



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 8, August 2025

A Systematic Review of Artificial Intelligence and Machine Learning in Healthcare Cost Prediction

Dr. S. V. Sonekar¹, Sanket Chaudhari², Sahil Lambat³, Bhumeswar katre⁴, Sushant Bansod⁵

The Principal, JD College of Engineering and Management, Nagpur, Maharashtra, India¹

U.G. Student, Department of Artificial Intelligence & Engineering, JD College of Engineering and Management,
Nagpur, Maharashtra, India²⁻⁵

ABSTRACT: The use of Artificial Intelligence (AI) and Machine Learning (ML) in the health sector has greatly enhanced cost prediction models to the advantage of healthcare providers, insurers, and policymakers in making financial decisions. The conventional cost estimation process is unable to detect the intricate relationships between patient attributes and medical expenditures, hence triggering inefficiencies in financial planning and risk evaluation. This review outlines existing studies on AI-based cost prediction approaches, analysing various supervised learning models used to forecast healthcare insurance costs. Age, body mass index (BMI), gender, smoking status, geographic location, and the number of dependents is established as the main factors that are essential inputs for predictive modelling.

The review also explores the implementation of AI-based cost forecasting in real-time applications, with a focus on web-based tools like Streamlit for interactive user engagement. Backend practices, such as data preprocessing, categorical encoding, numerical scaling, and joblib-based model serialization, are explored to identify the technical innovation in deployment tactics. Although present AI-based models provide better accuracy and efficiency in cost estimation, optimization of ensemble learning methods and deep learning strategies is still problematic for better predictive capabilities. The paper recognizes main gaps in past research and indicates future directions in improving AI-based financial forecasting in healthcare.

KEYWORDS: Artificial Intelligence in Healthcare, Machine Learning, Financial Implications, Systematic Review, Cost Assessment, Economic Feasibility

I. INTRODUCTION

The swift innovations in Machine Learning (ML) and Artificial Intelligence (AI) are transforming the healthcare sector, especially in the areas of financial prediction and cost estimation [1]. Increasing healthcare costs pose a challenge to insurance companies, hospitals, and policy makers, calling for precise prediction models to enhance budgeting and resource allocation [2]. Traditional cost estimation techniques rely on basic statistical models that fail to account for the complex relationships between specific health characteristics and insurance costs [3]. AI-driven models offer a more dynamic and data-centric approach to tackle these challenges [4]. This study investigates the role of AI in predicting healthcare costs by creating a supervised learning model that estimates insurance expenses based on many characteristics, including age, BMI, gender, smoking status, geography, and dependents [5].

The model is included into an intuitive Streamlit web application, enabling users to enter their information and obtain real-time forecasts [6]. Essential machine learning processes, including data pretreatment, categorical encoding, and feature scaling, are conducted to ensure accurate cost estimation [7]. The model is efficiently built with Joblib to provide seamless integration and performance [8]. Unlike traditional cost estimation, AI-driven predictive modelling facilitates a deeper comprehension of the expenses associated with individual healthcare. This study demonstrates how AI may enhance cost transparency and financial planning for healthcare consumers through user-friendly interfaces and real-time forecasts [10]. Although the model does not yet apply deep learning or complex ensemble techniques, it provides a platform for future development, such as incorporating further health data and regulatory requirements [11]. The growing use of AI in financial forecasting within healthcare points to its promise to drive efficiency and cost-effectiveness [12]. Data privacy, regulatory compliance, and interpretability are among the challenges that need to be met to facilitate ethical and secure adoption of AI [13]. This project is a step towards an intelligent, data-driven method

of healthcare cost prediction, allowing for more equitable pricing and informed financial decision-making [14].

II. RELATED WORK

The referenced paper, titled "Adaptive traffic signal control system using composite reward architecture based deep reinforcement learning," focuses on optimizing traffic signal control using deep reinforcement learning (DRL). This paper introduces an adaptive traffic signal control (ATSC) system that leverages real-time traffic information to control traffic signals, thereby improving traffic flow.

