IJHOSCR

Review Article

International Journal of Hematology-Oncology and Stem Cell Research

Role of Three-Dimensional Convolution Neural Networks (3D- CNN) in Image Processing and Recognition in Oncology: A Systematic Review and Meta-Analysis

Sanjeev Kumar Jain, Shehzeen Afaq, Sonika Sharma

Department of Anatomy, Teerthanker Mahaveer Medical College and Research Centre, Teerthanker Mahaveer University Moradabad, Uttar Pradesh, India

Corresponding Author: Sanjeev Kumar Jain, Department of Anatomy, Teerthanker Mahaveer Medical College and Research Centre, Teerthanker Mahaveer University Moradabad, Uttar Pradesh, India
E-mail: jainsanjeevkumar77@gmail.com

Received: 31, Aug, 2025
Accepted: 18, Sep, 2025

ABSTRACT

Three-dimensional convolutional neural networks (3D CNNs) have transformed oncology imaging, excelling in tumor detection, classification, segmentation, and prognosis prediction. Unlike traditional two-dimensional CNNs, 3D CNNs effectively analyze volumetric medical imaging data, enhancing spatial feature extraction and diagnostic accuracy across modalities, including CT, MRI, PET, and ultrasound.

This systematic review and meta-analysis evaluates the diagnostic performance and clinical utility of 3D CNNs across 22 studies, of which 11 were eligible for quantitative synthesis. Pooled sensitivity, specificity, and AUC were 0.72, 0.73, and 0.77, respectively, with a diagnostic odds ratio of 10.38, indicating favorable discriminative ability. Subgroup analyses demonstrated superior accuracy in lung cancer and CT-based models, with DenseNet and ResNet architectures outperforming traditional CNNs.

Technical innovations—including multi-modal fusion, spatial context integration, and explainable AI techniques—enhance model robustness and clinician trust. However, substantial heterogeneity ($I^2 > 95\%$) across studies, attributable to differences in imaging protocols, dataset quality, and model design, underscores the need for standardized methodologies. Persistent challenges include computational demands, annotation variability, and generalization limitations.

Future directions should prioritize the integration of explainable AI, PACS-compatible user interfaces, and federated learning frameworks to bridge institutional gaps. This review highlights the considerable promise of 3D CNNs in advancing precision oncology, while also identifying the infrastructural and methodological refinements necessary for widespread clinical adoption.

Keywords: Three-dimensional convolutional neural networks; Oncology imaging; Cancer diagnostics; Tumor segmentation; Tumor classification

INTRODUCTION

The advancement of deep learning and artificial intelligence (AI) has revolutionized cancer imaging by making three-dimensional convolutional neural networks (3D CNNs) a powerful tool for picture processing and recognition[1]. By examining volumetric data, 3D CNNs outperform conventional two-dimensional (2D) CNNs in tumor detection, classification, and segmentation[2]. Their role in oncology is evaluated in this systematic review and meta- analysis, with a focus on cancer detection, segmentation, classification, and prognosis. Even

DOI: http://doi.org/10.18502/ijhoscr.v20i2.21784

though imaging methods such as CT, MRI, and PET are used extensively in cancer, traditional methods have interobserver variability and inefficiencies. CNNs and other AI-driven techniques enhance automated analysis[3]. 3D CNNs capture spatial connections, increasing diagnostic accuracy[4,5], in contrast to 2D CNNs utilized in lung cancer screening[6] and breast cancer screening[7].

Applications of 3D CNNs in Oncological Imaging

Three-dimensional convolutional neural networks (3D CNNs) have demonstrated remarkable performance in cancer detection, classification, and segmentation by effectively processing volumetric imaging data and extracting complex spatial features. Their ability to integrate multi-slice information enhances diagnostic precision across various cancer types. In lung cancer, for instance, 3D CNNs significantly improve the identification of malignant nodules in CT images, surpassing conventional radiological assessments through accurate tumor localization[8]. Advanced techniques such as gradient-weighted class activation mapping further reduce false positives and enable comprehensive tumor morphology assessment across entire volumetric datasets [4, 9].

