PATIENT BILLING TRACKER AND PREDICTION

[1] Dr. M Dileep Kumar, [2] R Kavya, [3] K Palvasha, [4] S Harshith Raj , [5] A Shashank
[1]Professor, [2345]Students
Department of CSE
Siddhartha Institute of Technology & Sciences, Narapally
alladirajashekar_cse@siddhartha.co.in, 23tq1a0526@siddhartha.co.in, 23tq1a0541@siddhartha.co.in, 23tq1a0531@siddhartha.co.in, 23tq1a0537@siddhartha.co.in,

Abstract

The Patient Billing Tracker and Prediction system is designed to manage, analyze, and predict patient billing information in a structured and efficient manner. Hospitals and healthcare organizations generate a large volume of billing records related to consultations, diagnostic tests, medicines, treatments, procedures, room charges, and other medical services. Managing these records manually can be time-consuming and may result in calculation errors, missing information, and difficulty in tracking payment status. The proposed system provides a centralized platform for recording patient billing details, monitoring expenses, tracking paid and pending amounts, and analyzing historical billing patterns. The collected billing data is first preprocessed to handle missing values, duplicate records, inconsistent information, and categorical attributes. Exploratory data analysis is performed to identify relationships between patient characteristics, medical services, treatment types, and total billing amounts. Data visualization techniques can be used to display billing trends, payment status, department-wise expenses, and monthly revenue. Machine-learning algorithms such as Linear Regression, Decision Tree, Random Forest, and Gradient Boosting can be applied to analyze historical billing data and predict the possible billing amount for new or ongoing patients. The system can also classify billing records based on payment status and identify patients with pending payments. Model performance can be evaluated using suitable metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score for billing amount prediction. The system provides useful dashboards and reports that help hospital administrators and authorized staff monitor financial information efficiently. Overall, the Patient Billing Tracker and Prediction system combines billing management, data analysis, visualization, and machine learning to improve billing transparency, support financial planning, and provide data-driven insights into patient healthcare expenses.

I. Introduction

The healthcare industry generates a large amount of financial information through patient consultations, diagnostic tests, medicines, treatments, surgeries, room charges, and other medical services. Managing patient billing information accurately is important for hospitals and healthcare organizations to maintain proper financial records and provide transparent billing services. Traditional billing methods may involve manual calculations, paper records, or basic software systems, which can make it difficult to track expenses, payments, and pending amounts efficiently. Therefore, an automated system that can manage, analyze, and predict patient billing information can improve the overall efficiency of healthcare billing operations.

The Patient Billing Tracker and Prediction system is developed to provide a centralized platform for managing patient billing records and analyzing healthcare

expenses. The system stores important information such as patient details, admission information, treatment charges, medicine costs, diagnostic expenses, room charges, total bill amount, payment status, and payment dates. By maintaining these details in a structured format, the system makes it easier for authorized hospital staff to access and manage billing information. It can also help reduce errors associated with manual billing calculations and record maintenance.

The system begins with the collection of historical patient billing data from hospital records or suitable datasets. The collected data may contain missing values, duplicate records, inconsistent information, and different types of categorical attributes. Data preprocessing techniques are applied to clean and transform the dataset into a suitable format for analysis. Missing values and duplicate records are handled appropriately, while categorical information can be converted into numerical form when required. This preprocessing stage improves the quality and reliability of the data used for billing analysis and prediction.

After preprocessing, Exploratory Data Analysis (EDA) is performed to understand patient billing patterns and identify relationships between different variables. The system can analyze factors such as treatment type, department, length of stay, diagnostic services, medicine expenses, and total billing amounts. Charts, graphs, and tables can be used to display monthly billing trends, department-wise expenses, payment status, and other financial information. These visualizations help administrators and authorized staff understand billing patterns and identify areas requiring attention.

The prediction component uses machine-learning techniques to estimate future or ongoing patient billing amounts based on historical billing records. Regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, and Gradient Boosting can be considered for predicting billing amounts. The models learn relationships between patient and treatment-related attributes and the corresponding billing amount. The trained model can then estimate the expected bill for a new patient based on the available input information.

