IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, January 2024

Early Cancer Detection using Prediction Algorithm: A Survey

Mr. Sujit Prakash Tadadikar[1], Mr. Pankaj Sakpal[2], Mrs. Sujata Shahabade[3]
Embedded Software Engineer, Department of Science and Technology, IDOL, Mumbai, India[1]
Assistant Manager, Department of Science and Technology, IDOL, Mumbai, India [2]
Assistant Professor, Department of Computer Science, PCACS, Navi Mumbai, India[3]
logic.codedebugrepeat@gmail.com[1], pankaj.sakpal1819@gmail.com[2], sujathashahabade@gmail.com[3]

Abstract*: This survey paper provides an overview of the advancements in early cancer detection through the application of prediction algorithms. Early detection of cancer plays a crucial role in improving patient outcomes and reducing mortality rates. With the rise of computational technologies and machine learning techniques, predictive algorithms have emerged as powerful tools for identifying potential cancer cases at an early stage. This survey explores various prediction algorithms, their applications, challenges, and future directions in the realm of early cancer detection.*

Keywords: Cancer detection, DNA, Machine learning, Prediction Algorithm

I. INTRODUCTION

Cancer, a complex and heterogeneous group of diseases characterized by uncontrolled cell growth, remains a formidable global health challenge. Its far-reaching impact on individuals, families, and societies necessitates a continuous quest for innovative approaches to combat its devastating effects. According to the World Health Organization (WHO), cancer is a leading cause of morbidity and mortality worldwide, with new cases expected to surge in the coming decades. As our understanding of cancer biology and technological capabilities advances, the focus has shifted towards early detection as a pivotal strategy in the fight against this formidable adversary.

The insidious nature of cancer lies in its ability to progress silently, often evading detection until advanced stages when treatment options become limited, and prognosis is less favourable. Traditional diagnostic methods, while essential, have inherent limitations in reliably identifying cancers at their earliest, most treatable stages. This inherent challenge has prompted a paradigm shift in cancer research and healthcare delivery, with an increasing emphasis on the integration of cutting-edge technologies, such as prediction algorithms, to revolutionize early detection strategies.

Through a systematic examination of existing literature and studies, this survey aims to shed light on the state of the art in prediction algorithms for early cancer detection. By synthesizing current knowledge, identifying gaps, and addressing challenges, we hope to contribute to the ongoing efforts to enhance our ability to detect and treat cancer at its earliest and most curable phases, ultimately improving patient outcomes and reshaping the landscape of cancer care.

1.1 Background

Using visual, biological, and electronic health records data as the soleinput source, pretrained convolutional neural networks and conventional machine learning methods have been heavily employed for the identification of various malignancies. Initially, a series of pre-processing steps and image segmentation steps are performed to extract region of interest features from noisy features. Then, the extracted features are applied to several machine learning and deep learning methods for the detection of cancer.

1.2 Significance of Early Cancer Detection

The significance of early cancer detection is immense and can have a profound impact on patient outcomes, treatment options, and overall healthcare costs. Here are several key reasons why early cancer detection is crucial:

Improved Treatment Options:

Early detection often allows for a wider range of treatment options, including less invasive procedures, targeted therapies, and a higher likelihood of successful outcomes.

Early-stage cancers are generally more localized and may be easier to surgically remove or treat with localized therapies.

Increased Survival Rates:

Early detection is directly correlated with higher survival rates. When cancer is diagnosed atan early stage, the chances of complete remission or long-term survival are significantly improved.

Some cancers, when detected early, may have cure rates close to 100%, while late-stage cancers may have much lower survival rates.

Reduced Treatment Intensity:

Early detection may reduce the need for aggressive and extensive treatments, such as chemotherapy and radiation therapy, which can have significant side effects.

Patients diagnosed at later stages often require more intense treatments, leading to a higher risk of complications and a decrease in overall quality of life.

Lower Healthcare Costs:

Early detection can result in lower healthcare costs by reducing the need for extensive and prolonged treatments.

1.3 Role of Prediction Algorithms

Prediction algorithms play a crucial role in early cancer detection by leveraging data to identify patterns, make predictions, and assist healthcare professionals in diagnosing and treating cancer at its earliest and most treatable stages. Here are some key roles of prediction algorithms in the context of early cancer detection:

Risk Prediction:

Prediction algorithms can assess an individual's risk of developing cancer based on various factors such as age, family history, lifestyle, and genetic markers.

