A Predictive Approach for Healthcare Expenditure Using Ensemble Techniques

Syed Musharaf, Sripathi Srujan Kumar, B Maruthi Reddy and Bidyutlata Sahoo
Department of CSE(AI&ML), Institute of Aeronautical Engineering, Dundigal, Hyderabad, Telangana, India

Keywords: Healthcare, Ensemble Techniques, Regression Algorithms, Decision Making, XBG Regressor.

Abstract: Accurate estimates of healthcare expenses are necessary due to the rising cost of healthcare brought on by the advent of new viruses and other health issues. These forecasts are essential for individuals and insurance providers to make educated decisions and plan for future medical requirements. Because of the exponential expansion of data and the complexity of computations involved, traditional methods of evaluating health insurance prices have become increasingly insufficient and frequently inaccurate. The objectives of this research are to determine the most accurate ensemble technique for forecasting changes in healthcare spending, assess the effectiveness of ensemble approaches in comparison to traditional models, and help prospective purchasers of medical insurance choose plans that best suit their individual requirements.Through rigorous experimentation and analysis, several regression algorithms, including KNN Regressor, Linear Regressor, Random Forest, Decision Tree, and XGB Regressor, were assessed for their predictive performance. Among these, the most successful performance was XGB Regressor, with an astounding accuracy of 89.7%. This result highlights the algorithm's robustness and effectiveness in handling the complexity inherent in healthcare expenditure prediction.

1 INTRODUCTION

Digital health is a field that is expanding rapidly on a worldwide scale. (Sudhir, Biswajit , Dolly, & Manomitha, 2022) Internationally, the number of digital health businesses has increased during the previous five years. In developed countries, there are two major challenges to health insurance: rising health care costs and a rise in the uninsured population. Medical insurance is a vital part of the medical sector. On the other hand, because patients with uncommon illnesses account for most of the medical cost, it is difficult to forecast spending. Numerous ML algorithms and deep learning approaches are used for data prediction. The factors of training time and accuracy are looked at. The bulk of machine learning algorithms only require a brief time of training. However, the prediction results from these techniques are not particularly accurate.

Deep learning models can also find hidden patterns, but their usage in real-time is constrained by the training period. (Nithya & Dr. V. Ilango, 2017) Regression models such as Linear Regression, XGBoost Regression, Random Forest Regression, Decision Tree Regression, KNN Model, Support Vector Regression, and Gradient Boosting Regression can all be used and used in this study. This study's main goal is to present a novel approach for accurately forecasting healthcare expenses.

2 LITERATURE SURVEY

Numerous research in various contexts on the estimation of healthcare costs have been published in the medical domain. While there are numerous likely assumptions in machine learning, its effectiveness depends on selecting an almost exact method for the given problem domain and adhering to the right protocols for model construction, training, and deployment.(Sudhir, Biswajit , Dolly, & Manomitha, 2022) In order to predict health insurance prices based on certain attribute values observed in the dataset, a number of machine learning regression models were used in this study.

As part of the technique, a dataset containing pertinent characteristics like age, BMI, and smoking status is gathered. Missing value management, feature normalization, and categorical variable encoding are examples of preprocessing procedures. Training and

Musharaf, S., Kumar, S. S., Reddy, B. M. and Sahoo, B.
A Predictive Approach for Healthcare Expenditure Using Ensemble Techniques.
DOI: 10.5220/0013586300004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 2*, pages 41-48
ISBN: 978-989-758-763-4

testing sets of the data are created, with noteworthy features selected. After training several regression models, Polynomial Regression reduces residual error by fitting a curve to the data, resulting in the best accuracy of 80.97%. (Nithya & Dr. V. Ilango, 2017) Six different classification algorithms were used in combination to predict which beneficiaries' inpatient claim amounts increased in 2008 and 2009.

Using the test dataset, the results showed 80% sensitivity, 77.56% overall accuracy, and 76.46% precision, the model proved to be useful in helping high-risk patients select the best insurance plan and in properly estimating cost and revenue for insurance providers.(A. Tike & S. Tavarageri, 2017) Utilizing Medicare payment data, a medical price prediction system was developed to help patients identify less expensive medical providers. The authors present a brand-new hierarchical decision tree method for estimating prices.