II. Literature Survey

The increasing use of digital technologies in healthcare has created large amounts of financial and patient-related data. Researchers have explored healthcare cost prediction and billing analysis using statistical methods and machine-learning techniques. These studies show that historical patient information, medical services, treatment details, and previous healthcare expenditure can be useful for predicting future healthcare costs. Such predictive systems can help hospitals, insurance organizations, and patients understand and plan for expected healthcare expenses.

Morid et al. conducted a systematic literature review on supervised learning methods for predicting healthcare costs. Their research compared different approaches for predicting future healthcare expenditure using a large healthcare dataset containing medical and pharmacy claims. The study found that supervised-learning methods can provide strong predictive performance, with gradient boosting performing particularly well overall and artificial neural networks and ridge regression showing strong

performance for high-cost individuals. The research demonstrates the potential of machine learning for healthcare cost prediction and financial planning.

Kulkarni and Ambekar investigated inpatient hospital cost prediction using machine learning. Their study used patient discharge data from New York State and applied feature engineering and different regression algorithms to predict total hospital charges. The researchers compared Random Forest, Stochastic Gradient Descent, K-Nearest Neighbors, Extreme Gradient Boosting, and Gradient Boosting regression models. Random Forest Regression achieved the best reported performance, with an R^2 value of 0.7753, and age group was identified as an important predictor. This study demonstrates how patient and hospitalization characteristics can be used to estimate hospital charges before or during treatment.

Gopukumar, Ghoshal, and Zhao studied predicting hospital readmission charges using machine learning. Their research compared Linear Regression, Lasso, Elastic Net, Ridge Regression, XGBoost, and a multilayer perceptron model using a large Nationwide Readmission Database. The results showed that XGBoost and the multilayer perceptron model performed better than the traditional regression approaches for predicting readmission charges. This research indicates that advanced machine-learning models can be useful for predicting hospital-related financial charges and supporting healthcare financial planning.

Sanz and Leon-Sanz examined the use of machine learning for financial management in healthcare organizations. Their research emphasized that hospitals hold large quantities of data that can be analyzed using Big Data techniques. The study identified billing prediction as an important component of healthcare financial management because predicting future costs can support financial planning and help avoid unexpected expenses. This research supports the development of systems that combine healthcare data analysis with billing prediction.

Recent research has also focused on interpretable healthcare cost prediction. A 2024 study analyzed more than 2.3 million patients using open healthcare data and developed machine-learning models for predicting healthcare costs. The study found that diagnosis and disease severity were important predictors and reported an R^2 score of approximately 0.85 for its best-performing model. The research also emphasized the importance of interpretable models so that healthcare cost predictions can be better understood by users and decision-makers.

Recent reviews indicate that healthcare cost prediction is continuing to move toward more advanced machine-learning and deep-learning techniques. Research comparing deep-learning models with classical regression approaches has investigated models such as LSTM and CNN-LSTM for longitudinal healthcare expenditure and hospitalization-related prediction. These approaches can capture patterns across sequences of healthcare events, although model complexity, interpretability, data quality, and validation remain important considerations.

III. System Analysis

The **Patient Billing Tracker and Prediction** system is designed to manage, analyze, and predict patient billing information in an efficient manner. The system maintains

important billing details such as patient information, treatment charges, medicine expenses, diagnostic costs, room charges, and payment status. Historical billing records are used to identify patterns and relationships between healthcare services and total billing amounts. The system first collects and organizes patient billing data for further processing. Data preprocessing is performed to handle missing values, duplicate records, inconsistent information, and categorical attributes. Exploratory Data Analysis is conducted to understand billing patterns and variations in healthcare expenses. Data visualization techniques are used to represent billing trends, payment status, department-wise expenses, and other financial information. Relevant features are selected and transformed into a suitable format for machine-learning algorithms. Regression algorithms can be trained using historical billing records to estimate the expected billing amount for patients. The trained model can predict the approximate billing amount based on available patient and treatment information. Model performance can be evaluated using metrics such as MAE, MSE, RMSE, and R^2 score. Overall, the system converts patient billing data into useful financial insights and predictions that can support efficient billing management and planning.