By identifying individuals at higher risk, healthcare providers can implement targeted screening and prevention strategies.

Early Detection from Imaging Data:

Machine learning algorithms can analyze medical imaging data, such as mammograms, CT scans, and MRIs, to detect subtle patterns indicative of early-stage cancer.

Computer-aided detection (CAD) systems, powered by prediction algorithms, can assist radiologists in identifying abnormalities that might be missed during manual interpretation.

Genomic Analysis:

Prediction algorithms analyze genomic data to identify genetic mutations or markers associated with an increased risk of certain cancers.

Molecular profiling and genomic analysis enable personalized medicine by tailoring treatments to the specific genetic characteristics of a patient's tumor.

Prediction of Treatment Response:

Algorithms can predict how individual patients will respond to different treatment options based on their genetic makeup, tumor characteristics, and other clinical factors.

This personalized approach helps optimize treatment plans, leading to more effective outcomes and minimizing unnecessary side effects.

In summary, prediction algorithms play a multifaceted role in early cancer detection, contributing to risk assessment, early diagnosis, treatment optimization, and ongoing monitoring. Their integration into healthcare systems has the potential to significantly improve patient outcomes and contribute to the advancement of precision medicine in the field of oncology.

II. TYPES OF CANCER PREDICTION ALGORITHMS

2.1 Machine Learning Algorithms

2.1.1 Support Vector Machines (SVM)

- **Description:** Classifies data points by finding the hyperplane that best separates them into classes.
- **Application:** Binary and multiclass cancer classification tasks.
- **Advantages:** Effective in high-dimensional spaces, versatile due to different kernel functions.

2.1.2 Random Forests

- **Description:** Ensemble learning method that builds multiple decision trees and combines their predictions.
- **Application:** Improved accuracy and robustness in cancer prediction.
- **Advantages:** Reduces overfitting, handles missing data, and captures complex relationships.

2.1.3 Neural Networks

- **Description:** Computational model inspired by the human brain, consisting of interconnected nodes (neurons).
- **Application:** Deep learning models for cancer prediction tasks.
- **Advantages:** Can capture complex patterns, suitable for large datasets.

2.1.4 Decision Trees

- **Description:** Tree-like model where each node represents a decision based on input features.
- **Application:** Classification and regression tasks in cancer prediction.
- **Advantages:** Easy to interpret, handles both categorical and numerical data.

2.2 Deep Learning Algorithms

- **Description:** Neural network architectures with multiple layers for hierarchical feature learning.
- **Application:** Image-based cancer prediction, time-series data analysis.
- **Advantages:** Effective for complex data structures, automatic feature learning.

III. APPLICATIONS OF PREDICTION ALGORITHMS IN EARLY CANCER DETECTION

3.1 Breast Cancer:

3.1.1 Mammogram Analysis:

- Prediction algorithms analyze mammograms to detect early signs of breast cancer.
- Techniques include deep learning models for image classification and anomaly detection.

3.1.2 Genomic Data:

- Integrating genomic data for predicting breast cancer risk and subtype classification.
- Models may utilize gene expression patterns and mutations associated with breast cancer.

3.2 Lung Cancer:

3.2.1 CT Scan Analysis:

- Early detection through the analysis of chest CT scans using machine learning.
- Algorithms identify nodules and patterns indicative of lung cancer at its early stages.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, January 2024

3.2.2 Predictive Biomarkers:

- Identification of predictive biomarkers in blood samples or respiratory fluids.
- Algorithms analyze molecular and genetic data to assess lung cancer risk.

3.3 Colorectal Cancer:

3.3.1 Colonoscopy Image Analysis:

- Image recognition algorithms analyze colonoscopy images to identify precancerous lesions.
- Deep learning models aid in the early detection of colorectal cancer.

3.3.2 Stool DNA Testing:

- Prediction algorithms assess DNA markers in stool samples for the early detection of colorectal cancer.
- Machine learning helps interpret complex molecular patterns.

3.4 Prostate Cancer:

3.4.1 PSA Levels Prediction:

- Algorithms predict prostate-specific antigen (PSA) levels in blood to assess prostate cancer risk.
- Machine learning models analyze historical PSA data and other clinical features.

