

# A PRECISION HEALTHCARE APPROACH USING AGENTIC AI FOR EARLY CANCER DETECTION IN PAEDIATRIC PATIENTS BELOW 10 YEARS

CHEMBROLU SATISH<sup>1\*</sup>, Dr.ALUGOLU AVINASH<sup>2</sup>

<sup>1</sup>M.Tech Student,Rno:24A31D5809, Dept. of Computer Science Engineering

<sup>2</sup>Associate Professor, Dept. of Computer Science Engineering  
Pragati Engineering College<sup>1,2</sup>, A.P, India

\*\*\*

**Abstract** - Early detection of pediatric cancer remains a significant healthcare challenge due to non-specific symptoms, limited availability of annotated datasets, and delays in clinical decision-making. This study presents a comprehensive Agentic Artificial Intelligence framework designed for early cancer detection in children below ten years of age. The proposed system integrates multimodal medical data, including clinical records, laboratory reports, and patient history, to enhance diagnostic accuracy. Multiple intelligent agents collaboratively perform key tasks such as data preprocessing, feature extraction, classification, and result interpretation. Several machine learning models, including Support Vector Machine, Logistic Regression, Random Forest, and XGBoost, are implemented and evaluated for performance comparison. Experimental results indicate that the XGBoost model achieves the highest accuracy of 93.68 percent along with an improved F1-score, demonstrating its effectiveness for early diagnosis. The framework also incorporates explainable AI techniques to improve transparency and support clinical trust in automated predictions. By reducing diagnostic delays and enabling timely intervention, the proposed system contributes to improved early-stage cancer detection in paediatric patients. This work highlights the potential of Agentic AI in advancing paediatric oncology diagnostics and supporting precision healthcare solutions.

**Key Words:** Agentic AI, Paediatric Cancer Detection, Machine Learning, XGBoost, Explainable AI, Multimodal Data Analysis, Early Diagnosis, Healthcare AI

## 1. INTRODUCTION

Paediatric cancer detection presents unique challenges compared to adult oncology. Children often exhibit vague and non-specific symptoms such as fatigue, fever, or mild pain, which are frequently misinterpreted as common illnesses. This often results in delayed diagnosis and reduced survival rates<sup>[9]</sup>. With the advancement of healthcare technologies, large volumes of medical data are now available through electronic health records, laboratory reports, and imaging systems. Artificial Intelligence (AI) has emerged as a promising solution for analysing such complex and heterogeneous data to support early diagnosis<sup>[5,7]</sup>. However, existing AI-based systems face several limitations, including the inability to effectively integrate multiple data sources, lack of adaptability, and absence of explainability

<sup>[1,3]</sup>. Most conventional models function as isolated systems and generate predictions without providing clear reasoning, thereby reducing trust<sup>[10,14]</sup> among clinicians. To address these challenges, this research proposes an Agentic AI-based framework in which multiple intelligent agents collaborate to process multimodal data and generate accurate as well as interpretable predictions.

The proposed system aims to enhance early detection accuracy, minimize false positives, and assist clinicians with transparent and reliable decision-making tools.

## 1.1 DATASET DESCRIPTION AND PREPROCESSING

The dataset<sup>[16]</sup> used in this study consists of structured paediatric medical data, including demographic attributes, clinical symptoms, and laboratory measurements. Key features include age, gender, weight, height, hemoglobin levels, white blood cell count, platelet count, and environmental exposure factors. Proper organization of data before analysis is essential; therefore, the content should initially be prepared separately before applying formatting. Text and graphical elements should be handled independently until the final formatting stage. Authors should avoid unnecessary use of hard tabs and excessive line breaks, and no manual pagination should be added, as the template manages formatting automatically.

Data preprocessing plays a critical role in improving the performance and reliability of machine learning models. Several preprocessing steps were implemented to ensure data quality and consistency. Missing values were handled using mean and median imputation techniques. Numerical features were normalized to maintain uniform scale across variables. Categorical attributes were encoded into numerical formats to enable model compatibility. Outliers were detected and removed to prevent skewed predictions. Additionally, feature selection techniques were applied to retain only the most relevant attributes, thereby enhancing model efficiency.

## 2. METHODOLOGY AND ML PIPELINE

The proposed system adopts a systematic machine learning pipeline consisting of multiple stages to ensure accurate and reliable predictions. Initially, patient data is collected from clinical records and user input interfaces. The collected data undergoes preprocessing, which includes data cleaning,

normalization, and feature selection to improve data quality. In the feature engineering stage, significant attributes are extracted to enhance model learning capability. Subsequently, multiple machine learning models such as Support Vector Machine, Logistic Regression, Random Forest, and XGBoost are trained using the processed [8,9] dataset. The trained models are then evaluated based on standard performance metrics to determine their effectiveness. Among the evaluated models, XGBoost demonstrates superior performance and is selected as the final model for deployment. The system then generates real-time predictions based on new input data. Additionally, explainable artificial intelligence techniques are integrated to provide interpretability and reasoning behind the model predictions, thereby improving trust and usability [1,3,13] in clinical settings.

