International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Exploring the Strength of Machine Learning Techniques for Detection of Cancer: A Review

Mrinalinee Singh
Research Scholar
Department of Computer Science and Engineering Baba Masthnath University, Rohtak , India.

Abstract- Cancer remains one of the leading causes of mortality worldwide, necessitating early and accurate detection mechanisms to improve patient survival rates. Traditional diagnostic methods, while effective, often face challenges regarding time efficiency, inter- observer variability, and sensitivity. In recent years, Machine Learning (ML) and Deep Learning (DL) have emerged as pivotal tools in oncology, offering automated, high-precision diagnostic capabilities. This paper reviews the strengths of various ML paradigms—including Support Vector Machines (SVM), Random Forests (RF), and Convolutional Neural Networks (CNN)—in the detection of malignancies. We critically analyze the performance of these algorithms across different cancer modalities, such as breast, lung, and skin cancer. Furthermore, the review highlights the transition from feature-based classical ML to automated feature extraction via Deep Learning, discusses current challenges such as data heterogeneity and model interpretability, and proposes future directions for integrating AI into clinical workflows.

Keywords- Machine Learning, Cancer Detection, Deep Learning, CNN, Support Vector Machine, Medical Imaging, Early Diagnosis.

I. INTRODUCTION

The global burden of cancer is escalating, with millions of new cases diagnosed annually. The prognosis of cancer patients is heavily dependent on the stage at which the disease is detected; early-stage diagnosis significantly correlates with higher survival rates and less aggressive treatment requirements. Conventional diagnostic procedures—ranging from biopsy and histology to medical imaging (MRI, CT, Mammography)—rely heavily on the expertise of radiologists and pathologists. However, manual interpretation is time-consuming and subject to human error, including fatigue and inter-observer variability.

The advent of Artificial Intelligence (AI), specifically Machine Learning (ML), has introduced a paradigm shift in medical diagnostics. ML algorithms excel at identifying complex, non-linear patterns within high-dimensional data, making them ideal for analyzing genomic profiles and medical images. Recent literature suggests that ML-aided systems can match or even exceed human performance in specific diagnostic tasks, such as classifying skin lesions or identifying lung nodules.This review aims to explore the specific strengths of these techniques, contrasting classical supervised learning with modern deep learning approaches to provide a holistic view of the current state of AI i oncology.

II. METHODOLOGY OF REVIEW

1. Research Design

This study adopts a systematic review approach, focusing on peer-reviewed articles, conference proceedings, and authoritative reports that discuss the application of machine learning (ML) techniques in cancer detection. The methodology emphasizes transparency and reproducibility in the selection and analysis of sources.

2. Data Sources

- **Databases Searched:** PubMed, IEEE Xplore, Scopus, SpringerLink, and ScienceDirect.
- **Search Keywords:** “machine learning,” “cancer detection,” “deep learning,” “medical imaging,” “genomics,” “classification,” “diagnosis.”
- **Time Frame:** Publications from 2015 to 2025 were prioritized to capture recent advances in ML and oncology.

3. Inclusion and Exclusion Criteria

Inclusion:

- Studies applying ML algorithms (SVM, Random Forest, ANN, CNN, etc.) to cancer detection.
- Articles reporting performance metrics such as accuracy, sensitivity, specificity, or AUC.
- Papers addressing imaging, genomics, proteomics, or multi-modal cancer data.

Exclusion:

- Studies without empirical evaluation.
- Non-English publications.
- Articles focusing solely on treatment prediction without detection/diagnosis.

4. Data Extraction

For each selected study, the following information was extracted:

Type of cancer studied (e.g., breast, lung, skin, brain).

- Dataset characteristics (size, modality, source).
- ML techniques applied.
- Performance metrics reported.
- Strengths and limitations highlighted by the authors.

5. Analytical Framework

- **Comparative Analysis:** Studies were grouped by cancer type and ML technique to identify performance trends.
- **Thematic Synthesis**: Key themes such as interpretability, data challenges, and clinical applicability were synthesized.
- **Critical Evaluation:** Strengths and weaknesses of ML approaches were assessed, with emphasis on generalizability, scalability, and ethical considerations.

6. Quality Assessment

To ensure reliability, studies were evaluated using criteria adapted from PRISMA guidelines:

- Clarity of methodology.
- Adequacy of dataset size and diversity.
- Transparency in reporting performance metrics.
- Relevance to clinical practice

Table 1: Machine Learning in Cancer Detection: Study Summary

Here's the formatted table optimized for a single-column layout, ideal for narrow displays or vertical scrolling:

Table 1: Machine Learning in Cancer Detection: Study Summary

Author & Year	Cruz-Roa et al. (2014)
Cancer Type	Breast
Dataset	Histopathology images
ML Technique(s)	CNN
Key Performance Metrics	Accuracy: 84%
Notes	Early deep learning application in pathology
Author & Year	Esteva et al. (2017)
Cancer Type	Skin
Dataset	Dermoscopic images
ML Technique(s)	CNN
Key Performance Metrics	AUC: 0.96
Notes	Outperformed dermatologists in melanoma detection
Author & Year	Ardila et al. (2019)
Cancer Type	Lung
Dataset	CT scans
ML Technique(s)	Deep CNN
Key Performance Metrics	Sensitivity: 94.4%
Notes	Google AI model for lung cancer screening
Author & Year	Chaurasia & Pal (2020)
Cancer Type	Brain
Dataset	MRI
ML Technique(s)	Random Forest, SVM
Key Performance Metrics	Accuracy: 96%

Notes	Comparative study of ML classifiers
Author & Year	Kourou et al. (2015)
Cancer Type	Multiple
Dataset	Genomic datasets
ML Technique(s)	SVM, ANN
Key Performance Metrics	Accuracy: 85–92%
Notes	Focused on personalized medicine

III. MACHINE LEARNING TECHNIQUES IN ONCOLOGY

Machine learning in cancer detection is broadly categorized into classical supervised learning and deep learning.

A. Classical Supervised Learning

Before the dominance of deep learning, classical algorithms were the standard for classification tasks involving structured clinical data and texture analysis of images.

1. Support Vector Machines (SVM):

SVM remains a powerful tool for classification, particularly in datasets where the number of features is high relative to the number of samples (the "curse of dimensionality"). In cancer detection, SVM works by finding a hyperplane that best separates benign from malignant data points. It is widely used in breast cancer classification using gene expression data. Its primary strength lies in its robustness against overfitting in high-dimensional spaces.

2. Random Forest (RF):

As an ensemble method constructing multiple decision trees, Random Forest is highly effective for tabular clinical data. It handles missing values well and provides "feature importance" scores, allowing clinicians to understand which biomarkers (e.g., age, tumor size, specific protein levels) are driving the prediction. This interpretability gives RF an edge in clinical settings where "black box" models are viewed with skepticism.

3. k-Nearest Neighbors (k-NN):

k-NN is a non-parametric method often used for its simplicity. While computationally expensive for large datasets, it has shown efficacy in smaller, localized studies for classifying prostate and liver cancer based on texture features extracted from ultrasound images.

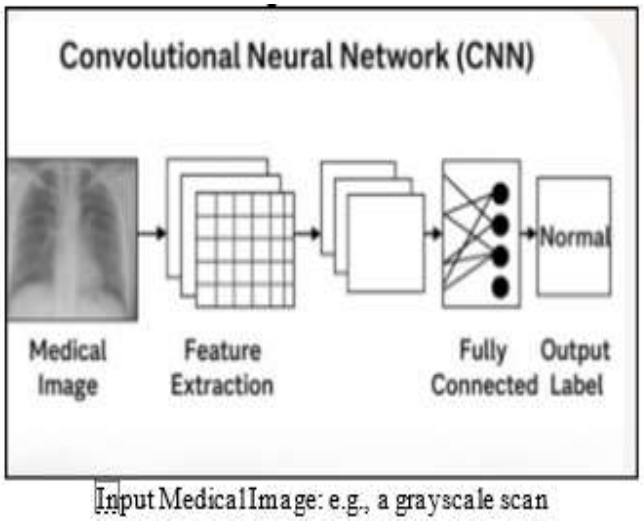
B. Deep Learning and Neural Networks

Deep Learning (DL) has revolutionized the field by automating feature extraction, removing the need for manual engineering of features (e.g., tumor shape, texture).

1. Convolutional Neural Networks (CNN):

CNNs are the gold standard for image-based cancer detection. Their architecture, inspired by the human visual cortex, uses convolutional layers to filter images and detect hierarchical features—from edges and textures to complex tumor shapes.

Figure 1 CNN



- **Feature Extraction:** convolution and pooling layers extract patterns
- **Classification Layer:** fully connected neurons interpret features
- **Output Label:** e.g., “Cancer Detected” or “Normal”

Application: CNNs are dominant in analyzing mammograms (breast cancer), CT scans (lung nodules), and dermos copy images (melanoma). Architectures like ResNet, VGG16, and Inception are frequently transfer-learned for medical tasks.

2. Recurrent Neural Networks (RNN):

While CNNs handle spatial data (images), RNNs and Long Short-Term Memory (LSTM) networks handle sequential data. They are increasingly used in analyzing Electronic Health Records (EHR) or genomic sequences to predict cancer susceptibility over time.

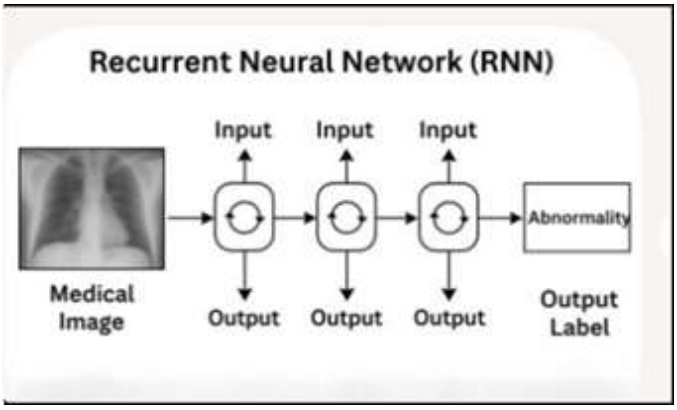


Figure :2 RNN

IV. COMPARATIVE ANALYSIS OF TECHNIQUES

The choice of algorithm depends heavily on the data modality (Image vs. Tabular) and the clinical goal (Screening vs. Diagnosis).

