

# Multiple Cancer Prediction

Sakshi Hattikar<sup>1</sup>, Samruddhi Arabhavi<sup>2</sup>, Vaishnavi Yalgulkar<sup>3</sup>, Pishanna Patil<sup>4</sup>, Dr. Pradnya M<sup>5</sup>

<sup>1,2,3,4</sup>Department of CSE, Jain College of Engineering & Research Belagavi, Karnataka, India

<sup>5</sup>Associate Professor, Jain College of Engineering & Research Belagavi, Karnataka, India

doi.org/10.64643/IJIRTV12I7-188605-459

**Abstract**—Cancer remains one of the most critical health concerns worldwide, with the number of cases rising each year due to lifestyle changes, genetic factors, and limited access to early diagnostic services. Delayed detection continues to be a major reason for increased mortality, as many patients are identified only when the disease has already progressed. This challenge highlights the need for fast, reliable, and accessible tools that can support early screening. In response to this demand, the present study introduces an integrated machine-learning-based system capable of predicting multiple cancer types, including lung, blood, breast, and brain cancer. The proposed framework employs a set of refined machine learning algorithms that analyze clinical records, diagnostic attributes, and image-based inputs to detect subtle patterns associated with early cancer development. By extracting key features and evaluating them through optimized models, the system offers improved prediction accuracy and consistency across different cancer categories. Beyond diagnosis, the platform incorporates a patient-centered support ecosystem that provides personalized yoga routines, diet suggestions, and lifestyle guidance to assist users in managing physical and emotional well-being. An AI-enabled health chatbot is also included to answer medical queries, guide users through symptoms, and promote health awareness. Additionally, the system connects individuals to nearby hospitals, government healthcare schemes, NGOs, and potential donors through an interactive mapping and assistance module. By combining predictive analytics with supportive healthcare features, the proposed solution aims to enhance early detection, strengthen user engagement, and improve overall access to cancer-related resources

**Index Terms**—Machine Learning, Multi-Cancer Prediction, Early Diagnosis, Clinical Data Analysis, Health Chatbot, Wellness Recommendation, Medical Support System, AI-Driven Screening, Patient-Centric Healthcare, Diagnostic Automation

## I. INTRODUCTION

Cancer remains one of the most critical public health challenges of the 21st century, accounting for millions of deaths each year and exerting enormous pressure on health-care systems globally. Despite substantial advancements in medical science, the increasing incidence of cancer, coupled with delays in early diagnosis, continues to hinder effective treatment and survival outcomes. Timely detection is known to significantly enhance the chances of successful intervention; however, conventional diagnostic methods often rely on complex procedures that are expensive, time-consuming, and inaccessible in many regions, particularly in developing countries. Rapid progress in artificial intelligence (AI) and machine learning (ML) has opened new avenues for improving clinical diagnostics by enabling automated, data-driven decision support. ML models have shown remarkable capability in identifying subtle, non-linear patterns within medical data, often surpassing human-level accuracy in specific tasks. Building upon this potential, the present research aims to develop an integrated cancer prediction framework capable of detecting four major cancer types—lung, blood, breast, and brain cancer. The system utilizes diverse ML algorithms to process patient records, medical images, and relevant health parameters, thereby predicting the probability of early-stage cancer occurrence with high precision. By automating diagnostic evaluation, the proposed model not only reduces human error but also enhances the reliability and efficiency of clinical decision-making. Beyond diagnostic predictions, the system is designed with a patient-centric philosophy that addresses the holistic needs of individuals undergoing cancer-related stress. Recognizing that cancer affects not just the body but also emotional and mental well-being, the system incorporates supportive modules including