In breast cancer diagnosis, 3D CNNs effectively distinguish benign from malignant lesions in MRI and mammography, with feature selection strategies boosting classification accuracy and enabling early, reliable detection[7]. When combined with image fusion techniques, 3D CNNs also improve ultrasound-based classification performance, underscoring their utility in breast cancer screening[10,11]. In liver cancer, these models predict microvascular invasion in hepatocellular carcinoma—a critical factor for treatment planning and prognosis—and outperform traditional methods in MRI-based tumor differentiation by capturing intricate spatial relationships [12, 2].

For brain tumors, 3D CNNs achieve high accuracy in glioma classification and segmentation on MRI, facilitating precise tumor boundary delineation for radiotherapy and surgical planning[13]. Integrating long-range 2D context into 3D architectures further refines segmentation outcomes and ensures reliable tumor identification [14]. The growing adoption of 3D CNNs has been driven by advances in CT, MRI, and PET imaging, which automate labor-intensive processes and reduce interobserver variability[1,15]. Their integration into clinical workflows improves lesion detection, minimizes diagnostic delays, and enhances treatment decisions across multiple cancer types[16–18]. As 3D CNNs continue to evolve, they are transforming oncological imaging by delivering unprecedented accuracy in cancer diagnosis, prognostication, and therapy planning.

3D CNNs in Tumor Segmentation and Treatment Planning

Precise tumor segmentation is essential for radiation therapy and surgical planning, and 3D CNNs have substantially improved segmentation accuracy and spatial delineation across multiple malignancies. In brain tumor imaging, 3D CNN-based models leveraging MRI data have surpassed traditional approaches by enhancing tumor boundary delineation[14]. The integration of long-range 2D context into 3D architectures has further refined segmentation performance [19].

In liver cancer, deep learning-based 3D CNNs applied to MRI have provided critical prognostic insights by predicting microvascular invasion in hepatocellular carcinoma[12], while dual-scale 3D CNNs have outperformed conventional methods in tumor delineation[20]. Similarly, in cervical cancer, automated tumor segmentation in CT images enabled by 3D deep neural networks has increased the precision of radiation planning [21].

Despite their transformative impact, 3D CNNs face persistent challenges related to model generalization, data availability, and computational complexity. Widespread clinical adoption will require large, multi-center datasets and improved model interpretability. Nevertheless, AI-driven decision support systems incorporating 3D CNNs have shown promise in multi-center cancer diagnosis, advancing AI-assisted clinical workflows. Ongoing developments in hardware infrastructure and deep learning algorithms are gradually addressing these limitations.

This systematic review evaluates the significance of 3D CNNs in cancer detection, classification, and

segmentation, as well as their broader relevance in oncological imaging.

Despite significant advancements, barriers to the widespread clinical adoption of 3D CNNs persist, including limited multi-center validation, poor model generalization, and variability in dataset quality. An objective synthesis of existing evidence is therefore essential to critically evaluate their diagnostic accuracy and clinical applicability.

This meta-analysis systematically synthesizes findings from diverse studies to assess the effectiveness of 3D CNNs in oncology imaging, with a focus on their potential to improve cancer diagnostics and treatment planning. In doing so, it also identifies critical challenges that must be addressed for successful clinical implementation. Given that recent reviews have highlighted the growing role of 3D CNNs in multi-modal integration and radiogenomics[2], a consolidated, up-to-date synthesis of performance metrics is urgently needed—a gap this study aims to fill.

MATERIALS AND METHODS

Research Question

What is the role of three-dimensional convolutional neural networks (3D CNNs) in oncology imaging?

