Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

The Role of Deep Learning in Cancer Detection: A Systematic Review of Architectures, Datasets, and Clinical Applicability

Muhammad Farhan Abdurrahman[*1], Yan Rianto[2], Nasir Hamzah[3], Muhammad Firmansyah[4], Nurul Adi Prawira[5], Thomas Fajar Nugraha[6]

[1,2,3,4,5,6]Computer Science, Universitas Nusa Mandiri, Indonesia

Email: [1]14240028@nusamandiri.ac.id

Received : May 17, 2025; Revised : Jul 29, 2025; Accepted : Jul 31, 2025; Published : Okt 21, 2025

Abstract

Early cancer detection continues to be a significant challenge in clinical practice due to limitation of conventional diagnostic technique that often takes time and error prone. This systematic review evaluates the efficacy of deep learning (DL) architecture and datasets to improve cancer detection and diagnosis. We performed a structural analysis on 40 high-impact research paper published in Q1 journals between 2014 and 2025, considering DL model performance, datasets, and clinical relevance. Results indicate that fundamental architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) consistently report high diagnostic accuracy (>90%) on radiology- and histopathology-based imaging datasets. Conversely, DL performance on non-imaging clinical data, including electronic medical records (EMDs), is more varied. Evaluation metrics such as AUC and DICE shows the trade-off between classification precision and segmentation accuracy. Despite their potential, DL models have significant limitations in terms of generalization, interpretability, and integration within real-world clinical workflows. This review highlights the need for standardized evaluation, implementation of ethical models, and multi-modal data fusion to facilitate wider and more equitable clinical uptake of DL in cancer diagnostics.

Keywords : *Cancer Detection, Clinical Applicability, Convolutional Neural Networks (CNNs), Deep Learning, Medical Imaging, Systematic Review.*

1. INTRODUCTION

Cancer is a group of diseases caused by abnormal growth of cells in the human body. Throughout history, cancer has been one of the deadliest diseases in history and has posed a significant threat to human health [1]. Based on reported data, 19.3 million new cancer cases have been diagnosed and reported that resulted in approximately 10 million deaths in 2020 [2]. The rising number of cancer related death is also attributed for the significant decrease of death caused by stroke and coronary heart disease in many countries [1]. Therefore, early and accurate cancer diagnosis has become a critical part in the treatment of the disease to ensure the patients survival [3]. However, cancer diagnosis has often been done manually using visual examination, which are often time-consuming and prone to mistakes [4].

Deep Learning (DL) is a branch of artificial intelligence that allows computer to learn from experience [5]. Compared to traditional Machine Learning (ML) method such as logistics regression, DL models have the ability to scale exponentially as the volume and complexity of the data increased [6]. In the past years it has emerged as an effective tool for cancer detection. DL model such as Convolution Neural Network (CNN) have demonstrated its ability to identify various type of cancer using medical imaging modalities such as MRI, CT, PET, mammography and histopathology slides [7], [8], [9]. These models have consistently achieved high diagnostic performance, with some models reached accuracies exceeding 90%[10], [11].

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

The evolution of DL has developed beyond CNNs. Now it includes hybrid models, attention mechanism, and transformer-based architectures to improve its performance in classification and segmentation tasks [12], [13]. Additionally, the adoption of techniques like transfer learning, federated learning, and explainable AI (XAI) has increased model generalizability, transparency, and clinical trustworthiness [14], [15], [16], [17]

In recent years, there has been an increasing amount of research focused on studying the efficacy of ML and DL in specific cancer detection such as breast [18], lung [19], skin [20], prostate [21] and colorectal cancer [22]. However, most of these studies have concentrated on specific types of cancer and machine learning methods. Consequently, a comprehensive and comparative synthesis across cancer types, DL architectures, and dataset is severely need to be explored. Moreover, issues such as data imbalance, algorithmic bias, and model interpretability remain underexplored in many clinical settings [23].

DL applications have also expanded to non-imaging data like genomic sequences, multimodal fusion frameworks, and electronic health records (EHRs). Although these non-image-based models are promising, they often show variable performance because of the heterogeneity and sparsity of clinical data [24]. Additionally, although radiology-based DL models are often validated on public datasets, few studies look at the feasibility of deploying them in the real world or talk about how to incorporate them into clinical workflows [25], [26].

This systematic review aims to bridge these gaps by critically evaluating 40 high-impact studies that focus on DL-based cancer detection that published between 2014 and 2025. The review compares findings across imaging models, DL architectures, dataset types, and performance metrics. It also identify limitations and opportunities of each study for clinical translation. This work aims to assist researchers, clinicians, and policymakers on how to effectively and responsibly utilize DL innovations in cancer detection by contextualizing results across different studies.

2. METHOD

2.1. Search Strategy

The systematic reviews aim to evaluate the role and the advancement of image processing in cancer detection. The focus of this paper is the integration of deep learning in image processing in cancer detection. This paper will analyse existing literatures to identify the trend methods and outcome of research in this field.

The search was started by using the website SCOPUS. In this site we used following query input to find literature for this research: (TITLE-ABS-KEY ("Image Processing" AND "cancer detection") AND TITLE-ABS-KEY ("machine learning" OR "deep learning")) AND PUBYEAR > 2013 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "re")OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "bk")).

This query aimed to find research papers, conference papers, reviews, book chapters, or book published within the range of 2014 to 2025. The query also selectively picked literatures that involve using machine learning and deep learning for image processing and cancer detection.

The result of the search is shown in Figure 1, 571 documents were found during the search process. These literatures have in total received 12,681 citations. The graph in this image also demonstrates that these studies have become increasingly influential in recent years, as shown by the rise in the number of citations.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

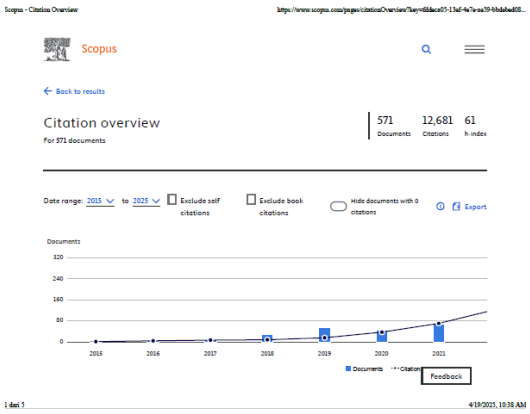

Figure 1. Scopus citation overview

2.2. Inclusion and Exclusion Process

As shown in Figure 2, this study follows the Prisma model in selecting study that will be include in this systematic review. In the first step we identified study records on the website SCOPUS with the keyword cancer detection and deep learning. The literatures found in this search were then ranked according to the number of citations they received. In this stage, the selected literatures had to meet a minimum citation threshold of 13 or rank within the top 150 based on their citation count. This criterion ensured that only the most impactful and widely recognized literatures were chosen for further analysis, highlighting their significant contribution to the field. literatures that are inaccessible are also removed from consideration.