**Evaluation Metrics:**

Accuracy is used to measure the proportion of correctly classified instances among the total observations.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

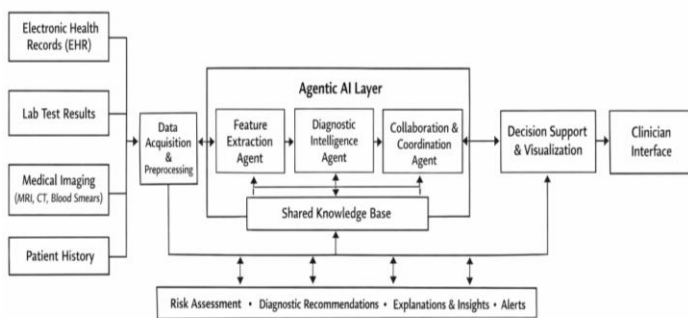
F1 Score is used to balance precision and recall, especially in cases of imbalanced datasets.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

**2.1 SYSTEM DESIGN AND ARCHITECTURE**

The system architecture consists of four main layers:

1. Data Layer – Collects patient data from multiple sources.
2. Preprocessing Layer – Cleans and prepares data.
3. Agent Layer – Includes Data Agent, Feature Agent, Diagnostic Agent, and Explainability Agent.
4. Decision Layer – Generates predictions and recommendations.



**Fig -1:** System Architecture for Agentic AI-Based Paediatric Cancer Detection

**3. CONCLUSIONS**

This section presents the performance evaluation of the proposed machine learning models along with system predictions based on real-time patient data inputs. The comparison of models indicates that XGBoost outperforms other algorithms in terms of both accuracy and F1-score.

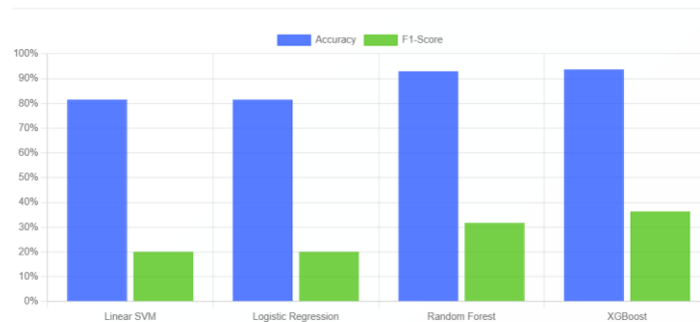
From the experimental results, XGBoost achieves the highest accuracy of **93.68%** and an F1-score of **36.31%**, followed by Random Forest with **92.90% accuracy** and **31.73% F1-score**. Linear SVM and Logistic Regression show comparatively lower performance, both achieving approximately **81.4% accuracy** and **20% F1-score**. This demonstrates that ensemble-based models provide better predictive capability for paediatric cancer detection compared [7,8] to traditional models.

The system also supports real-time prediction using patient medical inputs such as age, gender, laboratory values, and clinical symptoms. Based on the provided sample input, the model predicts “**No Cancer**” with a confidence score of **89.92%**, indicating a low probability of cancer presence. Additionally, Explainable AI (XAI) techniques are integrated into the system to provide interpretability [1,2,4]. The explanation output indicates that the prediction is based on the absence of strong cancer-related indicators in the input data. It also emphasizes that the result is not a definitive diagnosis and recommends further clinical evaluation, thereby supporting responsible and informed medical decision-making.

Overall, the results demonstrate that the proposed Agentic AI framework effectively improves prediction accuracy, supports real-time clinical usage, and enhances transparency through explainable insights [3,10,11].

**4. RESULTS**

**Model Performance Comparison**



The chart compares different ML models. The system automatically selects the best performing model for predictions.

**Fig -2:** Comparison of different ML models

Model	Accuracy	F1-Score
<b>XGBoost</b>	<b>93.68%</b>	<b>36.31%</b>
<b>Random Forest</b>	<b>92.90%</b>	<b>31.73%</b>
<b>Linear SVM</b>	<b>81.49%</b>	<b>20.04%</b>
<b>Logistic Regression</b>	<b>81.47%</b>	<b>20.03%</b>