The research question was structured using the PICOS framework as follows:

- Population (P): Patients with various malignancies, including ovarian tumors, cervical cancer, head and neck squamous cell carcinoma, lung cancer, oral cancer, breast cancer, liver tumors, and hepatocellular carcinoma, examined across multiple imaging modalities.
- Intervention (I): Application of 3D CNNs to CT, MRI, ultrasound, PET, and microscopic imaging for tumor detection, localization, classification, segmentation, and prognosis prediction.
- Comparison (C): 3D CNN performance compared with standard 2D CNNs, conventional machine learning techniques, radiomics-based models, traditional imaging-based diagnostic methods, and human radiologists or pathologists.
- Outcomes (O): Diagnostic accuracy, classification performance (accuracy, sensitivity, specificity, AUC-ROC), segmentation efficacy (Dice score, IoU), prognostic prediction (C-index, recurrence prediction), tumor localization precision, and overall clinical workflow efficiency.
- Study Design (S): Retrospective studies, experimental studies, cohort studies, and deep learning-based evaluations using internal/external test sets or cross-validation methods.

Search Strategy

Following PRISMA 2020 guidelines, a comprehensive literature search was conducted across PubMed, Scopus, Web of Science, Embase, Cochrane Library, and Google Scholar for studies published between 2017 and 2024 (Figure 1). The search approach was improved by using Boolean operators (AND, OR) to make sure pertinent papers were found. "3D CNN," "three-dimensional convolutional neural network," "deep learning in oncology," "AI in cancer imaging," "tumor segmentation using AI," "oncology image classification," "radiomics and deep learning," "medical image analysis with CNNs," "MRI-based cancer detection," "CT-based tumor identification," "PET imaging AI," and "computer-aided diagnosis in oncology" were among the keywords used. Only peer-reviewed English-language research on quantitative performance measures in cancer imaging was included in the search.

Eligibility Criteria

Studies were included if they applied 3D CNNs to oncology imaging, reported quantitative performance metrics, evaluated diagnostic accuracy, segmentation precision, or classification efficacy Eligible imaging modalities comprised CT, MRI, PET, and digital pathology. Exclusion criteria included studies without a clear methodology or lacking performance metrics, non-English publications, case reports, reviews, and editorials. Additionally, articles that solely used 2D CNNs without volumetric analysis were excluded. This ensured that only relevant studies focusing on the application of 3D CNNs in cancer imaging were considered, providing a comprehensive assessment of their effectiveness in tumor detection, classification, and segmentation.

To better interpret the diagnostic performance of 3D CNNs in oncology imaging, subgroup analyses were conducted based on cancer type, imaging modality, and CNN architecture.

1. Cancer Type: Subgroup-specific sensitivity and specificity revealed that models applied to lung cancer imaging (n = 5 studies) achieved the highest pooled sensitivity (0.81, 95% CI: 0.74–0.88) and specificity (0.84, 95% CI: 0.76–0.90). Cervical cancer models showed moderate sensitivity (0.71, 95% CI: 0.61–0.81) but lower specificity (0.58, 95% CI: 0.44–0.72). Breast cancer models demonstrated balanced performance (sensitivity: 0.77, specificity: 0.72).

2. Imaging Modality: CT-based 3D CNNs showed higher diagnostic performance (AUC: 0.81, 95% CI: 0.72–0.89) compared to MRI-based (AUC: 0.74, 95% CI: 0.63–0.84) and ultrasound-based models (AUC: 0.69, 95% CI: 0.57–0.81).

3. CNN Architecture: Architectures like DenseNet and ResNet variants achieved superior pooled AUCs (DenseNet: 0.84; ResNet: 0.81), whereas traditional vanilla CNNs demonstrated lower AUCs (0.68–0.72), particularly when applied to complex classification tasks.

Diagnostic Odds Ratio (DOR)

The pooled diagnostic odds ratio (DOR) across studies using 3D CNNs in oncology imaging was estimated to be 10.38 (95% CI: 6.17–17.47). This indicates that patients with cancer were over 10 times more likely to be correctly identified by 3D CNN-based systems compared to those without cancer. Despite notable heterogeneity ($I^2 = 95\%$), the overall effect remained statistically significant ($Z = 7.11$, $p < 0.00001$).