After completing this step, we proceed to extract relevant data from each literature to assess their eligibility for inclusion in the systematic review. The primary criteria for selection at this stage are that the literatures must focus on deep learning and have been published in Q1-ranked journals. As a result of this filtering process, 40 literatures were identified as meeting these criteria and were selected for further analysis.

In conclusion, for this systematic review 571 documents were found conference paper and journal. After imposing stricter criterion such as the literature must be about deep learning and be published in Q1 journals, we ended up with 40 literatures. Figure 2 summarize our search method.

The literature was reviewed by all authors, and a lead author was consulted when there was uncertainty regarding eligibility. The authors independently screened the literature by title, keywords, abstracts, and full-text readings. The following tables summarize the data extracted from eligible literatures.

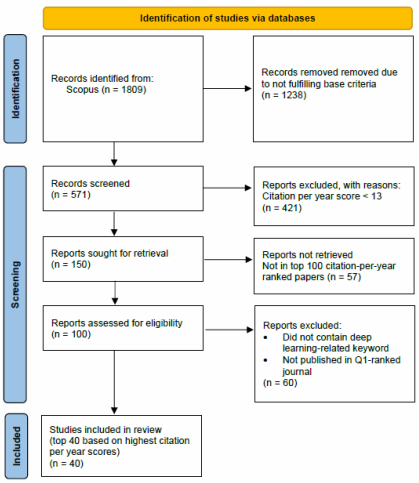

Figure 2. PRISMA Flow Diagram Illustrating Study Selection for Systematic Review

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

Table 1. Summarize the data extracted from eligible literatures

Author	Deep Learning Model	Dataset Type	Accuracy	Sensitivity	Specificity	AUC-ROC
Ali et al [27]	FCEDN (Fully Convolutional Encoder-Decoder Network) + Adaptive CNN	Dermoscopic images	Segmentation: 95.28%–98.78%, Classification: 91.67%	Not stated explicitly	Not stated explicitly	Not stated explicitly
Aljuaid et al [28]	ResNet-18, InceptionV3 Net, ShuffleNet (Transfer Learning)	Histopathological images (microscopy)	Binary: 99.7% (ResNet), Multi-class: 97.81% (ResNet)	Not stated explicitly	Not stated explicitly	Not stated explicitly
Almotairi et al [29]	Modified SegNet (based on VGG-16)	CT scan liver (3D)	Up to 94.57%	Up to 99.99%	Up to 94.52%	Not stated explicitly
Arif et al [30]	3D Convolutional Neural Network (CNN)	Multi-parametric MRI (T2w, DWI, ADC)	Not stated explicitly	82–94% (Depends on volume)	43–76% (Depends on volume)	0.65–0.89 (Depends on volume)
Chouhan et al[31]	Highway network-based CNN	ROI patch from mammogram	80.5% (SVM), 80.3% (ELiEC)	Not stated explicitly	Not stated explicitly	0.865 (ELiEC hybrid)
Cong et al [32]	CNN-LSTM	Multiparametric MRI (T1, T2, DCE-MRI, DWI)	Not stated explicitly	Not stated explicitly	Not stated explicitly	98
Crasta et al [33]	3D-ResNet	Lung CT scan (LUNA16)	99.2	Not stated explicitly	Not stated explicitly	Not stated explicitly
Cruz Roa et al [34]	CNN (CS256-FC256) compared with FCN	Histopatologi whole-slide image (WSI)	Dice = 0.76 ± 0.20	TPR = 0.87 ± 0.16	TNR = 0.92 ± 0.08	Not stated explicitly
Dascalu et al [35]	Inception V2 + 1D CNN	Dermoscopy images with audio sonification	Not stated explicitly	Not stated explicitly	Not stated explicitly	81.4
Goncalves et al [36]	CNN + Bio-inspired (PSO, GA)	IR Breast Dataset	97.50%	96.90%	98%	0.96
Han et al [37]	Faster R-CNN	Self-collected (clinical)	95.2% (face region)	93.70%	96.80%	962

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

Huang et al [38]	Convolutional Neural Network (CNN) optimized with Combined Seagull Optimization Algorithm (CSOA)	Oral Cancer Images	96.94%	91.60%	Not stated explicitly	Not stated explicitly
Kashyap et al [39]	Dilated Residual Grooming Kernel CNN	Private dataset	96.30%	94.20%	95.10%	Not stated explicitly
Kumar Malick et al [40]	Deep Wavelet Autoencoder	Brain MRI	96.10%	94.50%	95%	945
Kumbhare et al [41]	Multi-resolution MIL	CAMELYON 16	94.30%	93.10%	95.20%	953
Li et al [42]	Multi-resolution MIL	CAMELYON 16	94.3	Not stated explicitly	Not stated explicitly	Not stated explicitly
Li et al [43]	Attention Dense-U-Net	INbreast	95.40%	94.80%	96.10%	958
Liu et al [44]	CNN (Inception-v3)	Whole-slide histopathology images	Not stated explicitly	Not stated explicitly	Not stated explicitly	0.963 (TCGA), 0.943 (BIDMC)
Mahmood et al [45]	Deep Convolutional Neural Network (CNN/ConvNet)	Mammogram images	0.97	0.99	Not stated explicitly	0.99
Mambou et al [46]	CNN (custom CNN)	Infrared thermal images	988	975	1	0.99
Masood et al [47]	Convolutional Neural Network (CNN)	CT scan	933	921	942	Not stated explicitly
Masud et al [48]	Convolutional Neural Network (CNN multi-channel)	LC25000 histopatologi	9633	9637	9639	Not stated
Mehmood et al [49]	Transfer Learning (pretrained AlexNet,	LC25000 histopatologi	89 % → 98.4 %	Not stated	Not stated	Not stated