The proposed system employs RFID technology and Arduino microcontrollers, whereas the referenced paper focuses on deep reinforcement learning and composite reward architecture. The proposed system aims to incentivize drivers through a reward-based approach, offering points redeemable for discounts. In contrast, the referenced paper emphasizes optimizing traffic signal control using a reward architecture. The proposed system seeks to encourage compliance with traffic signals and promote responsible driving behavior. In contrast, the referenced paper aims to optimize traffic signal timing and coordination to enhance traffic flow[1].

Over the past decade, significant advancements have been made in intelligent traffic control systems aimed at reducing traffic congestion and supporting the development of smart cities. A key innovation in this area is edge computing, which enables real-time data exchange between vehicles and roadside units to help alleviate congestion. This study presents a novel edge- computing-based deep reinforcement learning (DRL) system that uses a carefully designed, multi-objective reward function to optimize various traffic management goals simultaneously. [2]

In the past decade, intelligent traffic control technologies have made significant strides in addressing the ongoing challenge of traffic congestion, which poses a major issue for smart cities. One notable advancement in this field is edge computing, which enables real-time data transmission between vehicles and roadside units, effectively helping to manage and reduce congestion. This study introduces a novel deep reinforcement learning (DRL) system powered by edge computing, designed to optimize traffic flow through a multi- objective reward function. The system overcomes the challenge of evaluating actions with simple numerical rewards by carefully selecting reward functions that guide the agent to learn the ideal behavior for managing multiple traffic signals in a large-scale road network.[3]

A *Reward-Based Adaptive Traffic Light System* is an innovative approach to traffic management that uses real-time data and adaptive algorithms to optimize traffic flow at intersections. Traditional traffic lights often operate on fixed timers or basic sensors, which can lead to inefficiencies such as long waiting times, congestion, and wasted fuel. In contrast, a reward-based adaptive system leverages techniques from reinforcement learning, where the system "rewards" certain actions—like minimizing wait times or reducing stops—to achieve optimal outcomes. In this setup, traffic lights "learn" to make better decisions by adjusting signal timings based on the current traffic conditions and the outcomes of previous actions.[4]

Gamification of Traffic Management introduces a novel approach to alleviating traffic congestion and promoting safer, eco-friendly driving habits by incentivizing drivers through a rewards system. Unlike traditional traffic management, which relies solely on infrastructural changes and enforcement, gamification taps into drivers' intrinsic motivation by offering rewards for positive driving behaviors, such as obeying speed limits, avoiding aggressive acceleration, or choosing less congested routes.

In this system, drivers can earn points or rewards (such as discounts, coupons, or digital badges) for behaviors that align with traffic management goals, like reducing congestion or lowering emissions. For example, a driver who consistently maintains safe speeds or chooses less crowded roads during peak hours could accumulate points redeemable for real-world rewards.[5]

An **Intelligent Traffic Control System Using Reinforcement Learning** represents a cutting-edge approach to modernizing traffic management by enabling traffic lights to make real-time decisions based on continuously evolving traffic conditions. Traditional traffic control systems, which often rely on fixed timings or basic sensor data, lack the adaptability needed to handle varying traffic volumes and unpredictable congestion patterns effectively. Reinforcement learning (RL), a type of machine learning, offers a solution by allowing the traffic control system to "learn" optimal signal timings through trial and error..[6]

A **Blockchain-Based Traffic Reward System for Smart Cities** is an innovative solution that combines blockchain technology and traffic management to incentivize drivers for positive behaviors that improve urban mobility. With the rapid growth of urban populations, smart cities are looking for ways to address congestion, pollution, and road safety. Traditional traffic management solutions often rely on fines or infrastructure changes, but a blockchain-based reward system offers an alternative by using incentives to encourage drivers to make traffic-friendly choices..[7]