Existing System

In the existing system, hospitals generally manage patient billing information using manual records, spreadsheets, or basic hospital management software. Billing information is recorded based on consultations, medicines, diagnostic tests, treatments, room charges, and other healthcare services. Traditional billing systems mainly focus on calculating the current bill and recording payment information. Hospital staff may manually verify individual charges and update payment status for each patient. When the number of patients increases, managing and tracking billing records can become difficult and time-consuming. Historical billing information may be stored but is not always effectively analyzed for identifying financial patterns. Traditional systems generally have limited capabilities for predicting future patient billing amounts. Generating detailed billing reports and visualizations may also require additional manual effort. It can be difficult to identify unusual billing patterns or compare expenses across different departments and treatments. The absence of predictive analytics makes it challenging to estimate expected healthcare expenses in advance. Manual data handling can also increase the possibility of calculation and record-management errors. Therefore, an automated system is required to combine billing tracking, data analysis, visualization, and prediction in a single platform.

Disadvantages of Existing System

- Manual billing management can be time-consuming.
- Large numbers of patient billing records are difficult to manage efficiently.
- Greater possibility of errors during manual calculations and data entry.
- Limited ability to analyze historical billing patterns.
- Limited prediction of future patient billing amounts.
- Difficulty in tracking paid and pending payments efficiently.
- Generating detailed billing reports may require additional manual effort.
- Limited visualization of healthcare expenses and billing trends.
- Difficult to compare expenses across departments and treatment types.
- Historical billing data may not be fully utilized.
- Difficult to identify unusual or unexpected billing patterns.

- Limited support for data-driven financial planning.

Proposed System

The proposed **Patient Billing Tracker and Prediction** system provides an integrated platform for managing, analyzing, and predicting patient billing information. The system stores patient details, treatment information, diagnostic charges, medicine expenses, room charges, total bill amounts, and payment status in a structured manner. Historical billing data is used to identify patterns between patient characteristics, healthcare services, and total expenses. The collected data is first preprocessed by handling missing values, duplicate records, inconsistent information, and categorical attributes. Exploratory Data Analysis is performed to understand billing distributions, treatment costs, payment patterns, and department-wise expenses. Visualization techniques are used to display billing trends and financial information through charts, graphs, and dashboards. Relevant features are selected and transformed into a suitable format for machine-learning models. Regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, and Gradient Boosting can be applied to predict billing amounts. The trained models learn relationships between historical patient and treatment information and their corresponding billing amounts. The best-performing model is selected based on suitable evaluation metrics and used to generate billing predictions. The system can also track paid and pending amounts and provide detailed billing reports to authorized users. Overall, the proposed system helps improve billing management, financial analysis, and estimated healthcare-cost prediction.

Advantages of Proposed System

- Automates patient billing record management.
- Reduces manual effort in maintaining billing information.
- Helps minimize billing calculation and data-entry errors.
- Efficiently manages large numbers of patient billing records.
- Provides prediction of expected patient billing amounts.
- Analyzes historical billing patterns and healthcare expenses.
- Helps track paid, unpaid, and pending billing amounts.
- Provides charts, graphs, and dashboards for better understanding.
- Supports department-wise and treatment-wise expense analysis.
- Helps identify important factors affecting billing amounts.
- Allows comparison of different machine-learning prediction models.
- Supports data-driven financial planning and decision-making.
- Can be extended with online payments and automated invoice generation.