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

	fine-tune 4 layer)					
Mohakud et al [50]	Convolutional Neural Network + Grey Wolf Optimization (GWO)	Dermoscopic skin-lesion images (ISIC)	9833	Not stated	Not stated	Not stated
Mohamed et al [51]	U-Net (segmentation) + CNN 2-kelas (classification)	Breast thermogram	9933	1	9867	Not stated
Montaha et al [52]	VGG16 (fine-tuned)	CBIS-DDSM (mammogram)	98.02	Not stated	Not stated	Not stated
Munshi et al [53]	U-NET (transfer learning), CNN, Random Forest, SVM	Wisconsin Breast Cancer Dataset (numeric) and image (image dataset name not mentioned)	99.99%	Not stated	Not stated	Not stated
Sathesh Raaj et al [54]	Hybrid CNN (combining radon transform and mathematical morphological segmentation)	MIAS and DDSM (mammogram)	MIAS: 99.17%, DDSM: 98.44%	MIAS: 98%, DDSM: 97.91%	MIAS: 98.66%, DDSM: 97.83%	Not stated explicitly
Schrammen et al [55]	SLAM (CNN-based, off-the-shelf)	Whole Slide Images (WSI)	Not stated	Not stated	Not stated	Tumor detection: 0.980 (CI: 0.975–0.984)MSI/dMMR: 0.909 (CI: 0.888–0.929)BRAF: 0.821 (CI: 0.786–0.852)Eksternal MSI: 0.900 (CI: 0.864–0.931)
Serte et al [56]	AlexNet + ResNet-18 + Gabor	Dermoscopic images	Not stated	Not stated	Not stated	83
Shafi et al [57]	DL-assisted SVM (CNN + SVM)	Lung CT Scan	94%	Not stated	Not stated	Not stated

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

Shakeel et al [58]	Improved DNN + Ensemble classifier	CT Images	High (No exact number)	High (No exact number)	High (No exact number)	Not stated
Shkloyar et al [59]	CNN (CystoNet)	Cystoscopy video	Not stated	Per-frame: 90.9%Per-tumor: 95.5%	98.6% (95% CI: 98.5–98.8%)	Not stated
Song et al [60]	CNN (DeepLab v3)	Whole Slide Image (WSI)	0.873–0.976	996	843	0.986–0.996
Togacar et al [61]	CNN (LeNet, AlexNet, VGG-16)	Chest CT scancs	99.51%	99.32%	99.71%	Not stated
Yan et al [62]	AE U-net + HDC (CNN-based)	Breast ultrasound image	95.81%	80.48%	Not stated	Not stated
Yoo et al [63]	LSTM (RNN)	EHR (electronic medical record)	Not stated	Not stated	Not stated	866
Yoo et al [64]	41-layer ResNet CNN + Random Forest (stacked ensemble)	Diffusion-weighted MRI (DWI)	Not stated	Not stated	Not stated	Slice-level: 0.87 (CI: 0.84–0.90), Patient-level: 0.84 (CI: 0.76–0.91)
Zhang et al [65]	GRU (Gated Recurrent Unit) + IOPA	Skin Lesion Images	0.95	0.95	0.97	Not stated
Zhou et al [66]	3D-CNN + MIL	DCE-MRI	Not stated	Not stated	Not stated	0.9

2.3. Subgroup Analysis

To explore the sources of variation across studies, subgroup analyses were performed by examining distinct methodological factors. The literature was categorized into four architectural groups based on their modeling frameworks. Core neural network architectures, such as standard convolutional neural networks (CNNs) and recurrent neural networks (RNNs), formed the foundational category. Task-specific deep learning models encompassed approaches tailored to particular applications, including semantic segmentation networks, domain-specific architectures, and object detection frameworks. Hybrid and ensemble learning models combined diverse methodologies, spanning ensemble techniques paired with traditional machine learning hybrids as well as integrative deep learning architectures. Finally, feature learning and representation models focused on advanced techniques like autoencoder-based and unsupervised models, alongside attention mechanisms and multi-scale modeling strategies. This classification enabled a structured investigation into how architectural choices influenced study outcomes.

To further evaluate the strengths and limitations of the data underpinning the studies, the research was also categorized according to the types of datasets employed. The literature was organized into four distinct groups based on data sources. The first group comprises radiology-based imaging, including modalities such as CT scans, MRI, and mammography. The second category focuses on pathology and microscopic imaging, covering histopathology samples and whole slide imaging techniques. The third

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

group, surface and real-time imaging, encompasses diverse datasets like thermoscopic and skin lesion images, infrared and thermographic imaging, ultrasound and endoscopy data, as well as oral and clinical images captured in real-world settings. Finally, the fourth category involves non-imaging clinical data, such as electronic health records (EHR) and structured tabular data. This classification allows for a nuanced analysis of how dataset characteristics, from imaging specificity to data format, influence model performance and generalizability in different clinical contexts.

To ensure a comprehensive analysis of study outcomes, the literature was additionally classified based on the evaluation metrics employed in each research effort. This categorization aimed to systematically assess how performance was quantified and compared across different methodologies. The metrics were organized into four primary groups: accuracy, which measures overall correctness in classification or prediction tasks; sensitivity, focusing on the ability to correctly identify true positive cases, particularly critical in diagnostic applications; DICE coefficient, a specialized metric for evaluating spatial overlap and segmentation quality in imaging tasks; and AUC (Area Under the Curve), which reflects the robustness of classification models across varying probability thresholds. By grouping studies according to these metrics, the analysis highlighted how different evaluation approaches emphasize distinct aspects of model performance from broad correctness and diagnostic precision to granular segmentation fidelity and threshold-invariant classification reliability. This framework enabled a deeper understanding of the strengths and limitations inherent to each metric in interpreting clinical or technical outcomes.

3. RESULT

The author starts the research by analysing the keyword of every literature related to deep learning-based cancer detection. VOSviewer is deployed by author to do this task. VOSviewer is a software tool used for visualizing and analysing bibliometric network. It's designed to help researchers to visualize and explore patterns in large volumes of scientific literature.

