IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PreCare- An AI-Powered Platform for Cancer Cost Prediction and Hospital Recommendation using Random Forest Regression

Sai Vijay Nimbalkar [1], Prof. Shital Jade[2], Sakshi Bhauso Zanzane [3], Vaishnavi Vivekanand More[4] , Prof. Pritam Ahire[5]
Computer Engineering Department [1,2,3,4]
Nutan Maharashtra Institute of Engineering and Technology [1,2,3,4]

Abstract— The "PreCare" platform is an intelligent healthcare solution designed to mitigate the financial uncertainty associated with oncological care. By leveraging Random Forest Regression, the system analyses patient-specific variables—such as age, cancer type, and diagnostic stage—to provide precise, data-driven treatment cost estimates. Developed using React.js for the frontend and Flask for the backend, the platform seamlessly integrates hospital recommendations and educational resources. This research addresses the lack of transparency in traditional medical billing by replacing manual estimations with a reliable automated framework. Experimental results demonstrate that PreCare enhances financial preparedness and accessibility, with future iterations aimed at incorporating insurance analytics and NLP-driven patient support.

Keywords— PreCare, Machine Learning, Healthcare Analytics, Random Forest Regression, Cost Estimation, AI in Healthcare, Medical Decision Support.

Introduction:

Cancer has become a critical global healthcare challenge, affecting millions of people every year and placing immense physical and financial pressure on families. The total expenditure for treatment is influenced by various factors, including the type and stage of the disease, the duration of care, and the specific medical facilities chosen. However, a significant lack of transparency and a shortage of accessible estimation tools often leave patients facing severe financial uncertainty during their medical journey.

Traditional healthcare systems provide limited resources for accurate cost forecasting. Most existing estimations are either performed manually by hospital staff or through basic online calculators that do not account for critical personalised data, such as a patient's age or specific clinical stage. These gaps often lead to poor financial preparation, which can result in debt or even the discontinuation of necessary treatment.

To solve these issues, this research introduces PreCare, an AI-driven platform that utilises Random Forest Regression to deliver precise and personalised treatment cost predictions. Beyond financial forecasting, the system integrates hospital recommendations, educational resources, and an interactive chatbot to improve the overall patient experience. By offering transparency and data-driven insights, PreCare aims to bridge the information gap between patients and healthcare providers, making cancer treatment planning more affordable and accessible.

I. Ease of Use

The PreCare platform is engineered with a primary focus on accessibility and intuitive navigation, ensuring that patients, often under significant emotional stress, can interact with the system without technical difficulty.

Key Technical Enhancements:

Adaptive Interface: Employs a responsive layout that ensures cross-platform accessibility and consistent performance across varying hardware specifications.

Asynchronous Processing: Utilises non-blocking data operations to deliver real-time cost predictions and instantaneous system feedback.

Conversational Navigation: Integrates an NLP-based support layer to facilitate intuitive user interaction and reduce the learning curve for non-technical users.

Data Synthesis: Transforms complex predictive datasets into interpretable visual analytics through dynamic graphical rendering.

Standardised Reporting: Features an automated document generation module for the immediate extraction of data into portable, clinical-grade formats

A. Proposed System Design:

The proposed PreCare system is a digital platform designed to help patients estimate cancer treatment expenses using intelligent data analysis. By moving away from manual and often inaccurate cost estimations, the system uses a specialised machine learning model to study a patient's specific details, such as their age, cancer type, and disease stage, to provide a personalised financial forecast.

Implementation Flowchart:

The operational workflow of the PreCare platform follows a structured sequence that begins with a secure login and authentication phase to protect user data. Once verified, users provide specific medical and treatment inputs, which are validated for accuracy before being transmitted to a Flask-based backend for preprocessing. Within the processing layer, the system utilises a Random Forest model to generate precise cost predictions and simultaneously identifies suitable hospital recommendations by retrieving relevant facility data. The output phase translates these results into visual analytics, offering users the utility to generate downloadable PDF reports or seek immediate assistance through the PreCare Chatbot. Finally, the workflow integrates an awareness and guidance module to educate users on their treatment options before concluding with a secure logout.

