learning. Electronics. 2022;11(22):3836. https://doi.org/10.3390/electronics11223836
22. Anmol, Aggarwal S, Badhon AJ. Medical Insurance Cost Prediction Using Machine Learning Algorithms. Lect Notes Networks Syst. 2023;467:271-281. https://doi.org/10.1007/978-981-19-2538-2_27
23. Hanafy M, Mahmoud OMA. Predict Health Insurance Cost by using Machine Learning and DNN Regression Models. Int J Innov Technol Explor Eng. 2021. https://doi.org/10.35940/ijitee.c8364.0110321
24. Siddique S, Chow JCL. Machine Learning in Healthcare Communication. Encyclopedia. 2021. https://doi.org/10.3390/encyclopedia1010021
25. Yeruva S, *et al*. Smart Machine Health Prediction Based on Machine Learning in Industry Environment. Information. 2023. https://doi.org/10.3390/info14030181