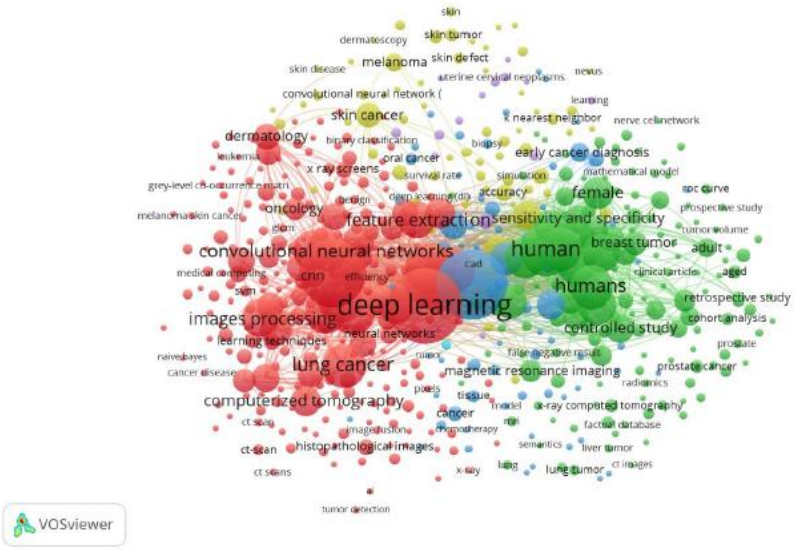

Figure 3. Literature Co-Occurance map on the role of deep learning in cancer detection

As we can see in Figure 3. that the keywords are divided into several clusters that are coloured differently. The red cluster, on the left, contains predominantly technical terms such as deep learning, CNN, image processing, and lung cancer. This shows the research focus on model approaches and architectures.

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

The green cluster, on the right, contains more keywords from clinical studies such as humans, controlled study, breast tumour, and prostate cancer. This reflects that many studies also relate the technology to real-world applications or patients. The blue cluster describes the types of images used, such as MRI, CT, and histopathological images. And other smaller clusters, such as purple and yellow, show specific topics such as skin tumours or other learning approaches such as k-nearest neighbour.

The size of the circles indicates the frequency of occurrence of the keywords, and the lines between the dots show how often the words appear together. From this visualization, we can see that deep learning is closely associated with various image processing techniques and certain types of cancer, as well as having a close relationship with clinical evaluations such as sensitivity, specificity, and accuracy.

The results of unweighted outcome values analysis show different results across different literatures. The outcome values of each study, as measured by metrics such as accuracy or sensitivity, varied significantly. This shows the performance of varying model on different type of cancer and datasets. The scatter plot (Figure 4.) gives visual representation of this variation.

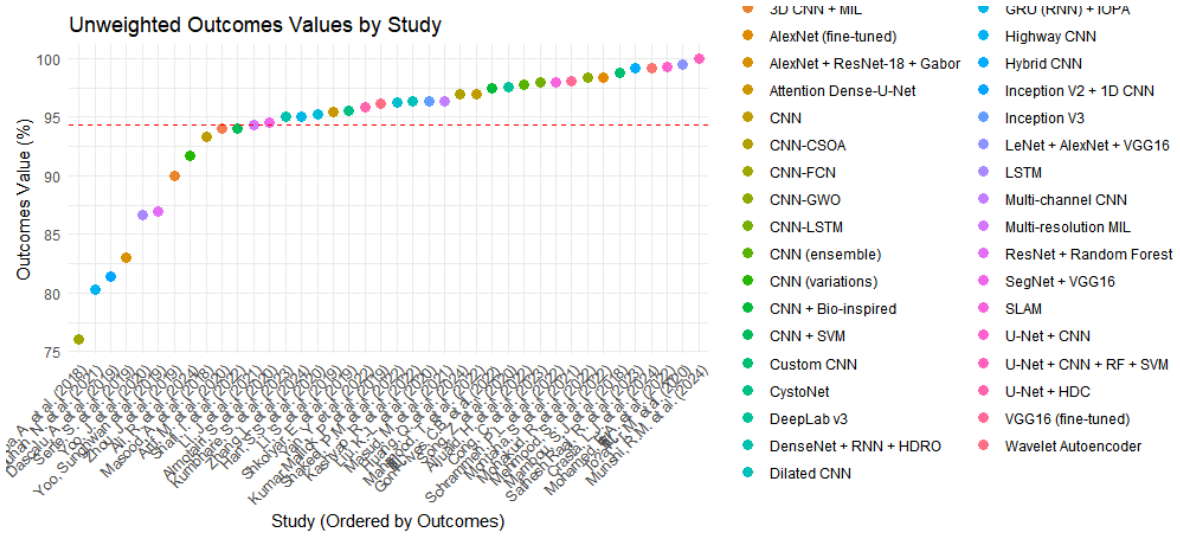

Figure 4. Unweighted outcomes values by study

In conclusion, unweighted outcome values analysis shows inconsistency and variability of model performance across research studies. This highlighted the need for further research into why certain models perform better under certain conditions.

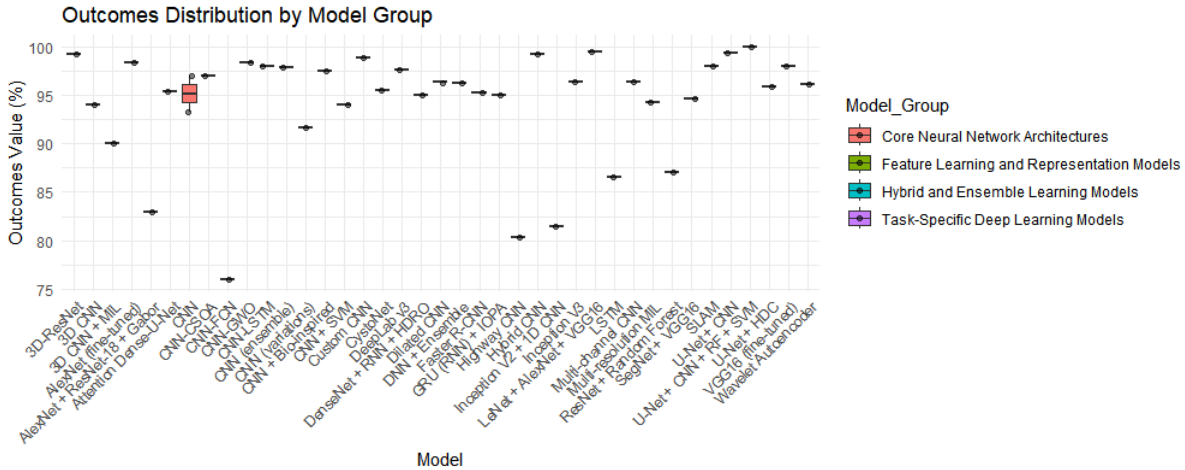

Figure 5. Outcome Distribution by Model Group

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

The Outcome Distribution by Model Group analysis reveals the performance of each model varies even among their own group (Figure 5). The boxplot depicts the range of outcome value to each model group, showing median, inter-quartile range (IQR) and outlier value. This highlighted that some model group such as Task-Specific Deep Learning Models and Hybrid and Ensemble Learning Models have wide range in performance, indicating that this model group may be useful in specific type of cancer or datasets. Conversely, other model group shows consistency with the range of outcome. However, this group still have some outliers that indicates that while the model of this group is in consistent in general. Some model may lag behind. In conclusion, while this outcome distribution may provide valuable insight into reliability of each model performance. It also suggests that certain model group may offer more stable result, while other group may still need for further improvement.