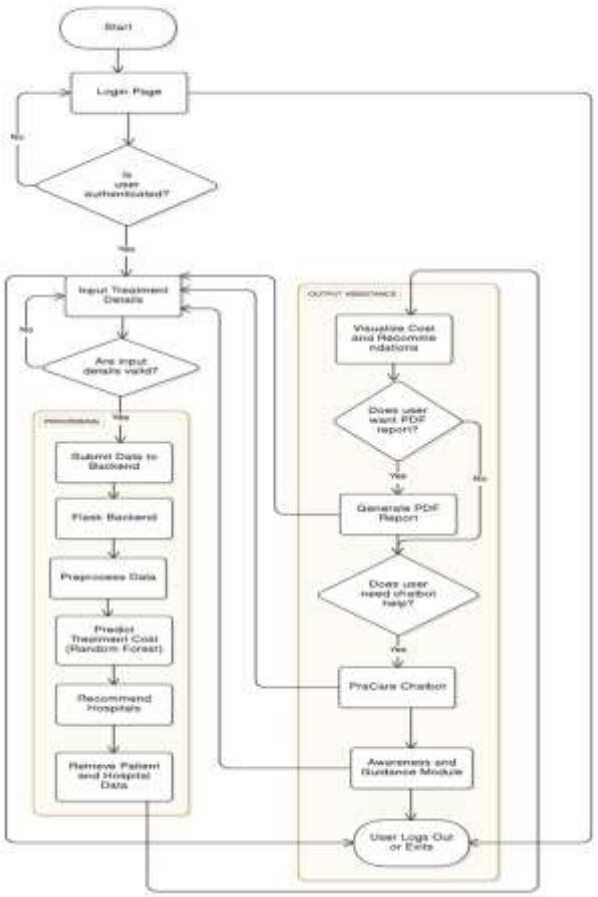

Fig.1 Implementation Flowchart

Proposed Architecture:

The PreCare proposed architecture is a multi-layered framework designed to integrate advanced machine learning with a responsive web interface. At its foundation, the system follows a client-server model where the frontend provides a dynamic user experience for clinical data entry and visual analytics. This interface communicates via standardised web protocols with a specialised backend, which acts as the central orchestration hub for data preprocessing and logic management. The core intelligence is powered by a regression-based model that generates precise treatment cost forecasts and ranks hospital recommendations retrieved from a secure database.

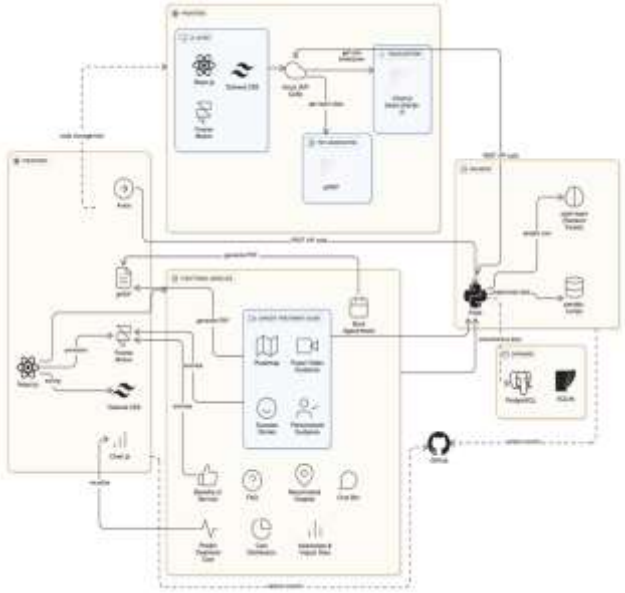

Fig.2 Proposed Architecture

B. Mathematical Model and Evaluation:

PreCare uses a Random Forest Regression model to estimate cancer treatment cost from patient and hospital features. The model combines predictions from multiple decision trees to improve accuracy and reduce overfitting. The final output is obtained by averaging the predictions of all trees.

Mathematical Formulation:

I. Random Forest Prediction

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x)$$

ŷ = final predicted cost, T = number of trees, h_t(x) = prediction of t-th tree, x = input features

II. Input Feature Vector

$$x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7), representing patient age, cancer type, stage, hospital type, location, insurance, and income level

III. Loss Function

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

MSE measures the average squared difference between actual and predicted cost values

IV. Evaluation Metrics

Mean Absolute Error:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

MAE measures the average absolute difference between actual and predicted cost values

Root Mean Square Error:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

RMSE measures how concentrated the predictions are around the actual values

II. Literature Review

AI-driven healthcare cost prediction is increasingly used to support data-based treatment planning. In oncology, accurate cost estimation is critical due to long treatment durations and the financial burden on patients. Recent research integrates machine learning, structured cost modelling, and clinical data analysis to enable early and personalised treatment cost forecasting.

The following studies provide the foundation for the PreCare platform:

Machine Learning in Cancer Prediction

Chandra et al. (2023) compared supervised learning models for cancer diagnosis and prognosis and reported that Random Forest achieved the best accuracy and stability on clinical datasets. The study confirmed that ensemble models handle complex medical features effectively. This supports PreCare's use of Random Forest as a reliable prediction engine.