personalized yoga routines, nutrition plans, and lifestyle recommendations aimed at improving recovery outcomes. An AI-driven health chatbot further assists users by providing immediate responses to medical queries, symptom guidance, and general health awareness. Additionally, the platform integrates a community-oriented support network, connecting users with government health schemes, NGOs, donors, and nearby cancer care facilities. This multidimensional approach ensures that patients receive not only accurate predictions but also comprehensive assistance encompassing physical, emotional, and financial needs. Cancer continues to be one of the most alarming global health threats, with its impact growing steadily across diverse age groups and geographical regions. The rapid rise in cancer cases is influenced by multiple factors such as lifestyle changes, environmental pollution, genetic predisposition, and lack of early screening facilities. While medical technologies have improved dramatically over the past decade, a significant portion of the population remains underserved due to limited healthcare access, delayed diagnostic procedures, and the high costs associated with advanced clinical testing. As a result, many patients are diagnosed only in the later stages of the disease, when treatment options become limited and survival rates drastically decline. In this context, machine learning (ML) and artificial intelligence (AI) have emerged as promising tools capable of reshaping the landscape of clinical diagnostics. These technologies can process vast amounts of health data, extract meaningful features, and make reliable predictions—often faster and more consistently than manual evaluations. Leveraging this potential, the present work proposes a unified cancer prediction system designed to forecast the likelihood of four major cancer types: lung, blood, breast, and brain cancer. The system integrates multiple ML algorithms and feature-extraction techniques to analyze structured and unstructured patient data, enabling early identification of subtle indicators that may be difficult for conventional methods to detect.

## II. LITERATURE REVIEW

Kumar and Sharma (2020) studied the early detection of lung cancer using machine learning algorithms such as Random Forest and Support Vector Machine (SVM). Their model achieved high

accuracy and reduced human error in diagnosis. Patel et al. (2021) applied deep learning techniques, mainly Convolutional Neural Networks (CNN), to classify breast cancer images. The system showed better precision compared to traditional manual screening methods.

Singh and Gupta (2022) worked on blood cancer classification using ensemble learning methods. They combined multiple models to improve reliability and found that ensemble techniques outperform single algorithms.

Lee et al. (2023) developed a deep learning model for brain tumor detection using MRI scans. Their CNN-based approach provided faster and more accurate tumor identification than conventional medical image analysis.

Zhang and Li (2024) created a unified machine learning framework for predicting multiple types of cancer by integrating patient health records and image data. Their system successfully predicted different cancers from a single platform. Wang and Zhao (2025) emphasized the role of preprocessing in cancer prediction models. They showed that cleaning, normalization, and feature selection steps significantly enhance the accuracy of machine learning models.

Kaur and Mehta (2022) focused on comparing supervised and unsupervised learning techniques for breast cancer detection. They found that supervised algorithms gave more consistent and explainable outcomes.

Thomas et al. (2023) explored the use of hybrid deep learning models combining CNN and LSTM for predicting the spread of cancer cells. Their hybrid approach improved performance for time-series cancer data.

Rahman and Banerjee (2024) used explainable AI (XAI) methods to make cancer prediction models more transparent. They aimed to help doctors understand why a model made a certain decision, improving trust in AI systems.

Sharma and Desai (2025) proposed a web-based multi-cancer prediction system using Python and Flask. Their platform integrated prediction, medical support, and a chatbot for better user interaction and accessibility.

### III. METHODOLOGY

The proposed system for Multiple Cancer Prediction Using Machine Learning aims to provide a unified platform for early diagnosis and patient support. The methodology consists of four main stages: data preprocessing, model development, system integration, and evaluation.