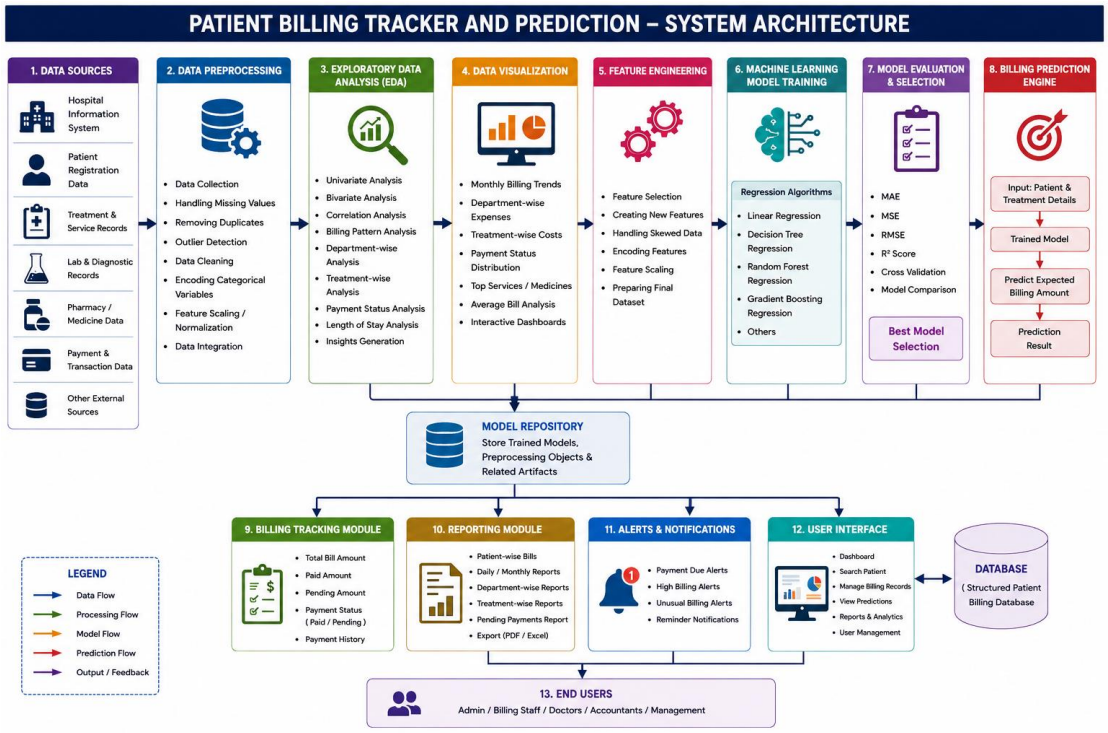
IV. Methodology

The methodology of the **Patient Billing Tracker and Prediction** system consists of several stages, beginning with data collection and ending with billing prediction and reporting. First, historical patient billing data containing patient, treatment, service, and payment information is collected. The dataset is examined to understand its attributes, data types, missing values, and target billing variable. Data preprocessing is performed to handle missing values, duplicate records, inconsistent information, and categorical attributes. Categorical variables are converted into numerical

representations using suitable encoding techniques. Exploratory Data Analysis is performed to understand billing distributions, treatment expenses, payment patterns, and relationships between different variables. Data visualization techniques are used to identify important trends and patterns in patient billing information. Relevant features such as treatment type, length of stay, diagnostic expenses, medicine costs, and other available attributes are selected for prediction. The processed dataset is divided into training and testing datasets for developing and evaluating machine-learning models. Regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, and Gradient Boosting can be trained using the prepared data. The models are evaluated using MAE, MSE, RMSE, and R^2 score to determine their prediction performance. Finally, the best-performing model is selected to predict the expected billing amount, while the system displays billing information, payment status, analytical charts, and prediction results through a user-friendly interface.