Several experiments are conducted to compare this method against several machine learning techniques, including linear regression, Random Forests, and gradient-boosted trees, and the results demonstrate that the hierarchical decision tree achieves a high accuracy. (M Mohammed Hanafy & Omar M.A. Mahmoud, 2021)This study demonstrates how insurance premiums can be predicted using a variety of regression models. They also compared the results of several other models, such as the DNN, Random Forest Regressor, Support Vector Machine, XGBoost, CART, Randomized Additive Model, and k-Nearest Neighbors.

The stochastic gradient boosting model, which produces an R-squared value of 84.8295, an MAE of 0.17448, and an RMSE of 0.38018, is found to be the most successful model. (Hossen, 2023) In this research, they anticipate insurance amounts for different groups of people using both individual and local health data. The effectiveness of these techniques was investigated using nine regression models: Linear Regression, XGBoost Regression, Gradient Boosting, KNN Model, Random Forest Regression, Ridge Regression, Lasso Regression, Decision Tree Regression, and Support Vector Regression. (Lahiri & N. Agarwal, 2014) They use about 114,000 beneficiaries and over 12,400 attributes in publicly available Medicare data. It solved the problem of accurately predicting which customers' inpatient claim amounts increased between 2008 and 2009 by using six different classification algorithms.

The study predicts which beneficiaries inpatient claim amounts grew from 2008 to 2009 using publicly available Medicare data with over 114,000 beneficiaries and 12,400 attributes. Employing an ensemble of six classification techniques, it obtains an overall accuracy of 77.56%, a precision of 76.46%, and a sensitivity of 80% with the test dataset. (Gregori, et al., 2011) After being introduced, regression techniques suitable for healthcare cost analysis were applied in two experimental settings: hospital care for diabetes and cardiovascular treatment (the COSTAMI trial). The study presents regression approaches created especially for healthcare cost analysis.

These techniques are applied in observational settings (diabetes hospital care) as well as experimental ones (the COSTAMI trial in cardiovascular treatment). (Bertsimas, et al., 2008) This study investigates data mining techniques, which are a potent tool for estimating health-care expenses and offer precise estimates of medical costs.

In order to predict medical costs, the study used data-mining techniques. It evaluates their accuracy in two contexts: observational (diabetic hospital care) and experimental (COSTAMI research in cardiovascular medicine). The study looks at the value of medical data in predicting costs, especially for high-priced members, and emphasizes the predictive ability of past cost trends. (Mukund Kulkarni, et al., 2022) This work uses a range of machine learning regression models on a personal medical cost dataset to anticipate health insurance costs based on certain features.

Regression models including gradient boosting, polynomial, decision tree, random forest, multiple linear, and other regression models are studied in this work. All the models are trained on a subset of the dataset, and their performance is evaluated using metrics such as root mean square error (RMSE), mean absolute error (MAE), and accuracy. (Sahu Ajay, Sharma Gopal, Kaushik Janvi, Agarwal Kajal, & Singh Devendra, 2023) “With a score of 85.776%, it was discovered that Gradient Boosting Decision Tree Regression had the best accuracy rate for estimating the amount. Although around 80% of the time, both random forest and linear regression could produce accurate forecasts.

3 DESIGN AND PRINCIPLE OF MODEL

3.1 METHODOLOGY

Using ensemble learning approaches, we created a predictive model for healthcare spending in this study. To

improve prediction accuracy, we used the XGB Regressor, Decision Tree, Random Forest, KNN Regressor, and Linear Regression algorithms. This was a reliable and scalable approach. In order to maximize model performance, our method included preprocessing the dataset, which included addressing missing values and scaling features, and then hyperparameter tweaking. To ensure generalizability, cross-validation techniques were used during the model's training and evaluation process on historical healthcare expenditure data. Important performance indicators including RMSE and MAE were used to evaluate how well the model predicted future healthcare expenses.