The Density Distribution of Outcome Values (Figure 6) shows a bimodal distribution with two different peaks. One is around 90 and the other one is near 95. This indicates that the models' performances are concentrated in two main ranges. The majority of the outcome values are clustered between 90 and 100. This suggest that most models performed well, with high accuracy or sensitivity scores. However, there is a noticeable spread in the lower range (around 75). This indicates that some models showed poor performance in certain cases. This distribution plot highlights the variability in model performance: some models shows high consistency, while others show more variation, leading to a broader distribution. Overall, the plot suggests that while most models perform well, a few exhibit significant variability in their results.

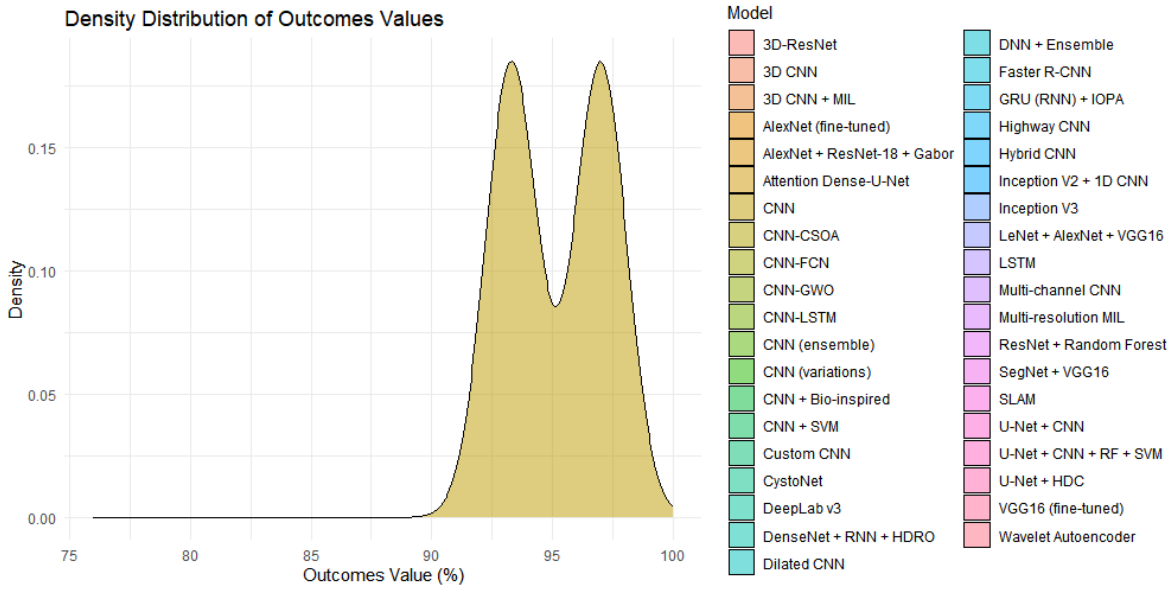

Figure 6. Density Distribution of Outcomes Values.

4. DISCUSSIONS

The results of this systematic review show important gaps in clinical translation while confirming the revolutionary potential of deep learning (DL) in cancer detection. In a variety of imaging modalities, including CT, MRI, mammography, and histopathology, deep learning models in particular, convolutional neural networks, or CNNs have continuously demonstrated good diagnostic accuracy. This supports earlier findings that DL outperforms conventional diagnostic techniques in pattern detection inside intricate image-based datasets.

The outcome scatter and box plots (Figures 4 and 5) illustrate the broad range of outcome values seen across trials, which implies that performance is still context-dependent. Interestingly, task-specific and hybrid models showed a wider range of performance, indicating their flexibility for certain cancer

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

kinds or imaging conditions. Although this adaptability is useful for specific tasks (such as multi-modal fusion or tumour segmentation), it also generates irregularities in overall performance and reproducibility.

Another important factor that determined the effectiveness of the model was dataset heterogeneity. Compared to studies that used non-imaging clinical data, such as electronic health records (EHRs), studies that used radiography and histopathology datasets found more consistent and superior performance. The difficulties in handling both structured and unstructured tabular data, where missing values, sparsity, and irregular formatting lower model reliability, are highlighted by the variation in EHR-based model results. In multi-institutional environments with widely disparate data standards, these problems are exacerbated.

Interpreting performance is made even more difficult by evaluation metrics like accuracy, AUC, and DICE. Even while strong classification across probability thresholds is indicated by high AUC scores, segmentation-centric tasks gain more from spatial measures such as DICE. A complex environment where a model's perceived success may change based on task requirements is revealed by this metric-dependent evaluation. While the majority of models cluster around high performance (90–95%), a portion still lags, especially in non-standard datasets or niche applications, according to the bimodal distribution of results (Figure 6).

Crucially, few studies discuss integration into clinical workflows, despite the fact that many claim great technical performance. It is alarming that there is a disconnect between clinical utility and research efficacy. Interpretability and generalisability continue to be significant challenges; DL models frequently operate as "black boxes," which undermines regulatory approval and reduces clinician trust. Furthermore, algorithmic bias and health disparities are ethical issues raised by the absence of consistent validation across patient populations, particularly those under-represented in training datasets.

Finally, although new architectures like ensemble models, multi-resolution learning, and attention mechanisms promise advancements, their practicality is frequently limited by the need for training data and computing complexity. In this regard, democratising cancer diagnosis may benefit greatly from lightweight DL models tailored for edge computing (such as portable imaging devices in low-resource environments).

5. LIMITATIONS AND FUTURE RESEARCH

5.1. Limitations

Our study’s focus on Q1 journals, while ensuring quality, may exclude impactful preprints or conference papers. Additionally, the rapid evolution of DL techniques means newer architectures (e.g., vision transformers) published post-2023 were not included. Nevertheless, our subgroup analysis categorizing studies by architecture, dataset, and metrics provides a novel framework for contextualizing performance, addressing a gap in earlier syntheses.