System Architecture

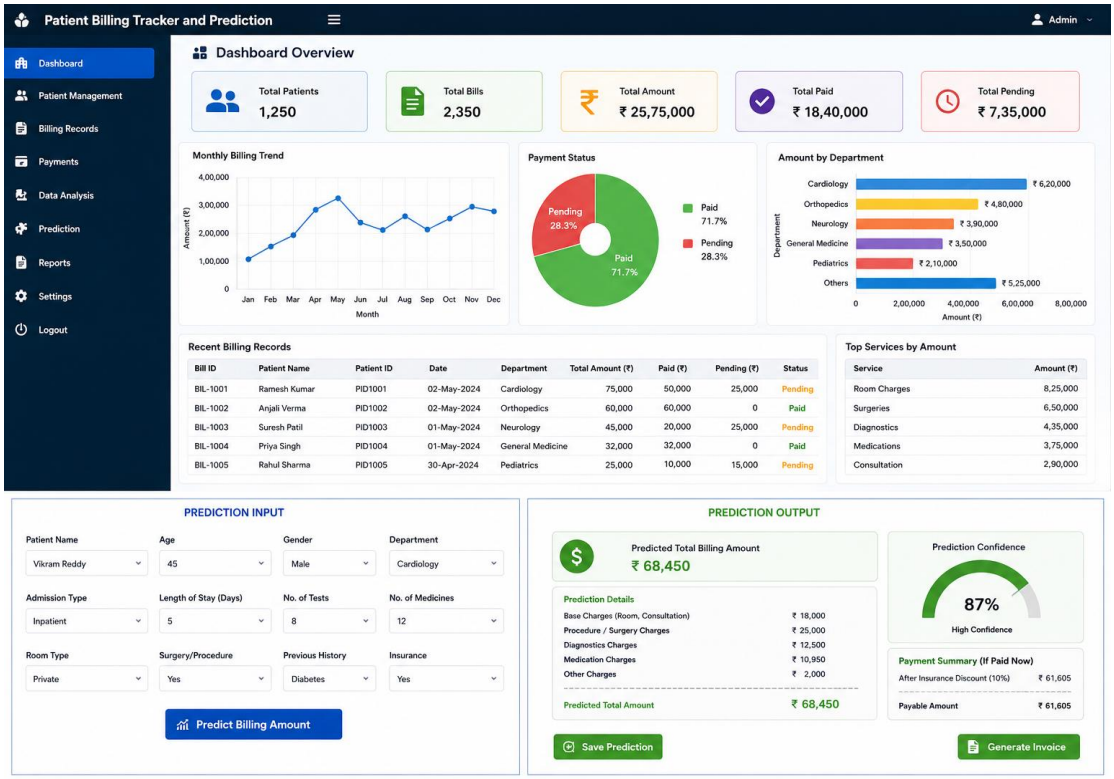
The Patient Billing Tracker and Prediction system architecture represents the complete flow of patient billing data from collection to prediction and reporting. The system starts with the Data Collection Layer, where patient, treatment, service, and payment information is collected from hospital records or datasets. The collected information is transferred to the Data Preprocessing Layer, where missing values, duplicate records, inconsistent data, and categorical attributes are handled. The cleaned data is then passed to the Data Analysis Layer, where billing patterns and healthcare expenses are analyzed.



The Visualization Layer presents billing trends, payment status, department-wise expenses, and other financial information using charts and graphs. The Feature Engineering Layer selects and transforms relevant patient and treatment attributes

required for billing prediction. The processed data is provided to the Machine Learning Layer, where suitable regression models are trained using historical billing records. The trained models are evaluated using metrics such as MAE, MSE, RMSE, and R^2 score to identify an effective prediction model. The selected model is connected to the Billing Prediction Layer, which receives patient and treatment information and estimates the expected billing amount. The Billing Tracking Layer maintains total bills, paid amounts, pending amounts, and payment status for individual patients. The Reporting Layer generates billing summaries, analytical reports, and prediction results for authorized users. Finally, the User Interface Layer allows hospital staff to manage billing records and view analysis and prediction results, completing the flow **Data Collection → Preprocessing → Analysis → Visualization → Feature Engineering → Machine Learning → Model Evaluation → Billing Prediction → Billing Tracking → Reporting → User Interface**.

V. Result and Output



VI. Conclusion

The PATIENT BILLING TRACKER AND PREDICTION system provides an efficient solution for managing, analyzing, and predicting patient billing information. The system maintains important details such as patient information, treatment charges, diagnostic expenses, medicine costs, room charges, total bill amounts, and payment status in a structured manner. Data preprocessing techniques help improve the quality and reliability of billing data by handling missing values, duplicate records, and inconsistent information. Exploratory Data Analysis and visualization provide useful insights into billing trends, treatment costs, department-wise expenses, and payment patterns. Machine-learning techniques can be used to analyze historical billing records

and predict the expected billing amount for new or ongoing patients. Different regression models can be evaluated using MAE, MSE, RMSE, and R^2 score to identify a suitable prediction model. The system also helps track paid and pending amounts, making billing management easier for authorized hospital staff. The dashboard provides a clear representation of billing information through charts, tables, and prediction results. By estimating expected healthcare expenses, the system can support better financial planning and improve billing transparency. It also reduces manual effort and minimizes the possibility of errors in maintaining and analyzing billing records. The system can be further enhanced with online payment integration, automated invoice generation, real-time billing updates, and advanced prediction algorithms. Overall, the project demonstrates how data analytics and machine learning can be combined with billing management to improve healthcare financial tracking, prediction, and decision-making.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.