3.1.1 Data Source Description

The healthcare expenditure dataset that is used is sourced from KAGGLE's repository. (Medical Cost Personal Datasets, 2018) There are 1338 rows and seven properties or features in the collected data set; Three of the seven characteristics or traits have categorical values, and the remaining four have numerical values. Next, the data set is split into two parts. Subsequently, the dataset is divided in half. For the first and second components, respectively, the words "training data" and "test data" are used. Predictions made by the model will be more accurate the more data it obtains throughout its training process on unknown data. Figure 1 displays the dataset's correlation matrix. Models are evaluated using test sets, while health insurance cost prediction models are developed using training datasets.

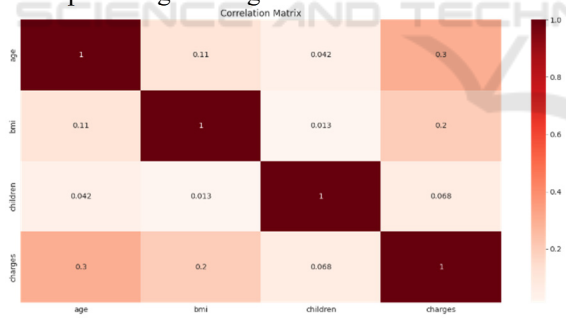

Figure 1: Correlation Matrix of Dataset.

A few fields in the dataset had missing values. Examining the distributions led to the conclusion that new attributes should be used to replace the missing variables, which implies that some data may be missing. The missing data mechanism must first choose the best approach to analyse the data because this is only possible if the data is lost completely at random.

3.1.2 Exploratory Data Analysis(EDA)

Using EDA, the dataset is quickly analysed to look for any hidden patterns, spot abnormalities, confirm assumptions, and test ideas. The data obtained from EDA aids in selecting suitable machine learning techniques to address the required challenge. Several kinds of columns that are present in the data are discovered after conducting an exploratory data analysis on the entire data set.

3.1.3 Preprocessing and Feature Engineering

The dataset has seven variables, as (Sudhir, Biswajit , Dolly, & Manomitha, 2022) Table 1 illustrates. The objective variable, the cost of a customer's charges, is determined by taking into consideration the values of the remaining six variables.

Table 1: Dataset.

Name	Description
Age	Customer's age
BMI	Body mass index of the customer
Number of kids	Total number of children
Gender	Male/Female
Smoker	Whether the customer is smoker or not.
Region	Where the person lives: southwest, southeast, northeast, northwest
Charges (target variable)	Medical charges the customer has to pay

Reviewing the data, accurately rebuilding it, and using machine learning techniques are all part of this step. The dataset was first checked for any missing values. It was found that the dataset contained missing values in the BMI and charges columns. The missing values were imputed using the relevant attribute values' mean values.

One-Hot encoder is used to encode the categorical columns and impute the null values. This process involves preparing the training, testing, and validation sets.

3.1.4 Machine Learning Model Selection KNN Regressor

To forecast a person's healthcare expenses, one machine learning technique used is the K-Nearest Neighbors (KNN) Regressor. After the dataset is prepared, this is trained on the pre-made training set

to encode categorical variables and normalize numerical features. The best 'k' value is found by averaging the costs of the nearest neighbors in the feature space, and this value is used by the model to anticipate healthcare spending using cross-validation.

Utilizing the KNN Regressor's ability to spot minute patterns and relationships in the data, this technique provides a benchmark to be utilized in conjunction with other ensemble approaches like XGBoost and Random Forest Regressor. Prior to hyperparameter change, the KNN model's loss function is ascertained by performance metrics such as Mean Absolute Error, Mean Square Error, and Root Mean Squared Error. We saw that after calculating these measures, the model's loss for testing data was roughly 8494, but its RMSE for validation data was roughly 9862. This means that by adjusting the hyperparameters, we may make our model better.

The following hyperparameters were tuned for the KNN model:

- n neighbors
- weights
- leaf size

Upon finalizing the hyperparameter for leaf size, the minimum RMSE training error is 9734.15 at leaf size of 50, whereas the minimum RMSE validation error is 9789.20 at leaf size of 10. Thus, we adjusted our model so that the leaf size is 10. Consequently, after adjusting the hyperparameter, we obtained the following values: leaf size = 10, weights = "uniform," and n neighbors = 11. This contributes to the KNN regressor's optimal performance.