Architectural Approaches and Technical Innovations

The reviewed studies employed a range of CNN architectures that vary in complexity, feature representation capacity, and diagnostic accuracy. Prominent models include:

3D U-Net: Utilized primarily for segmentation tasks, such as cervical and brain tumor delineation. U-Net-based models excel in capturing fine spatial details and boundary definitions due to their encoder-decoder structure and skip connections.

3D DenseNet: Frequently used for breast and lung cancer classification, DenseNet architectures allow efficient feature propagation and reuse by connecting each layer to every other layer. This design enhances gradient flow and reduces the number of parameters.

3D ResNet and ResNet Variants: Applied for classification and mutation prediction, ResNet-based architectures incorporate residual blocks that mitigate vanishing gradient issues, allowing the training of deeper networks.

Dual-Scale 3D CNNs: As demonstrated in liver tumor segmentation, these models process image features at both global and local scales, improving lesion detection and structural representation.

In addressing common challenges, several strategies were identified:

Class Imbalance: Resampling techniques, such as weighted loss functions and data augmentation (e.g., random flipping, cropping, rotation), were implemented to balance datasets. Some studies used synthetic minority oversampling or dropout regularization to enhance generalization.

Spatial Feature Capture: Long-range 2D context was integrated within 3D CNNs to enhance spatial representation, particularly in glioma segmentation and liver tumor classification.

Multi-Modal Fusion: Studies combining imaging modalities (e.g., CT + PET, multiparametric MRI) leveraged complementary structural and functional information, resulting in improved classification accuracy. Fusion was achieved through channel-wise concatenation or ensemble learning strategies.

These architectural advancements and technical methods address key limitations in clinical imaging applications, enhancing model robustness, generalizability, and diagnostic interpretability.

Study Selection and Data Extraction

A comprehensive search across multiple databases yielded a total of 381 articles, of which 22 met the eligibility criteria for inclusion in the systematic review (Figure 1). Among these, only 11 studies qualified for inclusion in the meta- analysis. The study selection process was conducted in two phases. Initially, two independent reviewers

(Reviewer 1 and Reviewer 2) screened the titles and abstracts to identify relevant studies. Full-text reviews were then performed on eligible articles, with any disagreements resolved through discussion or consultation with a third reviewer (Reviewer 3).

Risk of Bias and Quality Assessment

The risk of bias was evaluated using standardized tools depending on research design. To assess biases associated with missing data, intervention adherence, randomization, and selective reporting, the ROB2 tool was used in experimental investigations. The ROBINS-I[22] approach was used to evaluate observational research, with an emphasis on confounding issues, data classification, and participant selection. The Newcastle-Ottawa Scale (NOS)[23] was utilized for retrospective studies, while SYRCLE's Risk of Bias Tool[24] was also applied to experimental studies. Heterogeneity was examined through sensitivity analyses, and publication bias was investigated using funnel plots. The study adhered to PRISMA guidelines to ensure methodological rigor and transparency, enhancing the reliability and validity of the findings.

RESULT

Table 1 summarizes 22 studies investigating the role of 3D CNNs in oncology imaging across various malignancies, including lung cancer[25,27,35], breast cancer[38, 29], and cervical cancer[21,32]. CT was the most frequently used imaging modality, followed by MRI, ultrasound, and microscopy. The primary applications of 3D CNNs included tumor classification, segmentation, and prognosis. Several studies demonstrated superior performance of 3D CNNs compared to traditional 2D CNNs, particularly in classification accuracy and diagnostic efficiency[16,26]. Notable findings include improved tumor localization accuracy (<0.9 mm error)[27] and enhanced classification performance with preprocessing techniques[10]. Some models achieved exceptionally high diagnostic accuracy, exceeding 90% in specific applications[21,31]. These findings highlight the potential of 3D CNNs in improving cancer diagnosis and patient outcomes. Several studies report high classification accuracy, such as Mzoughi et al. (98.38% for NSCLC)[13], Abinaya K et al. (98.6% for cervical cancer)[21], and Chithra PL et al. (98% for lung tumor segmentation)[35]. 3D CNN models often outperform 2D CNNs, as was seen in some studies[16, 26]. Studies on tumor localization and prognosis demonstrate precise segmentation and predictive capabilities. Advanced architectures, including 3D DenseNet and Vision Transformer-based models (K A et al.), show significant improvements. Most studies validate results using cross-validation or external test sets, with statistical significance reported in several cases ($p < 0.05$), reinforcing the robustness of 3D CNN applications in medical imaging (Table 2).