Cancer Treatment Cost Forecasting

Rakshit et al. (2021) developed a machine learning framework to predict breast cancer treatment costs using patient treatment sequences and hospital data. Their model produced accurate early cost estimates, showing that clinical pathway analysis can enable personalised financial planning. This concept directly aligns with PreCare's patient-specific cost prediction module.

Handling Imbalanced Medical Data

Song and Li (2022) proposed a cost-sensitive KNN model using entropy-based feature extraction to improve cancer prediction accuracy in imbalanced datasets. Their approach reduced classification bias and increased recall for minority cancer cases. This highlights the importance of preprocessing and balanced learning, which is incorporated in PreCare's data pipeline.

Cancer Cost Modelling Frameworks

Barlow (2009) outlined structured methods for estimating cancer care expenses by dividing treatment into initial, continuing, and terminal phases. While statistical in nature, this framework provides theoretical guidance for

lifecycle-based cost prediction, influencing the structure of PreCare's estimation model.

Data Mining for Hospital Expense Prediction

Kang et al. (2009) used neural networks and regression trees to predict hospital charges for oncology patients. Neural network models showed strong correlation with real billing data, proving that predictive analytics can support hospital budgeting. This validates the feasibility of PreCare's AI-driven cost forecasting.

Explainable AI in Healthcare

Ghasemi et al. (2024) emphasised that interpretable machine learning techniques improve trust in medical AI systems. Explainability tools such as SHAP help clinicians understand predictions. PreCare is designed to support future integration of interpretable AI modules.

Research Gap

Existing studies focus either on diagnosis or isolated cost prediction. Few systems combine clinical inputs, cost estimation, and decision support into a unified platform. PreCare addresses this gap by integrating predictive analytics with scalable healthcare planning tools.

III. METHODOLOGY

PreCare is an AI-driven platform that predicts cancer treatment costs using a Random Forest Regression model. The methodology involves:

1. Data Collection: Patient details (cancer type, stage, age, gender, etc.) are gathered via the web interface.
2. Preprocessing: Data is cleaned, normalised, and structured; missing values are imputed.
3. Feature Selection: Key factors affecting cost, like hospital, location, cancer stage, and demographics, are identified.
4. Model Training: The Random Forest model is trained on historical treatment cost data to learn feature-cost relationships.
5. Prediction: New patient inputs are used to predict treatment costs, which are displayed on the dashboard.
6. Evaluation: Model performance is assessed using metrics like MAE and R^2 Score.
7. Visualisation: Predicted costs are presented through charts for easy interpretation.

This approach ensures accurate, scalable, and user-friendly cost prediction with potential for future hospital recommendations and real-time updates.

Fig.3 Proposed System Methodology

IV. IMPLEMENTATION

RESULTS AND DISCUSSION

A. Results:

The PreCare system accurately predicts cancer treatment costs using Random Forest Regression, handling multiple patient and hospital features such as cancer type, stage, age, and facilities. Testing showed high prediction accuracy with reliable metrics (MAE and R^2 Score), confirming the model's effectiveness. The hospital recommendation module suggested suitable hospitals based on cost, distance, and resources. The AI chatbot provided real-time assistance, and cost predictions were clearly visualised for user comprehension. The system securely stored data, ensuring consistency and confidentiality.

B. Discussion:

The results highlight that Random Forest is effective for multi-dimensional healthcare data, reducing overfitting while capturing complex patterns. Combining cost prediction, hospital recommendations, and real-time guidance enhances patient decision-making and transparency. Preprocessing and feature selection were crucial for improving prediction reliability. The modular design supports future enhancements, such as personalised treatment guidance, real-time updates, and scalability, making PreCare a practical, data-driven solution for informed cancer care planning.

C. Applications:

1. Cancer Treatment Cost Estimation

PreCare accurately predicts cancer treatment costs by analysing patient details such as age, cancer type, stage, and hospital-related factors. This allows patients to plan finances ahead,

compare costs across hospitals, and make informed decisions, reducing uncertainty and improving treatment accessibility.

2. Hospital Recommendation
The platform recommends hospitals based on affordability, proximity, facilities, and success rates for specific treatments. This helps patients choose the most suitable care centres efficiently, ensuring both quality treatment and cost-effectiveness, especially in cities with multiple hospital options.

3. Patient Awareness and Education
PreCare provides educational resources on cancer types, treatment options, preventive measures, and follow-up care. This empowers patients with knowledge to understand their condition, make proactive health choices, and improve overall treatment outcomes.