The estimated RMSE is incredibly high, indicating that the model performs horribly on the test data. Thus, in order to improve accuracy and lower loss/error, we will train our model using various regression models.

Linear Regressor.

The cost of individual healthcare is predicted using the Linear Regression model. Once the training dataset has been processed through a series of data preprocessing steps, such as normalizing numerical features and encoding categorical variables, the trained linear regression model is used. The objective of this model is to create a linear relationship between the goal variable (healthcare costs) and the input parameters (such as age, BMI, smoking status, and region).

The Linear Regression model fits a straight line to the data that minimizes the sum of squared errors to get an easy-to-understand estimate of healthcare expenditures. This model can be used as a benchmark for more intricate ensemble techniques such as XGBoost and Random Forest Regressor.

When we calculated the linear regression model's loss function, we found that it had a far lower RMSE than the KNN model for both training and validation data. To see if it performs any better, we adjusted its hyperparameters.The following hyperparameter were tuned for Linear Regression Model:

- normalize
- fit intercept

We left the hyperparameters at their default values because, even after fine-tuning them, the model is not influenced by the adjustments. This model performs noticeably better than the Knn Model, as evidenced by the noticeably greater RMSE of the Knn Model on the testing data.

Decision Tree.

In order to forecast healthcare expenses based on different individual attributes, the Decision Tree Regressor is used. The Decision Tree model is trained on the training data after the data has been pre-processed to allow for categorical variables and scale numerical features. This model iteratively splits the data into subsets based on the characteristic that decreases variance the maximum, resulting in a structure that resembles a tree. The decisions made based on attributes are represented by each node in the tree, while the expected expenses of healthcare are represented by the leaves. The decision tree model has been found to be entirely overfit to the training set. In order to lessen overfitting on our model, we thus carried out hyperparameter adjustment in the following stage.

The following hyperparameters were tuned for the Decision tree model:

- max-depth
- max-leaf-nodes

When max-leaf-nodes = 8, the training and validation sets produce the greatest results when the max leaf node's hyperparameters are tweaked. This is the way the model is adjusted.

Because of this, the decision tree regression model has outperformed the linear regression model. To explore whether we may obtain better outcomes, we nevertheless evaluated our data using two more models.

Random Forest Regressor.

One essential technique for predictive modelling is Random Forest. First off, it makes reliable data preprocessing easier by handling missing values and categorical variables well, which expedites the production of the dataset. In the modelling stage,

Random Forest's intrinsic capacity to compute feature relevance helps it find critical features that impact healthcare forecasts, like patient demographics and medical history. The group of decision trees in Random Forest then uses the combined predictive capability of several trees to improve accuracy and robustness when predicting healthcare outcomes, such as patient costs or illness progression.

In addition to enhancing prediction stability, this ensemble approach allows for thorough model evaluation, guaranteeing that the forecasts are accurate and suitably in line with the demands of healthcare planning and resource allocation. It is highly likely that the training data will cause this model to overfit. Hyperparameter tunning is thus carried out.

The following hyperparameters were tuned for the random forest regressor:

- n-estimators
- max-depth
- max-leaf-nodes

Max-Leaf-Nodes is set to None by default, meaning that there can be an infinite number of nodes. As a result, when max leaf nodes = 16, the model's training and validation RMSE is at its lowest. We thus adjusted our model in that manner. In comparison to the decision tree model, this has done even better. To check if we could make any more accurate predictions, however, we tested our data using an XGBRegressor.

Gradient Boosting Machines (XGBRegressor).

In order to handle complicated interactions in healthcare data and improve predictive accuracy, the XGBRegressor (Extreme Gradient Boosting Regressor) is essential. XGBoost is used because of its capacity to create a sequence of decision trees one after the other. These trees learn from each other's mistakes and improve predictions one iteratively. First, by determining and ranking the most important variables for healthcare predictions—such as patient demographics, medical history, and environmental factors—XGBoost performs exceptionally well in feature selection and dimensionality reduction. Gradient boosting, in which each new tree concentrates on the residual mistakes of the preceding ones, successfully increases predictive precision, is how it enhances the performance of the model during training.