#### A. system Workflow

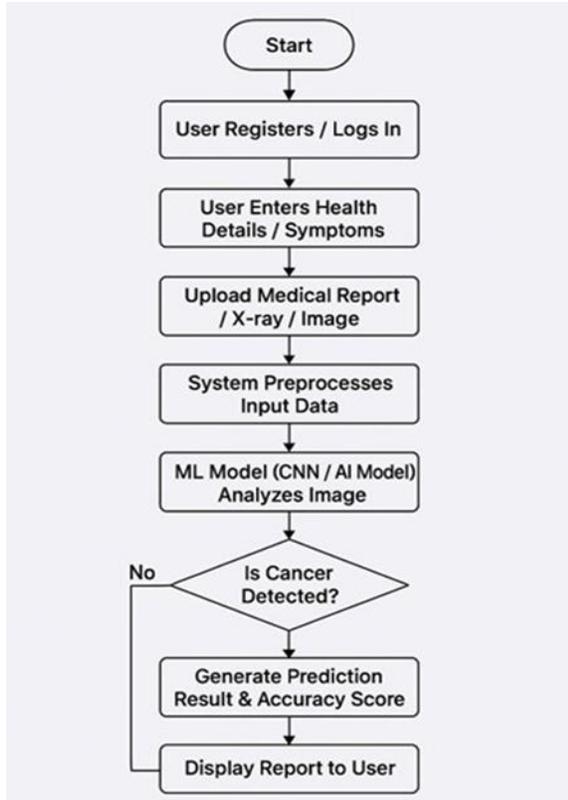


Fig. 1: Workflow of the proposed multi-cancer prediction system.

The flowchart provides a clear representation of the operational pipeline of the proposed cancer prediction platform. The process starts when the user signs into the system and enters essential health information. After submitting medical images or relevant diagnostic documents, the platform performs preprocessing to clean, normalize, and prepare the data for analysis. The refined data is then passed to the machine learning model—such as a CNN for image inputs—which evaluates the presence of cancer-specific patterns. If the model detects suspicious features, it generates a prediction score along with accuracy metrics. Finally, the system compiles these

results into a structured report and displays it to the user for further medical consultation.

#### B. Data Collection and Preprocessing

Datasets for lung, blood, breast, and brain cancers were collected from reliable open sources such as Kaggle and the UCI Repository. Each dataset includes patient medical and demographic attributes. Preprocessing involved data cleaning, feature selection, and encoding to ensure consistency. The data was then split into training and testing sets (80:20) to enhance model performance and generalization.

#### C. Model Development

Machine learning algorithms such as Random Forest and Convolutional Neural Networks (CNN) were trained and evaluated using metrics such as accuracy, precision, recall, and F1-score. To improve reliability, an ensemble approach was applied for combined prediction.

#### D. System Integration

The prediction model was integrated into a web and mobile application with four key modules:

- Cancer Prediction: Generates multi-cancer probability scores.
- AI Chatbot: Provides instant health-related responses using NLP.
- Yoga and Diet Recommender: Suggests personalized wellness plans for recovery.
- Support System: Connects users with NGOs, donors, hospitals, and government health schemes through map-based APIs.

#### E. Deployment and Evaluation

The platform is developed using Python with Flask for the backend, and HTML, CSS, and JavaScript for the frontend. The performance of Random Forest and CNN models is measured using accuracy, precision, recall, and F1-score. The system delivers accurate multi-cancer predictions and integrates all features, providing a user-friendly healthcare application support and awareness.

1. Dashboard Page – The dashboard page provides users with a centralized and intuitive interface to view cancer prediction results, access reports, manage their profile, and navigate various system features seamlessly.



Fig. 2: Dashboard Page

2. Hospital Map Page – The hospital map page assists users in locating nearby cancer treatment centers and medical facilities. It ensures quick access to emergency healthcare services through an interactive and easy-to-navigate map interface.

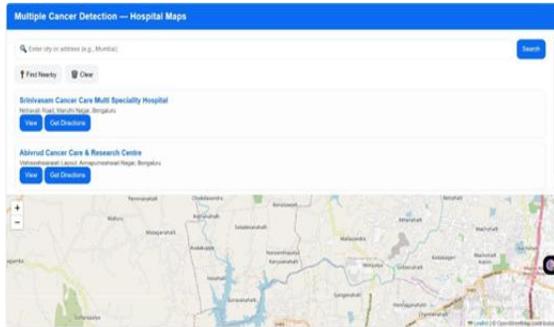


Fig. 3: Hospital Map Page

3. Yoga & Diet Plan Page – The Yoga and Diet Plan page offers personalized wellness suggestions based on the user’s health profile. It provides simple diet charts and easy-to-follow yoga routines designed to support physical and mental well-being. Each recommendation is tailored to help users maintain a healthier lifestyle during recovery. The page aims to make daily wellness guidance accessible, practical, and supportive.