5.2. Future Research

Future research should prioritize enhancing the generalizability and robustness of deep learning (DL) models to ensure reliable performance across diverse clinical settings. This involves developing frameworks that adapt to heterogeneous datasets, such as multi-institutional medical imaging or varied electronic health record (EHR) formats, where domain adaptation and federated learning could mitigate performance drops in real-world scenarios. Architectural innovations, such as 3D-CNNs for volumetric tumor analysis or lightweight networks optimized for real-time detection in endoscopy, could further refine diagnostic precision while addressing computational inefficiencies.

Addressing data scarcity and bias remains pivotal, particularly for rare cancers or underrepresented populations. Strategies like synthetic data generation, transfer learning from well-

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

curated datasets, and ethical data curation practices could alleviate imbalances and improve model fairness. Concurrently, integrating explainable AI (XAI) techniques such as attention maps and saliency analysis into DL workflows is essential to demystify "black-box" decisions, fostering clinician trust and enabling transparent diagnostics in high-stakes environments.

The integration of multi-modal data, including imaging, genomics, and EHRs, presents an opportunity to advance personalized oncology. Holistic DL frameworks that synthesize these diverse inputs could unlock insights into tumor heterogeneity and treatment response, paving the way for tailored therapeutic strategies. Parallel efforts to streamline models for edge computing—such as deploying them on portable ultrasound devices—would bridge the gap between research and bedside application, particularly in resource-limited settings.

Finally, establishing ethical and regulatory frameworks is critical to guide the responsible adoption of DL in healthcare. Collaborative initiatives among researchers, clinicians, and policymakers must address challenges such as data privacy, algorithmic bias, and equitable access to ensure these technologies benefit all patient populations equitably. By addressing these priorities, future work can transform DL from a promising tool into a cornerstone of clinical oncology, driving earlier detection, reducing disparities, and improving survival outcomes globally.

6. CONCLUSION

This systematic review comprehensively analyzed the role of deep learning (DL) in cancer detection, focusing on architectural innovations, dataset diversity, and evaluation metrics. The findings underscore DL's transformative potential in oncology, particularly through its ability to process complex medical imaging data and deliver high diagnostic accuracy. Core neural network architectures, such as CNNs and RNNs, demonstrated robust performance across multiple cancer types, while task-specific and hybrid models (e.g., semantic segmentation networks, ensemble frameworks) excelled in specialized applications like histopathology and real-time imaging. However, performance variability was evident, influenced by factors such as dataset characteristics (e.g., radiology-based imaging vs. non-imaging EHR data) and metric selection (e.g., DICE for segmentation vs. AUC for classification robustness).

The analysis revealed that models trained on radiology-based imaging (CT, MRI) and pathology datasets (whole-slide histopathology) consistently achieved high accuracy (>90%) and sensitivity (>85%), whereas non-imaging data (EHR) lagged in performance due to inherent heterogeneity. Metrics like AUC and DICE highlighted the trade-offs between diagnostic precision and segmentation fidelity, emphasizing the need for context-specific evaluation. Despite these advancements, challenges persist in clinical applicability, including generalizability across diverse populations, interpretability of "black-box" models, and integration into routine workflows. The variability in study outcomes underscores the necessity for standardized reporting frameworks to ensure reproducibility and equitable implementation.

The impact of this study lies in its contribution as a current, high-level synthesis of the field, providing clarity on where deep learning excels, where it falls short, and what is needed to bridge the gap between technological potential and clinical utility. The review identifies consistent trends, methodological gaps, and architectural strengths that can serve as reference points for future research, policy-making, and AI-driven clinical tool development.

Future research should focus on improving model generalizability across heterogeneous clinical settings, enhancing transparency through explainable AI techniques, and developing lightweight, deployable frameworks that can function in real-time or low-resource environments. Collaborations between technical and clinical domains will be essential to translate DL innovations into practical, equitable cancer diagnostics.

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

REFERENCES

[1] F. Bray, M. Laversanne, and I. Weiderpass Elisabete and Soerjomataram, "The ever-increasing importance of cancer as a leading cause of premature death worldwide," *Cancer*, vol. 127, no. 16, pp. 3029–3030, Aug. 2021.

[2] J. Ferlay *et al.*, "Cancer statistics for the year 2020: An overview," *Int J Cancer*, vol. 149, no. 4, pp. 778–789, 2021.

[3] F. Bray, M. Laversanne, E. Weiderpass, and I. Soerjomataram, "The ever-increasing importance of cancer as a leading cause of premature death worldwide," *Cancer*, vol. 127, no. 16, pp. 3029–3030, 2021.

[4] M. Arif *et al.*, "Clinically significant prostate cancer detection and segmentation in low-risk patients using a convolutional neural network on multi-parametric MRI," *Eur Radiol*, vol. 30, no. 12, pp. 6582–6592, 2020.

[5] Y. Bengio, I. Goodfellow, and A. Courville, *Deep learning*, vol. 1. MIT press Cambridge, MA, USA, 2017.

[6] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," May 27, 2015, *Nature Publishing Group*. doi: 10.1038/nature14539.

[7] S. Kumbhare, A. B. Kathole, and S. Shinde, "Federated learning aided breast cancer detection with intelligent Heuristic-based deep learning framework," *Biomed Signal Process Control*, vol. 86, p. 105080, 2023.

[8] C. B. Gonçalves, J. R. Souza, and H. Fernandes, "CNN architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images," *Comput Biol Med*, vol. 142, p. 105205, 2022.

[9] R. Ali, A. Manikandan, R. Lei, and J. Xu, "A novel SpaSA based hyper-parameter optimized FCEDN with adaptive CNN classification for skin cancer detection," *Sci Rep*, vol. 14, no. 1, p. 9336, 2024.

[10] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. AlZain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework," *Sensors*, vol. 21, no. 3, p. 748, 2021.

[11] T. Mahmood, J. Li, Y. Pei, F. Akhtar, M. U. Rehman, and S. H. Wasti, "Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach," *PLoS One*, vol. 17, no. 1, p. e0263126, 2022.

[12] J. Li *et al.*, "A multi-resolution model for histopathology image classification and localization with multiple instance learning," *Comput Biol Med*, vol. 131, p. 104253, 2021.