Furthermore, because of its robustness against overfitting and capacity to handle big datasets with high dimensionality, XGBoost is particularly well-suited for projecting healthcare costs and illness outcomes. This guarantees that the model not only produces accurate predictions but also improves decision-making processes in healthcare planning and resource allocation. Looking at the RMSE of training and validation data suggests that the model may have been overfit using the training set. Therefore, we adjusted the model's hyperparameters to lessen overfitting.

The following hyperparameters were tuned for the Gradient Boosting Machine:

- max-depth
- max-leaf-nodes
- n-estimators
- booster
- min-child-weight

When the max depth hyperparameter is set to 2, the model operates at its peak efficiency. Changing the value of the hyperparameter max-leaf-nodes has no effect on the model when it comes to tuning. In contrast, the model performs best on both the training and validation sets when n estimators equal 100 when tuned with n-estimator hyperparameters. Finally, when max-child-weight = 94 is selected as the hyperparameter for minimum child weight, the model performs best on the training and validation sets. We thus adjusted our model in that manner.

Thus, we observed that XGBRegressor model has the best performance on the testing data.

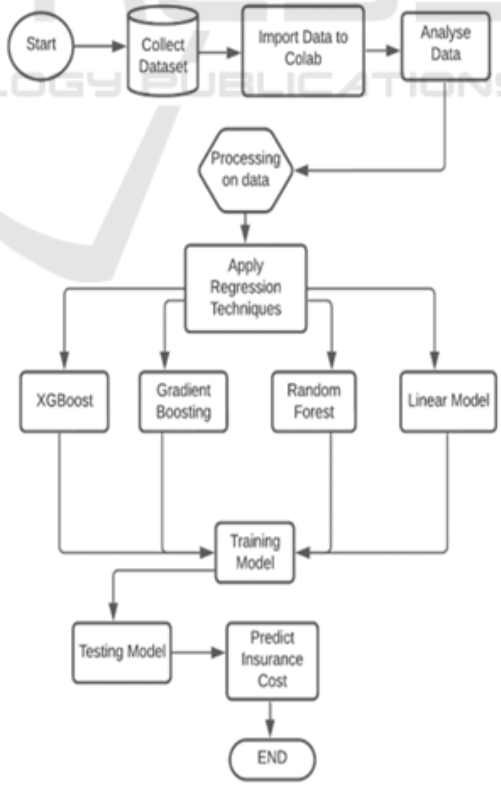

Figure 2: Flowchart of Model.

4 RESULTS AND DISCUSSION

4.1 Experimentation Results

The performance of the model is evaluated using the following metrics:

R^2 Score

Root Mean Square Error (RMSE)

R^2 Score: It is a useful indicator to evaluate the fitness of the model. The R-squared number ranges from 0% to 100%, or 0 to 1. A greater number denotes a better match.

$$R^2 = 1 - SSE/SST$$

The total squared residuals, or squared deviations from each observation's predicted value, is known as the SSE (Squared Sum of Error). $\sum(y_i - y^*)^2$.

Sum of Squared Total, or SST, is the squared deviations of each observation from the mean for the whole sample. $\sum(y_i - \hat{y})^2$.

RMSE: The Root Mean Square error is a commonly used method to estimate a model's prediction error. It shows the model's absolute fit to the data points and shows how close the observed data points are to the model's projected values. A better match is indicated by lower RMSE values. The accuracy of the basis paper model as calculated using these metrics is shown in (Sudhir, Biswajit , Dolly, & Manomitha, 2022) Table 2.

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(y_i - y^*)^2}{n}}$$

Table 2: Base Paper Results.

Model	R2_Score	RMSE	Accuracy(in %)
Simple Linear Regression	0.62	7523.98	62.86
Multiple Linear Regression	0.75	7523.98	75.86
Polynomial Regression	0.80	5100.53	80.97
Ridge Regression	0.75	6070.80	75.82
Lasso Regression	0.75	6066.31	75.86

Table 3: Calculated Results.