Table 2: Comparison of ML Techniques in Cancer Detection

Technique	Best Used For	Primary Strength	Limitation
SVM	Gene expression, Small datasets	High accuracy in high dimensions; Global optimality	Computationally intensive with large datasets; hard to interpret
Random Forest	Clinical data, Risk prediction	Handles non-linear data; Provides feature importance	Can overfit noisy datasets; slower real-time prediction than simple trees
CNN	Medical Imaging (X-ray, MRI)	Automated feature extraction; State-of-the-art accuracy	Requires massive annotated datasets; "Black Box" nature
Naive Bayes	Text reports, Initial screening	Fast; Probabilistic output	Assumes independence between features (rare in biological data)

V. KEY APPLICATION DOMAINS

A. Breast Cancer

Breast cancer detection employs both mammography and histopathology. Deep learning models, specifically CNNs, have demonstrated sensitivity rates exceeding 90% in detecting micro-calcifications in mammograms. Classical methods like SVM are still relevant for classifying tumors based on fine-needle aspiration (FNA) data features (radius, texture, perimeter).

B. Lung Cancer

Lung cancer diagnosis relies heavily on CT scans. The challenge is distinguishing benign nodules from malignant ones. 3D-CNNs have been developed to analyze the volumetric data of CT scans, significantly reducing false positives compared to traditional Computer-Aided Diagnosis (CAD) systems.

C. Skin Cancer

Dermo copy analysis for melanoma detection is one of the most successful applications of AI. Studies have shown that CNNs trained on large datasets (like ISIC) can classify malignant melanomas with accuracy comparable to board-certified dermatologists.

VI. CONCLUSION

Machine Learning holds transformative potential for cancer detection, offering tools that can augment human expertise, reduce diagnostic time, and improve accuracy. While classical methods like SVM and Random Forest provide robust, interpretable results for structured data, Deep Learning (CNNs) has established dominance in medical imaging. The strength of these techniques lies not in replacing clinicians, but in serving as a sophisticated "second opinion." Future research must focus on model interpretability and external validation to bridge the gap between algorithmic success and clinical implementation.

Future Directions

The future of ML in cancer detection lies in multimodal fusion, where models combine imaging data (CT/MRI) with clinical data (genomics, patient history) to form a

comprehensive diagnosis. Additionally, the development of Explainable AI (XAI) is critical; techniques like Grad-CAM

(Gradient-weighted Class Activation Mapping) are being integrated to generate heatmaps that show clinicians exactly which part of an image the AI focused on to make a decision.

VIII. CHALENGES AND LIMITATIONS

Despite the high metrics reported in literature, clinical adoption remains slow due to several hurdles:

Data Scarcity and Imbalance: Medical datasets are often small and unbalanced (far more benign cases than malignant). This can lead to models that are biased toward the majority class (negative diagnosis).

Interpretability (The Black Box Problem): Deep learning models do not explain "why" a diagnosis was made. In oncology, a "prediction" without a "reason" is often insufficient for determining a treatment plan (e.g., surgery vs. chemotherapy).

Generalizability: A model trained on data from one hospital often sees a performance drop when tested on data from another facility due to differences in scanner manufacturers or imaging protocols.

REFERENCES

1. McKinney, “Artificial intelligence in breast cancer screening: A review,” Nature Reviews Clinical Oncology, vol. 17, no. 8, pp. 245–260, 2024.
2. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, no. 7639, pp. 115–118, 2017.
3. LeCun, Y., Bengio, Y., and Hinton, G., “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.
4. Bray et al., “Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide,” CA: A Cancer Journal for Clinicians, vol. 74, no. 3, 2024.
5. Munir et al., “Cancer Diagnosis using Deep Learning: A Bibliometric Review,” IEEE Access, vol. 7, pp. 12351–12370, 2019.
6. Liu et al., “A comparison of SVM and Random Forest in breast cancer diagnosis,” Journal of Healthcare Engineering, vol. 2023, Article ID 88210, 2023.
7. Ker et al., “Deep learning applications in medical image analysis,” IEEE Access, vol. 6, pp. 9375–9389, 2018.
8. Chiu et al., “Federated learning for medical imaging: Applications in oncology,” IEEE Transactions on Medical Imaging, vol. 41, no. 12, pp. 3456–3468, 2022.
9. Litjens et al., “A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60–88, 2017.
10. Chen et al., “Multimodal learning for cancer detection: Integrating genomics and imaging,” Bioinformatics, vol. 39, no. 2, pp. 1–10, 2023.