3.4.2 MRI Imaging:

- Advanced image analysis techniques, including radiomics, analyze MRI scans for prostate cancer detection.
- Models may predict the likelihood of malignancy based on imaging features.

3.5 Ovarian Cancer:

3.5.1 CA-125 Biomarker Prediction:

- Prediction algorithms assess CA-125 blood marker levels for ovarian cancer risk.
- Machine learning models integrate CA-125 trends, patient history, and other factors.

3.5.2 Ultrasound Image Analysis:

- Early detection through the analysis of ultrasound images for ovarian masses.
- Algorithms help differentiate between benign and malignant tumors.

3.6 Pancreatic Cancer:

3.6.1 Genetic Risk Prediction:

- Algorithms predict pancreatic cancer risk based on genetic markers.
- Machine learning models analyze family history and genetic data.

3.6.2 CT and MRI Image Analysis:

- Imaging analysis aids in early pancreatic cancer detection using CT and MRI scans.
- Algorithms identify abnormalities and patterns associated with pancreatic tumors.

3.7 Challenges and Opportunities in Multimodal Approaches:

3.7.1 Data Integration:

- Challenge: Integrating diverse data sources, such as imaging, genomic, and clinical data.
- Opportunity: Multimodal approaches leverage the complementary nature of different data types for more accurate predictions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, January 2024

3.7.2 Interpretable Models:

- Challenge: Ensuring interpretability in complex multimodal models.
- Opportunity: Developing transparent models that provide insights into decision-making for clinicians.

3.7.3 Data Privacy and Security:

- Challenge: Handling sensitive medical data while ensuring patient privacy.
- Opportunity: Implementing secure and ethical data-sharing practices to foster collaboration.

3.7.4 Clinical Validation:

- Challenge: Validating prediction algorithms in real-world clinical settings.
- Opportunity: Collaborating with healthcare professionals to validate the effectiveness of early detection models.

3.7.5 Population-specific Considerations:

- Challenge: Addressing population-specific variations in cancer risk and biology.
- Opportunity: Customizing prediction algorithms to diverse populations for more personalized early detection strategies.

Comparative Analysis

The comparative analysis section highlighted the study of diferent researchers for cancer disease detection using AI techniques. The prediction outcomes are classifed on basis of parameters such as accuracy, sensitivity/recall, precision, specifcity, dice score, Area under the Curve. Figure 9 provides the description of multiple evaluation parameters. Table 1 comprises the comparative analysis based on multiple evaluation parameters for various cancer types. As shown in the comparative analysis, many research works have been analyzed for cancer diagnosis and detection using conventional machine and deep learning methods. It can be observed that most of the deep learning techniques have performed well and achieved high accurateness in terms of the prediction scores obtained. Also, most of the research articles have been published recently (2020). Also, most of the studies have worked on the diagnosis of breast cancer.

Table 1 Comparative analysis using AI techniques for different cancers

Authors	**Cancer types**	**Training data**	**Techniques**	**Challenges**	**Reported outcomes**
Sudharani et al. [1]	Brain	MRIs images	Fuzzy C-Means	The small and unstructured data were not used in the system, restricting the generality and clinical applicability	Accuracy =89.2% Sensitivity =88.9% Specifcity =90%
Patil et al. [2]	Breast	Mammogram images	CRNN FC-CSO	The current system did not work with blur images which should be improved by using a wiener flter	Accuracy=98.4% Specifcity=99.9% F1-score=74.5%
Wu et al. [3]	Cervical	Pathological images	CNN	The accuracy of the system can be improved by incorporating more training datasets	Accuracy=93.33%
Hoerter et al. [4]	Colorectal	ImageNet database	CNN	The current system is restricted to detect polyps that are smaller than 10 mm	per-polyp sensitivity =71%
Santini et al. [5]	Kidney	KiTS19	CNN	New training strategies will be designed to diferentiate	Mean Dice score=0.96

				between the data, and a diferent stage will be added for more detailed local features for escalating the current system's efficiency	
Chlebus et al. [6]	Liver	CT images	Deep-CNN	The present system requires more work to be done to match the performance of human expertise Detection rate=77%	Detection rate=77%
Le et al. [7]	Brain	BraTS 2018	CNN	Random Forest Regression The current system should add more datasets to increases the prediction rate in the future	Predict the survival rate
Shakeel et al. [8]	Lung CT	images	Improved-DNN	The present system can be improved by adding more datasets to it	Accuracy = 96.2% Specifcity =98.4% Precision =97.4%
Bur et al. [9]	Oral	NCDB dataset	Tumor depth of invasion (DOI) Model	The current system needs improved predictive algorithms to enhance accuracy in detecting oral cancer in patients	Sensitivity=86.6%
Liu et al. [10]	Prostate	MRI images	CNN	The current system worked on a limited dataset that should be increased to improve its efciency	AUC=0.84