Model	R2_Score	RMSE	Accuracy(in %)
KNN Regressor	0.1966	12150.10	19.66
Linear Regressor	0.8057	5975.16	80.57
Decision Tree Regressor	0.8905	4485.31	89.05
Random Forest Regressor	0.8995	4297.03	89.95
XGB Regressor	0.89.72	4345.61	89.72

Following model testing on the test dataset, the results of the previously stated models were recorded. Table 3 shows the accuracy, RSME, and R^2 metrics for the model that has been covered thus far. Figure. 2 depicts how this model operates.

4.2 Observations

When compared to other models, the KNN Regressor performs noticeably worse. The R2 value of 0.1966 indicates a weak link between the input features and anticipated healthcare spending, explaining only 19.66% of the dataset's variation. Since there appears to be a significant mistake in the forecasts, the RMSE of 12,150.10 renders it unreliable for precise forecasting. The model's 19.66% accuracy shows its limitations, which are probably brought on by the dataset's high dimensionality and complexity, which the KNN algorithm finds difficult to handle. KNN is not appropriate for forecasting healthcare spending in this situation due to its subpar performance.

With a R^2 value of 0.8057, the Linear Regression model shows a considerable improvement over the KNN, accounting for around 80.57% of the data's variance. The model successfully captures important patterns between characteristics and the target variable, as evidenced by this strong linear relationship. With a reduced error and an RMSE of 5975.16, it is evident that it can produce predictions that are more dependable than KNN. Although it still lacks the predictive precision and adaptability of more complex models like Decision Trees and Random Forests, the model's 80.57% accuracy makes it a useful baseline for comparison.

With an R^2 score of 0.8905, the Decision Tree Regressor outperforms the other models, accounting for 89.05% of the variance and indicating that it can handle more intricate, non-linear interactions in the

dataset. The RMSE of 4485.31, which is significantly less than that of linear regression, indicates improved prediction accuracy. The Decision Tree model is highly competitive with an accuracy of 89.05%.

The highest performance is the Random Forest Regressor, which explains 89.95% of the variation in the data with an R^2 value of 0.8995. With the lowest RMSE of 4297.03 out of all the models, this one makes the most accurate forecasts. With an accuracy of 89.95%, Random Forest is clearly the best model in terms of precision and generalization. This is partly because Random Forest is an ensemble model, which minimizes overfitting and increases robustness across various data patterns.

Strong performance is also obtained by the XGBoost Regressor, which rivals Random Forest closely with an R2 score of 0.8972, explaining 89.72% of the variance. Although it has a little larger RMSE of 4345.61 than Random Forest, it nevertheless exhibits very excellent prediction accuracy. The 89.72% accuracy rate confirms XGBoost's potency as a prediction model especially for intricate datasets. It is almost as good as Random Forest in this situation because it can reduce bias and variation using gradient boosting techniques.

The best models for forecasting healthcare costs are Random Forest and XGBoost, which have the lowest error and the highest accuracy. While KNN performs poorly, Linear Regression provides a solid baseline but lacks the accuracy of the ensemble models. Although it may be prone to overfitting, the Decision Tree Regressor produces results that are competitive. In general, the best methods for this forecasting task are ensemble models such as Random Forest and XGBoost.

With an accuracy rate of 89.7%, XGBoost's exceptional performance highlights its applicability for challenging predictive tasks in the analysis of healthcare spending. It differs from conventional regression techniques in that it can handle big datasets with high dimensionality, find important predictors, and improve model performance through repeated refining. Accurate spending forecasts are essential for optimizing budget allocation, boosting operational efficiency, and improving patient care in healthcare planning and resource allocation. Healthcare providers can improve overall healthcare management and service delivery by using ensemble approaches like XGBoost to make well-informed decisions based on dependable prediction insights.

5 CONCLUSIONS

This study set out to precisely forecast healthcare costs through the application of multiple ensemble learning strategies. Several algorithms, such as the KNN Regressor, Linear Regressor, Random Forest, Decision Tree, and XGBRegressor, were assessed based on their prediction accuracy through thorough testing and analysis. XGBRegressor stood up as the best performance among all, attaining an astounding accuracy of 89.7%. This outcome demonstrates how well the algorithm handles the complexity involved in predicting healthcare expenses.