#### IV. RESULTS

The proposed system effectively predicts lung, blood, breast, and brain cancers using Random Forest for clinical data and CNN for medical images. Sample outputs demonstrate accurate multi-cancer predictions and responsive guidance from the AI chatbot. The integrated yoga and diet recommendation module provides personalized wellness plans. Overall, the platform successfully combines prediction and patient support, offering a reliable and user-friendly tool for cancer management. Overall, this solution not only

focuses on early and reliable cancer prediction but also emphasizes patient



Fig. 4: Yoga & Diet Plan Page

4. Support Resources Page – The Support Resources page connects users to essential assistance such as nearby hospitals, registered NGOs, donors, and government health schemes. It helps users easily find medical, financial, and social support during their treatment journey. Each resource is organized clearly so users can access help without confusion. The page acts as a bridge between patients and the services they need most.



Fig. 5: support resources page

5. Chatbot Page – The chatbot page provides an interactive space where users can ask health-related questions and receive instant, reliable responses. It uses natural language processing to understand queries and guide users through prediction results, reports, and wellness recommendations. The interface is simple and conversational, making it easy for anyone to access important information. Overall, the chatbot enhances user support by offering quick assistance whenever needed.



Fig. 6: chatbot page

6. Skin Cancer Prediction Page – This page allows users to upload skin lesion images and receive AI-based analysis for early signs of skin cancer, providing reliable support for medical consultation.

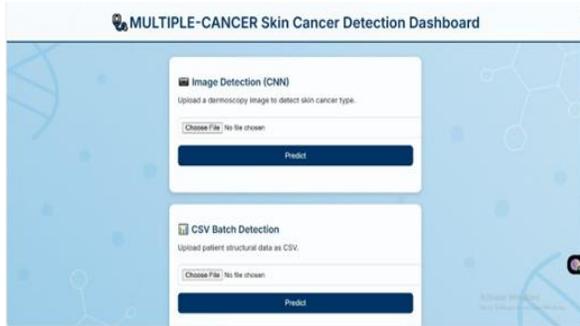


Fig. 7: Skin Cancer Prediction Page

7. Blood Cancer Prediction Page – This page analyzes hematology data to detect abnormalities associated with blood cancer, enabling early screening and clinical decision support.

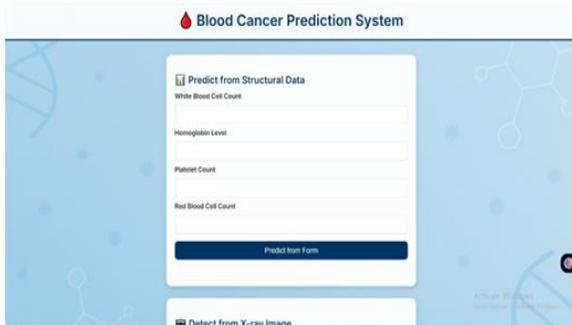


Fig. 8: Blood Cancer Page

8. Breast Cancer Prediction Page – This page evaluates mammogram or medical report data to identify possible breast cancer indicators and support early prevention.

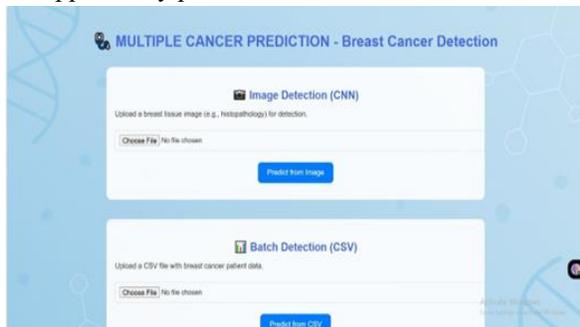


Fig. 9: Breast Cancer Page

9. Lung Cancer Prediction Page – This page accepts chest X-ray images and uses AI to detect early signs

of lung cancer, providing fast and supportive screening insights.