IV. CONCLUSION

This review study attempts to summarize the various research directions for AI-based cancer prediction models. AI has marked its significance in the area of healthcare, especially cancer prediction. The paper provides a critical and analytical examination of current state-of-the-art cancer diagnostic and detection analysis approaches—a thorough examination of the machine and deep learning models used in cancer early detection using medical imaging. The AI techniques play a signifcant role in early cancer prognosis and detection using machine and deep learning techniques for extracting and classifying the disease features. Our study concluded that most previous literature works employed deep learning techniques, especially Convolutional Neural Networks. Another signifcant factor noted in our study is that most studies have worked on breast cancer data. It was examined that when deep learning models are applied to pre-processed and segmented medical images, the images perform better in classifcation metrics such as AUC, Sensitivity, Dice-coefcient, and Accuracy. There is scope to work on early detection of head and neck cancers because less study has been conducted for both types of cancer. Also, the federated learning model can be used for cancer detection based on distributed datasets. hence, we intend to use a federated learning model for the detection of cancer disease by creating the decentralized training model for cancer datasets in remote places.

REFERENCES

[1]. Sudharani K, Sarma TC, Prasad KS (2016) Advanced morphological technique for automatic brain tumor detection and evaluation of statistical parameters. Procedia Technol 24:1374–1387. https://doi.org/10.1016/j.protcy.2016.05.153

[2]. 31. Patil RS, Biradar N (2020) Automated mammogram breast cancer detection using the optimized combination of convolutional and recurrent neural network. Evol Intel. https://doi.org/10.1007/ s12065-020-00403-x

[3]. Wu M, Yan C, Liu H, Liu Q, Yin Y (2018) Automatic classifcation of cervical cancer from cytological images by using convolutional neural network. Biosci Rep 38(6):1–9. https://doi.org/10.1042/BSR20181769

[4]. Hoerter N, Gross SA, Liang PS (2020) Artificial intelligenceand polyp detection. Current Treatment Options Gastroenterol18(1):120–136.https:// doi. org/ 10. 1007/ s11938- 020- 00274-2\

[5]. Santini G, Moreau N, Rubeaux M (2019) Kidney tumorsegmentationusing an ensembling multi-stage deep learning approach.A contribution to the KiTS19 challenge. 1–11.https:// doi.org/ 10.24926/ 548719. 023

[6]. Chlebus G, Schenk A, Moltz JH, Ginneken BV, Hahn HK, MeineH (2018) Automatic liver tumor segmentation in CT with fullyconvolutional neural networks and object-based postprocessing.Sci Rep 8:15497. https:// doi. org/ 10. 1038/ s41598- 018- 33860-7

[7]. Le TN, Bao PT, Huynh HT (2016) Liver tumor segmentationfrom MR images using 3d fast marching algorithm and singlehidden layer feedforward neural network. Biomed Res Int2016:3219068. https:// doi. org/ 10. 1155/ 2016/ 32190 68

[8]. Shakeel PM, Burhanuddin MA, Desa MI (2020) Automatic lungcancer detection from CT image using improved deep neuralnetwork and ensemble classifier. Neural Comput Appl. https://doi. org/ 10. 1007/ s00521- 020- 04842-6

[9]. Bur AM, Holcomb A, Goodwin S, Woodroof J, Karadaghy O,Shnayder Y, Kakarala K, Brant J, Shew M (2019) Machine learningto predict occult nodal metastasis in early oral squamous cellcarcinoma. Oral Oncol 92:20–25. https:// doi. org/ 10. 1016/j. oraloncology. 2019. 03. 011

[10]. Liu S, Zheng H, Feng Y, Li W (2017) Prostate cancer diagnosisusing deep learning with 3D multiparametric MRI. Medicalimaging 2017: Comput Aid Diagn 10134:1013428. https:// doi.org/ 10. 1117/ 12. 22771 21