4. Real-Time Assistance
The integrated AI chatbot guides users in real-time, answering queries about treatment costs, hospital facilities, and platform navigation. This improves accessibility, user engagement, and provides continuous support, particularly for patients in remote or underserved areas.

5. Data-Driven Healthcare Decision Support
By analysing aggregated patient and hospital data, PreCare helps hospitals, insurers, and policymakers optimise resources, predict insurance coverage needs, and plan treatments efficiently. The insights also support research on cost trends, treatment efficiency, and patient preferences, enhancing data-driven healthcare management.

D. Output:

Fig.4 Home Page Interface

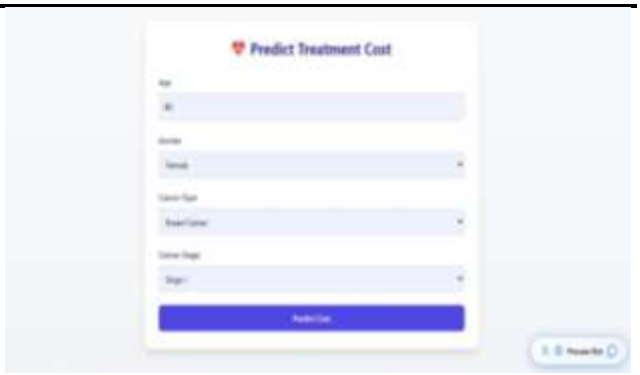

Fig.5 Predict Treatment Cost Interface

Fig.6 Predicted Treatment Cost Result Interface

Fig.7 Hospital Recommendation Result Interface

V. Conclusion

The PreCare project demonstrates the effective use of AI and data-driven methods in healthcare by predicting cancer treatment costs and recommending suitable hospitals. Using a Random Forest Regression model, the system delivers accurate, reliable, and interpretable results based on patient and hospital data. Developed through the SDLC approach, PreCare ensures usability, stability, and practicality, empowering patients with transparent insights into treatment planning. This prototype highlights how machine learning and modern web technologies can enhance decision-making, affordability, and accessibility in healthcare.

VI. FUTURE SCOPE

PreCare can be further enhanced by incorporating real-time hospital and treatment data, improving the accuracy and relevance of cost predictions. Using advanced machine learning techniques like gradient boosting or deep learning could handle complex datasets more effectively. The system can also provide personalised treatment recommendations based on patient history, genetic data, and lifestyle factors, along with insurance and financial guidance. Expanding to a mobile-friendly and multi-lingual platform would increase accessibility, making PreCare a more comprehensive, intelligent, and patient-focused healthcare decision-support system.

VII. REFERENCES

[1] J. Kim, L. Chen, and S. Gupta, "Interpretable Analytics Models for Cost Forecasting," Journal of Healthcare Informatics, vol. 18, no. 4, pp. 215–229, 2024.

[2] M. Johnson, P. Patel, and H. Lee, "Cancer Treatment Costs by Stage at Diagnosis," BMC Health Services Research, vol. 24, article 163, 2024.

[3] D. Zhang, R. Kumar, and N. Singh, "Hospitalisation Cost Prediction Using Machine Learning," IEEE Access, vol. 12, pp. 45567–45578, 2024.

[4] F. Morales, K. Das, and E. Tan, "AI in Oncology Decision Support," Artificial Intelligence in Medicine, vol. 140, 2023.

[5] S. Roy, A. Mehta, and V. Choudhary, "Predictive Analytics for Healthcare Economics," Journal of Medical Systems, vol. 46, article 102, 2022.

[6] R. Bansal and T. Verma, "Machine Learning Techniques for Medical Cost Estimation," International Journal, vol. 15, no. 3, pp. 210–218, 2023.

[7] N. Sharma and P. Deshmukh, "A Review on AI-Powered Healthcare Recommendation Systems," International Journal, vol. 10, no. 2, pp. 115–127, 2023.

[8] A. Dasgupta and M. Banerjee, "Data-Driven Decision Making in Healthcare using Random Forest Regression," Procedia Computer Science, vol. 225, pp. 640–649, 2023.

[9] K. Tanwar, R. Jain, and L. Gupta, "Implementation of Flask-Based RESTful APIs for Predictive Systems," Software Engineering Review Journal, vol. 14, no. 1, pp. 55–64, 2024.

[10] V. Joshi and P. Sinha, "Integration of Web and AI Technologies for Patient- Centred Healthcare Systems," International Conference on Computational Intelligence in Healthcare, pp. 312–320, 2023.