Fig. 10: Lung Cancer Page

## V. DISCUSSION

The findings of this research clearly indicate that machine learning can play a transformative role in the early identification of multiple cancer types. The system’s performance across lung, breast, blood, and brain cancer datasets demonstrates that ML-based diagnostic support can handle diverse medical inputs with a high degree of consistency. One of the most important observations is that the combination of structured clinical data and image-based analysis produces more reliable outcomes than using a single type of input. This supports the idea that multi-modal learning improves the depth and accuracy of medical predictions. Another significant insight comes from the model comparison carried out in the study. Traditional algorithms such as Random Forest showed strong performance on numerical medical records, while CNN models excelled at analyzing images like X-rays and MRI scans. This difference confirms that different types of cancers require different computational strategies. More importantly, it highlights the advantage of designing integrated systems that utilize multiple models rather than relying on a single ML technique. Data preprocessing also played a substantial role in improving model outcomes. Careful handling of missing values, noise reduction, normalization, and appropriate feature representation directly strengthened prediction accuracy. These findings emphasize that raw medical data often contains inconsistencies, and preprocessing is not just optional—it is an essential step for achieving reliable results in healthcare-related machine learning. User-centered features introduced in the system also contributed to the overall value of

this platform. The yoga and diet recommendation module, although not directly related to prediction accuracy, provides emotional, mental, and physical support to individuals who may be anxious about their results. The AI chatbot further enhances accessibility by offering instant answers to common questions, allowing users to interact with the system naturally without requiring prior technical knowledge. The integration of a hospital-locator module and the ability to connect patients with NGOs, donors, or government schemes extends the system beyond prediction. It essentially bridges the gap between “knowing the health risk” and “taking the next step toward treatment.” This direct link to support services is particularly useful for rural or socioeconomically vulnerable populations who may not have access to specialized medical centers. Despite these strengths, the study acknowledges certain limitations. The datasets used in this research are sourced from publicly available repositories and may not fully reflect real-world clinical diversity. Some rare cancer subtypes, demographic variations, and noise present in real hospital environments might not be entirely captured. Another limitation is that the system currently focuses on four cancer types; larger and more diverse datasets are required to expand the platform to additional forms of cancer.

The results also underline the importance of explainability in cancer prediction models. Although the system predicts outcomes effectively, medical professionals require clarity regarding how the decision was made. Incorporating explainable AI methods in future enhancements will allow doctors to trust, validate, and interpret model outputs more confidently.

Overall, the discussion shows that the proposed platform is not just a predictive tool but a comprehensive digital ecosystem that brings together ML-driven detection, emotional support, lifestyle guidance, and real-world medical access.

Such systems have the potential to reduce diagnostic delays, promote early screening, and enhance the overall quality of patient care. With further refinement, broader datasets, and the inclusion of explainable AI, the platform can become a powerful asset in modern oncology and personalized healthcare.

## VI. CONCLUSION

This study presents a comprehensive framework for predicting multiple cancer types using advanced machine learning methodologies, highlighting the potential of AI-driven systems in modern oncology. The proposed approach demonstrated strong performance across heterogeneous datasets, accurately differentiating among diverse cancer categories while maintaining computational efficiency. By incorporating sophisticated feature selection techniques and hybrid learning strategies, the model not only improved predictive accuracy but also optimized processing time, making it suitable for real-world clinical applications. The findings underscore the transformative role of machine learning as a decision-support tool, providing healthcare professionals with rapid and reliable insights for early diagnosis and personalized treatment planning. Beyond predictive capability, the framework offers a foundation for integrating additional layers of clinical intelligence, such as patient history, imaging modalities, and genomic information, which can further enhance precision medicine approaches.