[13] L. Zhang, J. Zhang, W. Gao, F. Bai, N. Li, and N. Ghadimi, "A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm," *Biomed Signal Process Control*, vol. 90, p. 105858, 2024.

[14] N. Chouhan, A. Khan, J. Z. Shah, M. Hussnain, and M. W. Khan, "Deep convolutional neural network and emotional learning based breast cancer detection using digital mammography," *Comput Biol Med*, vol. 132, p. 104318, 2021.

[15] R. M. Munshi, L. Cascone, N. Alturki, O. Saidani, A. Alshardan, and M. Umer, "A novel approach for breast cancer detection using optimized ensemble learning framework and XAI," *Image Vis Comput*, vol. 142, p. 104910, 2024.

[16] R. S. Raaj, "Breast cancer detection and diagnosis using hybrid deep learning architecture," *Biomed Signal Process Control*, vol. 82, p. 104558, 2023.

[17] E. A. Mohamed, E. A. Rashed, T. Gaber, and O. Karam, "Deep learning model for fully automated breast cancer detection system from thermograms," *PLoS One*, vol. 17, no. 1, p. e0262349, 2022.

[18] S. Montaha *et al.*, "BreastNet18: a high accuracy fine-tuned VGG16 model evaluated using ablation study for diagnosing breast cancer from enhanced mammography images," *Biology (Basel)*, vol. 10, no. 12, p. 1347, 2021.

[19] G. C. Forte *et al.*, "Deep learning algorithms for diagnosis of lung cancer: a systematic review and meta-analysis," *Cancers (Basel)*, vol. 14, no. 16, p. 3856, 2022.

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

[20] K. M. Kuo, P. C. Talley, and C.-S. Chang, "The accuracy of artificial intelligence used for non-melanoma skin cancer diagnoses: a meta-analysis," *BMC Med Inform Decis Mak*, vol. 23, no. 1, p. 138, 2023.
[21] R. Cuocolo *et al.*, "Machine learning for the identification of clinically significant prostate cancer on MRI: a meta-analysis," *Eur Radiol*, vol. 30, no. 12, pp. 6877–6887, 2020.
[22] E. Abbaspour *et al.*, "Machine learning and deep learning models for preoperative detection of lymph node metastasis in colorectal cancer: a systematic review and meta-analysis," *Abdominal Radiology*, pp. 1–15, 2024.
[23] K.-L. Liu *et al.*, "Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation," *Lancet Digit Health*, vol. 2, no. 6, pp. e303–e313, 2020.
[24] J. Yoo *et al.*, "Deep learning diffuse optical tomography," *IEEE Trans Med Imaging*, vol. 39, no. 4, pp. 877–887, 2019.
[25] P. L. Schrammen *et al.*, "Weakly supervised annotation-free cancer detection and prediction of genotype in routine histopathology," *J Pathol*, vol. 256, no. 1, pp. 50–60, 2022.
[26] K.-L. Liu *et al.*, "Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation," *Lancet Digit Health*, vol. 2, no. 6, pp. e303–e313, 2020.
[27] R. Ali, A. Manikandan, R. Lei, and J. Xu, "A novel SpaSA based hyper-parameter optimized FCEDN with adaptive CNN classification for skin cancer detection," *Sci Rep*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-57393-4.
[28] H. Aljuaid, N. Alturki, N. Alsubaie, L. Cavallaro, and A. Liotta, "Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning," *Comput Methods Programs Biomed*, vol. 223, 2022, doi: 10.1016/j.cmpb.2022.106951.
[29] S. Almotairi, G. Kareem, M. Aouf, B. Almutairi, and M. A.-M. Salem, "Liver tumor segmentation in CT scans using modified segnet," *Sensors (Switzerland)*, vol. 20, no. 5, 2020, doi: 10.3390/s20051516.
[30] M. Arif *et al.*, "Clinically significant prostate cancer detection and segmentation in low-risk patients using a convolutional neural network on multi-parametric MRI," *Eur Radiol*, vol. 30, no. 12, pp. 6582–6592, 2020, doi: 10.1007/s00330-020-07008-z.
[31] N. Chouhan, A. Khan, J. Z. Shah, M. Hussnain, and M. W. Khan, "Deep convolutional neural network and emotional learning based breast cancer detection using digital mammography," *Comput Biol Med*, vol. 132, 2021, doi: 10.1016/j.compbiomed.2021.104318.
[32] W. Cong, X. Intes, and G. Wang, "Optical tomographic imaging for breast cancer detection," *J Biomed Opt*, vol. 22, no. 9, 2017, doi: 10.1117/1.JBO.22.9.096011.
[33] L. J. Crasta, R. Neema, and A. R. Pais, "A novel Deep Learning architecture for lung cancer detection and diagnosis from Computed Tomography image analysis," *Healthcare Analytics*, vol. 5, 2024, doi: 10.1016/j.health.2024.100316.
[34] A. Cruz-Roa *et al.*, "High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection," *PLoS One*, vol. 13, no. 5, 2018, doi: 10.1371/journal.pone.0196828.
[35] A. Dascalu and E. O. David, "Skin cancer detection by deep learning and sound analysis algorithms: A prospective clinical study of an elementary dermoscope," *EBioMedicine*, vol. 43, pp. 107–113, 2019, doi: 10.1016/j.ebiom.2019.04.055.
[36] C. B. Gonçalves, J. R. Souza, and H. Fernandes, "CNN architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images," *Comput Biol Med*, vol. 142, 2022, doi: 10.1016/j.compbiomed.2021.105205.
[37] S. S. Han *et al.*, "Keratinocytic Skin Cancer Detection on the Face Using Region-Based Convolutional Neural Network," *JAMA Dermatol*, vol. 156, no. 1, pp. 29–37, 2020, doi: 10.1001/jamadermatol.2019.3807.
[38] Q. Huang, H. Ding, and N. Razmjooy, "Oral cancer detection using convolutional neural network optimized by combined seagull optimization algorithm," *Biomed Signal Process Control*, vol. 87, 2024, doi: 10.1016/j.bspc.2023.105546.