Figure 2 presents the risk of bias assessment for experimental studies using the ROB2 tool. Xu S et al. (2019) [16] demonstrated a low risk of bias across most domains; however, missing outcome data led to a moderate overall risk. Similarly, Zhang Z et al. (2019)[26] showed low bias but had concerns regarding selective reporting, potentially affecting result reliability. Wei R et al. (2019)[27]exhibited issues in randomization, missing data, and selective outcome reporting, suggesting weaknesses in transparency. Muñoz-Aseguinolaza U et al. (2023)[4] had the least bias, with only minor concerns about missing data, indicating a well-structured study. Figure 3 provides an overall summary, showing that while most studies maintained low bias in intervention adherence and randomization, missing data and selective reporting were prevalent, impacting result reliability. Figure 4 evaluates observational studies using the ROBINS-I tool. Alakwaa W et al. (2017)[25] and Perez G et al. (2020)[18] had a low risk of bias, though selective reporting remained a concern. Trivizakis E et al. (2019)[2] and Ding Y et al. (2022)[19] had missing details regarding participant selection and classification, affecting credibility. Camalan S et al. (2021)[29] and Zhang Y et al. (2021)[31] exhibited moderate bias due to confounding variables. Tan X et al. (2021)[32] and Huang X et al. (2022)[17] had concerns regarding missing data and intervention classification.

Table

Table 1: Characteristics of the Studies Included

Table 1

Author/Year	Study Design	Sample Size	Malignancy Type	Imaging Modality	3D CNN Application	Key Outcome Measure
1. Alakwaa W et al. (2017) [25]	Retrospective Analysis	1,000 patients; 1,397 CT scans; 248,580 slices	Lung Cancer	CT scans	Pulmonary nodule detection and classification	Accuracy: 86.6%; Misclassification rate: 13.4%; False positive rate: 11.9%; False negative rate: 14.7%
2. Xu S et al. (2019) [16]	Experimental Study	~7,000 CT images	Oral Cancer	Dynamic Contrast-Enhanced CT (DCE-CT)	Early diagnosis of benign/malignant oral cancer lesions	3DCNN outperformed 2DCNN, improving classification accuracy by over 6%; AUC increased to 0.801 in enhancement rate imaging
3. Trivizakis E et al. (2019) [2]	Retrospective Study	130 patients	Primary (HCC, Cholangiocarcinoma, Hepatic Adenocarcinoma) and Metastatic (from various primary sites)	MRI (DW-MRI) (b1000)	Classification of primary vs. metastatic liver tumors	3D CNN outperformed 2D CNNs, achieving 85.5% accuracy in distinguishing primary vs. metastatic liver tumors
4. Zhang Z et al. (2019) [26]	Experimental	1,920 tumor spheres	Breast Cancer (SUM159 cells)	Bright-field Microscopy	Classification, Prediction of Drug Inhibition Score	CNN predicted drug treatment concentration with 93.2% accuracy; IC50 estimation difference of 18% compared to LIVE/DEAD staining.
5. Wei R et al. (2019) [27]	Experimental	100,000 simulated images and data from three real patients for testing	Lung Cancer	Single X-ray Projection, CBCT, DRR	Tumor Localization & Volumetric Reconstruction	CNN-based method showed improved tumor localization accuracy (<0.9 mm error) compared to MM and MM-FD methods. Faster real-time tracking (36 ms per projection).
6. Perez G et al. (2020) [18]	Retrospective study	1890 subjects (1384 training, 506 testing)	Lung Cancer	Low-dose CT LDCT (DICOM Format)	Nodule Detection and Cancer Prediction	Ranked 1st in ISBI 2018 Challenge, High Sensitivity in Nodule Detection, 4% Performance Improvement with Feature Extraction
7. Mzoughi et al. (2020) [13]	Retrospective study	284 patients (160 training, 62 validations, 62 testing)	Non-small cell lung cancer (NSCLC)	CT	Classification	
8. Starke S et al. (2020) [28]	Retrospective	291 patients (206 exploratory, 85 validation)	Head and Neck Squamous Cell Carcinoma (HNSCC)	CT	Prognosis	Concordance Index (C-index), Loco-regional Tumor Control (LRC)
9. Camalan S et al. (2021) [29]	Retrospective Analysis	54 patients (30 UK, 24 Brazil)	Oral Dysplasia and Oral Squamous Cell Carcinoma	Photographic Images	Classification	Developed deep learning model for early detection of oral dysplasia with Class Activation Map analysis