The ability of ensemble learning methods, in particular Random Forest and XGBoost, to estimate healthcare costs with low error margins and high accuracy. These models perform better than more straightforward algorithms like KNN and Linear Regression, which are helpful as baselines but fall short in capturing the intricacy of the data. Because ensemble approaches reduce variation and bias, they are essential for producing accurate forecasts. This method offers a strong foundation for predicting healthcare expenditures, which is essential for budgeting, resource allocation, and policy planning in the healthcare industry. It does this by utilizing historical data and cutting-edge machine learning algorithms. Predictive healthcare analytics can benefit greatly from additional improvements made through feature engineering, deep learning integration, and real-time data application, which can result in even more accurate and flexible models.

Future research can go in a number of ways to improve the machine learning models' ability to anticipate healthcare costs. First, the accuracy of the forecasts could be increased by adding more features, like lifestyle indicators, geographic location, and socioeconomic aspects. In order to capture patterns and seasonality in healthcare expenses over time, time-series analysis could be used to enhance the present model to include temporal elements.

Investigating deep learning methods like neural networks, which have the potential to identify more intricate patterns in the data that conventional machine learning models might overlook, is another exciting direction. To strike a compromise between interpretability and accuracy, hybrid models that use both machine learning and deep learning techniques could potentially be investigated.

To further enhance performance, automatic hyperparameter tuning utilizing methods like grid search or Bayesian optimization might be applied to the present models. Better understanding model decisions would also benefit from the application of

explainable AI (XAI) techniques, particularly in a delicate industry like healthcare where openness is essential.

Furthermore, moving to real-time predictive analytics may provide managers of healthcare organizations with timely insights on spending patterns, allowing them to make responsive changes to their resource allocation plans. Maintaining the model's applicability and efficacy in changing healthcare environments requires ongoing validation against fresh datasets and collaboration with healthcare professionals. By seizing these chances, the initiative can improve patient outcomes, cost effectiveness, and healthcare management by advancing predictive analytics in healthcare spending forecasting.

REFERENCES

A. Tike & S. Tavarageri. (2017). A Medical Pricce Prediction System using Hierarchical Decision Trees. IEEE Big Data Conference

Bertsimas, M. V. Bjarnadottir, M. A. Kanre, J. C. Kryder, R. Pandey, S. Vempala, & G. Wang. (2008). Algorithmic prediction of healthcare costs. Operations Research, 56.

Gregori, M. Petrinco, S. Bo, A. Desideri, F. Merletti, & E. Pagano. (2011). Regression Model for Analyzing costs and their determinants in healthcare : an introductory review. International Journal for Quality in Healthcare, 23.

Hossen, S. (2023). Medical Insurance Cost Prediction Using Machine Learning. Dhaka, Bangladesh.

Lahiri, & N. Agarwal. (2014). Predicting Healthcare Expenditure Increase for an Individual from Medicare Data. ACM SIGKDD Workshop on Health Informatics.

M Mohammed Hanafy, & Omar M.A. Mahmoud. (2021). Predict Health Insurance Cost by using Machine Learning and DNN Regression Models. International Journal of Innovative Technology and Exploring Engineering(IJITEE), 10(3, January).

(2018). Medical Cost Personal Datasets. Kaggle Inc.

Mukund Kulkarni, Dhammadeep D. Meshram, Bhagyesh Patil, Rahul More, Mridul Sharma, & Pravin Patange. (2022). Medical Insurance Cost Prediction Using Machine Learning. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 10.

Nithya, B., & Dr. V. Ilango. (2017). Predictive Analytics in Healthcare Using Machine Learning Tools and Techniques. International Conference on Intelligent Computing and Control Systems ICICCS.

Sahu Ajay, Sharma Gopal, Kaushik Janvi, Agarwal Kajal, & Singh Devendra. (2023). Healtch Insurance Cost Prediction by Using Machine Learning . International Conference on Innovation Computing & Communication (ICICC).

Sudhir, P., Biswajit , P., Dolly, D., & Manomitha, C. (2022). Health Insurance Cost Prediction Using Regression Models. Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), May.