Incentive-based Traffic Management Using Mobile Applications is a modern approach to improving traffic flow and encouraging safer, eco-friendly driving by rewarding drivers directly through mobile apps. Unlike traditional traffic systems, which often rely on infrastructure changes or penalties for enforcement, this approach uses mobile technology to incentivize drivers for positive behaviors such as avoiding congested routes, obeying speed limits, or using eco-friendly driving techniques. In this system, a mobile app tracks drivers' behaviors in real-time, offering rewards like points, discounts, or special privileges for actions that align with broader traffic management goals. For instance, a driver who takes an alternative route to avoid high-traffic zones during peak hours may earn points redeemable for rewards, such as discounted fuel, toll credits, or vouchers. Mobile apps also enable cities to communicate real-time traffic conditions, alternative routes, and other relevant data, making it easier for drivers to make informed choices.[8]

Improving Traffic Signal Efficiency through a Reward-Penalty Framework introduces an innovative approach to optimizing traffic signal timings by incorporating a system of rewards and penalties. Traditional traffic signal systems often operate on pre- set timings or basic detection sensors, which may not adapt effectively to fluctuating traffic conditions. In contrast, a reward-penalty framework uses principles from reinforcement learning to enhance signal efficiency by continuously adapting to real-time traffic flow and patterns..[9]

Reward-Based Traffic Management with Real-Time Feedback is an innovative approach that uses rewards and immediate feedback to encourage drivers to adopt behaviors that improve traffic flow, reduce congestion, and promote road safety. Traditional traffic management relies heavily on passive enforcement methods and infrastructure changes, but a reward-based system takes a proactive approach by offering incentives for drivers who make traffic-friendly choices in real-time.[10]

Reinforcement Learning for Traffic Signal Control with a Focus on Driver Behavior is an advanced approach to traffic management that leverages machine learning to optimize signal timings based not only on traffic flow but also on the real-time behaviors of drivers at intersections. Traditional traffic signal systems are typically designed to follow fixed schedules or respond to basic sensor data, which may not be sufficient to handle the complex, dynamic interactions between vehicles. Reinforcement learning (RL), however, enables traffic signals to "learn" from experience, making it possible to adapt dynamically to changes in traffic patterns..[11]

Crowd-sourced Traffic Monitoring with Reward-Based Participation is an innovative approach to gathering real-time traffic data by incentivizing everyday drivers to contribute information about road conditions, traffic flow, and incidents. Traditional traffic monitoring systems often rely on costly infrastructure, like cameras and sensors, which may not provide full coverage of all areas. Crowd-sourcing offers a flexible, scalable alternative by allowing drivers to report and share live traffic updates through mobile apps. [12]

III. RESEARCH METHODOLOGY /TOOLS AND TECHNIQUES

A. Dataset Overview

Data used in this research consist of key patient characteristics affecting medical spending such as age, BMI, gender, current smoker status, geography, number of dependents, and respective medical spending. The data is applied in developing a cost estimation model using AI.

Preprocessing Steps:

- **Missing Data Treatment:** For ensuring data integrity and avoiding biases, missing values were treated with proper imputation methods like mean/mode replacement or predictive modeling.
- **Encoding Categorical:** Gender, smoking, and region categorical features were encoded as numeric values through one-hot encoding to increase model explainability and compatibility.
- **Feature Scaling:** We normalized continuous features such as age, body mass index (BMI), and health expenditure using min-max scaling or z-score normalization to stabilize the model and make it converge more quickly.

B. Model Selection

A supervised learning technique was employed to develop a robust cost estimation model through the help of advanced regression models to determine the non-linear associations between patient demographics and their corresponding medical expenses.

Key Processing Techniques:

- **Feature Engineering:** The most powerful features were extracted and optimized to improve predictive performance and, consequently, utilize only the correct attributes to make decisions for the model..
- **Encoding and Scaling:** All the categorical variables were encoded to make them machine learning algorithm-compatible, and numerical variables were scaled to ensure consistency of variables.
- **Model Tuning:** Hyperparameter tuning and cross-validation were used to fine-tune the model to generalize well and prevent overfitting.

C. Web Application Implementation

To show cost estimation in real time and make it more user-friendly, a web application was created using Streamlit. The website is distinguished by its interactive interface, where the customers can provide their details and their estimated cost of medical bills can be forecasted to them within seconds.

Implementation and Features

Joblib use for model deployment: The trained model was serialized and deployed via Joblib for improving processing time and lowering latency, thus making real-time prediction possible.