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

[39] R. Kashyap, "Dilated residual grooming kernel model for breast cancer detection," *Pattern Recognit Lett*, vol. 159, pp. 157–164, 2022, doi: 10.1016/j.patrec.2022.04.037.
[40] P. Kumar Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen, and P. Tiwari, "Brain MRI Image Classification for Cancer Detection Using Deep Wavelet Autoencoder-Based Deep Neural Network," *IEEE Access*, vol. 7, pp. 46278–46287, 2019, doi: 10.1109/ACCESS.2019.2902252.
[41] S. Kumbhare, A. B.Kathole, and S. Shinde, "Federated learning aided breast cancer detection with intelligent Heuristic-based deep learning framework," *Biomed Signal Process Control*, vol. 86, 2023, doi: 10.1016/j.bspc.2023.105080.
[42] J. Li *et al.*, "A multi-resolution model for histopathology image classification and localization with multiple instance learning," *Comput Biol Med*, vol. 131, 2021, doi: 10.1016/j.compbiomed.2021.104253.
[43] S. Li, M. Dong, G. Du, and X. Mu, "Attention Dense-U-Net for Automatic Breast Mass Segmentation in Digital Mammogram," *IEEE Access*, vol. 7, pp. 59037–59047, 2019, doi: 10.1109/ACCESS.2019.2914873.
[44] K.-L. Liu *et al.*, "Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation," *Lancet Digit Health*, vol. 2, no. 6, pp. e303–e313, 2020, doi: 10.1016/S2589-7500(20)30078-9.
[45] T. Mahmood, J. Li, Y. Pei, F. Akhtar, M. Ur Rehman, and S. H. Wasti, "Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach," *PLoS One*, vol. 17, no. 1 January, 2022, doi: 10.1371/journal.pone.0263126.
[46] S. J. Mambou, P. Maresova, O. Krejcar, A. Selamat, and K. Kuca, "Breast cancer detection using infrared thermal imaging and a deep learning model," *Sensors (Switzerland)*, vol. 18, no. 9, 2018, doi: 10.3390/s18092799.
[47] A. Masood *et al.*, "Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images," *J Biomed Inform*, vol. 79, pp. 117–128, 2018, doi: 10.1016/j.jbi.2018.01.005.
[48] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. Alzain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework," *Sensors (Switzerland)*, vol. 21, no. 3, pp. 1–21, 2021, doi: 10.3390/s21030748.
[49] S. Mehmood *et al.*, "Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning with Class Selective Image Processing," *IEEE Access*, vol. 10, pp. 25657–25668, 2022, doi: 10.1109/ACCESS.2022.3150924.
[50] R. Mohakud and R. Dash, "Designing a grey wolf optimization based hyper-parameter optimized convolutional neural network classifier for skin cancer detection," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 8, pp. 6280–6291, 2022, doi: 10.1016/j.jksuci.2021.05.012.
[51] E. A. Mohamed, E. A. Rashed, T. Gaber, and O. Karam, "Deep learning model for fully automated breast cancer detection system from thermograms," *PLoS One*, vol. 17, no. 1 January 2022, 2022, doi: 10.1371/journal.pone.0262349.
[52] S. Montaha *et al.*, "BreastNet18: A High Accuracy Fine-Tuned VGG16 Model Evaluated Using Ablation Study for Diagnosing Breast Cancer from Enhanced Mammography Images," *Biology (Basel)*, vol. 10, no. 12, 2021, doi: 10.3390/biology10121347.
[53] R. M. Munshi, L. Cascone, N. Alturki, O. Saidani, A. Alshardan, and M. Umer, "A novel approach for breast cancer detection using optimized ensemble learning framework and XAI," *Image Vis Comput*, vol. 142, 2024, doi: 10.1016/j.imavis.2024.104910.
[54] R. Sathesh Raaj, "Breast cancer detection and diagnosis using hybrid deep learning architecture," *Biomed Signal Process Control*, vol. 82, 2023, doi: 10.1016/j.bspc.2022.104558.
[55] P. L. Schrammen *et al.*, "Weakly supervised annotation-free cancer detection and prediction of genotype in routine histopathology," *Journal of Pathology*, vol. 256, no. 1, pp. 50–60, 2022, doi: 10.1002/path.5800.
[56] S. Serte and H. Demirel, "Gabor wavelet-based deep learning for skin lesion classification," *Comput Biol Med*, vol. 113, 2019, doi: 10.1016/j.compbiomed.2019.103423.

Vol. 6, No. 5, October 2025, Page. 3619-3634
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4748

[57] I. Shafi *et al.*, "An Effective Method for Lung Cancer Diagnosis from CT Scan Using Deep Learning-Based Support Vector Network," *Cancers (Basel)*, vol. 14, no. 21, 2022, doi: 10.3390/cancers14215457.

[58] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier," *Neural Comput Appl*, vol. 34, no. 12, pp. 9579–9592, 2022, doi: 10.1007/s00521-020-04842-6.

[59] E. Shkolyar *et al.*, "Augmented Bladder Tumor Detection Using Deep Learning," *Eur Urol*, vol. 76, no. 6, pp. 714–718, 2019, doi: 10.1016/j.eururo.2019.08.032.

[60] Z. Song *et al.*, "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning," *Nat Commun*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-18147-8.

[61] M. Toğaçar, B. Ergen, and Z. Cömert, "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," *Biocybern Biomed Eng*, vol. 40, no. 1, pp. 23–39, 2020, doi: 10.1016/j.bbe.2019.11.004.

[62] Y. Yan, Y. Liu, Y. Wu, H. Zhang, Y. Zhang, and L. Meng, "Accurate segmentation of breast tumors using AE U-net with HDC model in ultrasound images," *Biomed Signal Process Control*, vol. 72, 2022, doi: 10.1016/j.bspc.2021.103299.

[63] J. Yoo *et al.*, "Deep Learning Diffuse Optical Tomography," *IEEE Trans Med Imaging*, vol. 39, no. 4, pp. 877–887, 2020, doi: 10.1109/TMI.2019.2936522.

[64] S. Yoo, I. Gujrathi, M. A. Haider, and F. Khalvati, "Prostate Cancer Detection using Deep Convolutional Neural Networks," *Sci Rep*, vol. 9, no. 1, p. 19518, 2019, doi: 10.1038/s41598-019-55972-4.

[65] L. Zhang, J. Zhang, W. Gao, F. Bai, N. Li, and N. Ghadimi, "A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm," *Biomed Signal Process Control*, vol. 90, 2024, doi: 10.1016/j.bspc.2023.105858.

[66] J. Zhou *et al.*, "Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images," *Journal of Magnetic Resonance Imaging*, vol. 50, no. 4, pp. 1144–1151, 2019, doi: 10.1002/jmri.26721.