**Fig -3:** Comparison of different ML models in tabular form

**Patient Medical Data**

Age (Years)	Gender	Weight (kg)
<input type="text" value="3"/>	<input type="text" value="Female"/>	<input type="text" value="12"/>
Height (cm)	Family History of Cancer	Prenatal Radiation Exposure
<input type="text" value="14"/>	<input type="text" value="Yes"/>	<input type="text" value="No"/>
Environmental Toxin Exposure	Persistent Fever	Unexplained Weight Loss
<input type="text" value="Yes"/>	<input type="text" value="No"/>	<input type="text" value="Yes"/>
Bone Pain	Frequent Infections	Bruising or Bleeding
<input type="text" value="Yes"/>	<input type="text" value="No"/>	<input type="text" value="Yes"/>
Fatigue Level	White Blood Cell Count (cells/mcL)	Hemoglobin Level (g/dL)
<input type="text" value="Medium"/>	<input type="text" value="1200"/>	<input type="text" value="5"/>
Platelet Count (cells/mcL)	<input type="text" value="1000"/>	

**Analyze & Predict**

**Fig -4:** Prediction for cancer detection (Run time values)

**Prediction: Cancer**

**Confidence: 62.61%**

XAI: The analysis suggests high indicators associated with cancer. This finding requires immediate clinical validation and further diagnostic testing. Please consult an oncologist as soon as possible.

**Fig -5:** Prediction Output

**REFERENCES**

1) D. Jin, E. Sergeeva, W.-H. Weng, G. Chauhan and P. Szolovits, Explainable Deep Learning in Healthcare: A Methodological Survey from an Attribution View, **arXiv preprint**, Dec. 2021. arXiv

2) B. Dhanalaxmi and P. V. V. S. Srinivas, "Explainable Multimodal Deep Learning in Healthcare: A Survey of Current Approaches," International Research Journal on Advanced Engineering Hub (IRJAEH), vol. 3, no. 03, pp. 1055–1059, Mar. 2025. Irjaeh

3) A. Chaddad, J. Peng, J. Xu and A. Bouridane, "Survey of Explainable AI Techniques in Healthcare," Sensors, vol. 23, no. 2, pp. 634, Jan. 2023. MDPI

4) D. Bhati, F. Neha and M. Amiruzzaman, "A Survey on Explainable Artificial Intelligence (XAI) Techniques for Visualizing Deep Learning Models in Medical Imaging," J. Imaging, vol. 10, no. 10, pp. 239, Sept. 2024. MDPI

5) "Artificial Intelligence in Paediatric Clinical Medicine: Applications, Innovations and Practical Implications," Eur. J. Clin. Med., 2025. European Journal of Clinical Medicine

6) "Artificial intelligence for neuroimaging in paediatric cancer: current state and challenges," PubMed, 2025. PubMed

7) Y. Yang, Y. Zhang and Y. Li, "Artificial intelligence applications in paediatric oncology diagnosis," Explor. Target Antitumor Ther., vol. 4, no. 1, pp. 157–169, Feb. 2023. PMC

8) M. Hassan, S. Shahzadi and A. Kloczkowski, "Harnessing Artificial Intelligence in Paediatric Oncology Diagnosis and Treatment: A Review," Cancers, vol. 17, no. 11, 1828, May 2025. PMC

9) N. N. Elsayid et al., "The Role of Machine Learning Approaches in Paediatric Oncology: A Systematic Review," Cureus, vol. 17, no. 1, e77524, Jan. 2025. PMC

10) "A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges," Informatics Med. Unlocked, vol. 51, 101587, 2024. ScienceDirect

11) S. Nerella, S. Bandyopadhyay, J. Zhang et al., "Transformers in Healthcare: A Survey," **arXiv preprint**, Jul. 2023. arXiv

12) H. Daldrup-Link, "Artificial intelligence applications for paediatric oncology imaging," Pediatr. Radiol., vol. 49, no. 11, pp. 1384–1390, Oct. 2019. PMC

13) "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," **arXiv preprint**, Jul. 2021. arXiv

- 14) "Explainable deep learning models and interpretability in medical AI," PubMed, various authors, ongoing surveys in 2023-24. PubMed
- 15) "Artificial intelligence applications in paediatric oncology," World Neurosurg., highlighted by systematic reviews.
- 16) [https://drive.google.com/file/d/1JZmbSr6IZEF4cZA3k50fje00bLDPJ\\_EB/view?usp=drive\\_link](https://drive.google.com/file/d/1JZmbSr6IZEF4cZA3k50fje00bLDPJ_EB/view?usp=drive_link)

## BIOGRAPHIES



Mr Ch Satish working as Assistant professor in Electronics and communication department, Pragati Engineering college, Surampalem with 16 years of experience, with 3 years of software industry experience at Synergy MultiTech, Hyderabad. He has published 3 international journals and 1 book chapter and 1 UGC journal paper. Attended several workshops and guest lectures, done NPTEL courses in MPMC, Communication skills, PYTHON courses.



Dr. A. Avinash, Associate Professor, CSE department with 15 years of experience in teaching with machine learning -NLP specialization, published over 35 Plus research articles in reputed journals, published 3 patents, 5 book chapters, 3 text books, He's a member of IAENG and ISRD. Reviewer for Springer Nature journal neural computing and applications (NCAA).