- **User-Friendly Interface:** The app has incorporated interactive components such as sliders, dropdowns, and input fields to allow simple input of suitable personal data by users.
- **Real-Time Cost Estimating:** The program reacts to user input in real time and produces real-time, data-driven estimates, enhancing end-user usability and accessibility.
- **Scalability and Performance:** The web application has been designed in a way that it can handle large numbers of user interactions efficiently, providing a responsive and seamless interface and delivering accuracy and computational effectiveness.

This study design provides data-driven, systematic costing of healthcare by combining AI-based analytics with an interactive, user-friendly application providing functional usability as well as technological effectiveness.

IV. CONCLUSION

1. Age raises medical costs disproportionately, with those aged 56+ having the greatest costs.
2. Obesity is the largest cost factor, with obese persons paying more than overweight or normal-weight persons.
3. Smoking has the greatest cost, with smokers paying considerably more in medical costs than non-smokers.
4. Men usually have more cases than women, but the difference is quite minimal.
5. There are also regional variations, with the Southeast being the most expensive and the Southwest being the least expensive.
6. Greater family sizes have greater medical costs, with families with 3+ children paying the highest. High-BMI non-smokers are also costly, suggesting the unique contribution of obesity to the cost of health
7. The BMI-age relationship is weight gain with age, rising health expenditure with rising age.
8. Healthcare expenditures of the 40+ age group significantly rise, particularly in the obese, because of the synergistic effect of the diseases brought about by elevated age and high BMI.
9. We established a predictive analytics dashboard with the supervised machine learning model trained across gender,

BMI, age, smoker, region, and dependents to be able to approximate predict insurance premiums and spend on healthcare.

REFERENCES

1. T. Takeshima, S. Keino, R. Aoki, T. Matsui, and K. Iwasaki, "Development of Medical Cost Prediction Model Based on Statistical Machine Learning Using Health Insurance Claims Data," *Value in Health*, vol. 21, pp. S97, 2018.
2. C. Yang, C. Delcher, E. Shenkman, and S. Ranka, "Machine Learning Approaches for Predicting High-Cost High-Need Patient Expenditures in Health Care," *BioMed Eng OnLine*, vol. 17, no. 1, p. 131, 2018.
3. R. Rozenblum et al., "Using a Machine Learning System to Identify and Prevent Medication Prescribing Errors: A Clinical and Cost Analysis Evaluation," *The Joint Commission Journal on Quality and Patient Safety*, vol. 45, no. 9, pp. 1-8, 2019.
4. H. M. Alzoubi, U. Malik, N. Sahawneh, A. Majid, A. Q. AlHamad, and A. Atta, "Analysis of Cost Prediction in Medical Insurance Using Modern Regression Models," *IEEE Xplore*, 2022.
5. Y. Huang, S. Li, M. Chen, and T.-S. Lee, "The Prediction Model of Medical Expenditure Applying Machine Learning Algorithm in CABG Patients," *Healthcare*, vol. 9, no. 6, p. 710, 2021.
6. D. Y. Lin, "Regression Analysis of Incomplete Medical Cost Data," *Statistics in Medicine*, vol. 22, no. 8, pp. 1181-1200, 2003.
7. R. Rao, R. Jain, M. Singh, and R. Garg, "Machine Learning Models for Patient Medical Cost Prediction and Trend Analysis Using Open Healthcare Data," *IEEE International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*, 2023.
8. T. R. Juday, "Cost-Effectiveness Analysis: Problems and Promise for Evaluating Medical Technology," *Project HOPE Center for Health Affairs*, 1994.
9. V. Lin, "The Economic Impact of AI in Healthcare: Cost Prediction and Resource Optimization," *International Journal of Healthcare Analytics*, vol. 5, no. 3, pp. 45-62, 2022.
10. J. Brown, M. Taylor, and L. Wilson, "AI-Driven Predictive Analytics for Healthcare Cost Estimation: Challenges and Opportunities," *Journal of Medical Informatics and Decision Science*, vol. 10, no. 2, pp. 87-102, 2023.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details