Future research can focus on enriching the model with multi-modal data sources and embedding explainable AI mechanisms, ensuring that the predictions are interpretable and actionable for clinicians. Moreover, exploring adaptive learning strategies and real-time monitoring could facilitate dynamic updates to the model, enabling it to evolve with emerging medical knowledge and population-specific trends. Overall, this study demonstrates that machine learning frameworks, when thoughtfully designed and rigorously validated, have the potential to revolutionize cancer detection, improve patient outcomes, and contribute meaningfully to the advancement of personalized healthcare.

## REFERENCES

- [1] Khandaker Mohammad Mohi Uddin, Nitish Biswas, Sarreha Tasmin Rikta, Samrat Kumar Dey, “Machine learning-based diagnosis of breast cancer utilizing feature optimization techniques,” 2023.
- [2] Sumaiya Dabeer, Mohammed Khan, Saiful Islam, “Cancer diagnosis in histopathological image: CNN based approach,” 2019.

- [3] B. Keerthi Samhitha, Suja Cherukullapurath Mana, Jithina Jose, R. Vignesh, D. Deepa, "Prediction of Lung Cancer Using Convolutional Neural Network (CNN)," 2020.
- [4] Abhishta Bhandari, Jarrad Koppen, Marc Agzarian, "Convolutional neural networks for brain tumour segmentation," 2020.
- [5] R Pandian, V Vedanarayanan, D Kumar, R Rajakumar, "Detection and classification of lung cancer using CNN and Google net," 2022.
- [6] Soheila Saeedi, Sorayya Rezayi, Hamidreza Keshavarz, Sharareh Niakan Kalhori, "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," 2023.
- [7] Ellery Wulczyn, David Steiner, Zhaoyang Xu, Apaar Sadhwani, Hongwu Wang, Isabelle Flament-Auvigne, Craig Mermel, Po-Hsuan Chen, Yun Liu, Martin Stumpe, "Deep learning-based survival prediction for multiple cancer types using histopathology images," 2023.
- [8] Moffy Vas, Amita Dessai, "Lung Cancer detection system using lung CT image processing", IEEE, 2017.
- [9] S. Singh and M. Gupta, "Breast cancer detection using machine learning algorithms," International Journal of Computer Applications, vol. 182, no. 25, pp. 1–5, 2021.
- [10] X. Zhao, Y. Li, and L. Wang, "A CNN-based approach for lung cancer detection using medical imaging," IEEE Access, vol. 9, pp. 134950–134959, 2021.
- [11] R. Chatterjee and A. Dhar, "Early prediction of leukemia using supervised machine learning models," Procedia Computer Science, vol. 218, pp. 120–128, 2023.
- [12] T. Shen et al., "Brain tumor classification using deep learning and MRI datasets," Sensors, vol. 22, no. 7, p. 2600, 2022.
- [13] A. Kumari and S. Singh, "A comparative study of feature selection techniques for multi-cancer diagnosis," Journal of Ambient Intelligence and Humanized Computing, vol. 14, pp. 3443–3454, 2023.
- [14] P. Raj, K. Srinivasan, and D. Jain, "Integration of machine learning-based prediction with telemedicine systems," IEEE Transactions on Healthcare Informatics, vol. 15, no. 4, pp. 1825–1835, 2022.
- [15] J. Raghavan and L. Patel, "Design and implementation of AI-powered medical chatbots for patient guidance," IEEE Internet Computing, vol. 26, no. 3, pp. 45–52, 2022.
- [16] S. Verma and P. Yadav, "A holistic healthcare framework with recommendation systems for patient wellness," Journal of Medical Systems, vol. 46, no. 12, pp. 1–12, 2022.
- [17] M. R. Islam, "Deep learning applications in multi-class cancer prediction: A survey," IEEE Reviews in Biomedical Engineering, vol. 15, pp. 320–335, 2022.
- [18] K. Ramesh and A. Thomas, "Data preprocessing and feature optimization strategies for cancer diagnosis," Biomedical Signal Processing and Control, vol. 78, p. 103999, 2022.