Author/Year	Study Design	Sample Size	Malignancy Type	Imaging Modality	3D CNN Application	Key Outcome Measure
10.Yanagawa M et al. (2021) [30]	Retrospective Study	285 patients (75 AIS, 58 MIA, 152 IVA)	Pulmonary adenocarcinoma (AIS, MIA, IVA)	Chest CT	Classification of EGFR mutation status	3D-CNN model is a non-invasive method for predicting pulmonary invasive adenocarcinoma in CT images with high sensitivity. Diagnostic accuracy by less-experienced radiologists improved with the 3D- CNN model support.
11. Zhang Y et al. (2021) [31]	Retrospective	237 (158 training, 79 validation)	Hepatocellular Carcinoma (HCC)	Multiparametric MRI (T2WI, T2-SPIR, PVP)	Prediction of Microvascular Invasion (MVI)	Fusion model achieved AUC 0.81 (training) and 0.72 (validation), with sensitivity 69% and specificity 79% (training), sensitivity 55% and specificity 81% (validation)
12. Tan X et al. (2021) [32]	Retrospective, multicenter, multicohort study	16,000 TCT Images (Training: 13,775, Validation: 2,301, Test: 408,030 images from 290 slides)	Cervical Cancer (ASCUS, LSIL, HSIL)	ThinPrep Cytologic Test (TCT)	Faster R-CNN	Sensitivity: 99.4%, Specificity: 34.8%, AUC-ROC: 0.67, Reduced workload for pathologists, High efficiency in classification
13. Huang X et al. (2022) [17]	Retrospective Study	1074 patients	Non-Small Cell Lung Cancer	CT	Prediction of EGFR mutation status	Radiomic and deep learning models effectively predict EGFR gene mutation status non-invasively
14. Tao J et al. (2022) [33]	Retrospective Cohort Study	203 patients (112 men, 91 women)	Non-small cell lung cancer (NSCLC)	Contrast-enhanced CT	Prediction of spread through air spaces (STAS)	The convolutional neural network (CNN) model demonstrated superior performance in predicting STAS status compared to conventional radiomics and computer vision models.
15. Ding Y et al. (2022) [19]	Technical Study	130 (90 training, 10 validation, 30 testing)	Cervical Cancer	CT	Segmentation	V-Net model achieved better segmentation performance than U-Net for CTV and OARs; improved efficiency in radiation therapy workflow
16. Jung Y et al. (2022) [34]	Retrospective Cohort Study	1613 ultrasound images from 1154 patients	Ovarian tumors (benign: cystadenoma, mature cystic teratoma, endometrioma; malignant: borderline/cancer)	Ultrasound	CNN with Convolutional Autoencoder (CNN- CAE)	Accuracy, Sensitivity, Specificity, AUC-ROC, Grad-CAM visualization
17. Cui H et al. (2022) [3]	Retrospective	220 (178 training, 42 test)	Hepatocellular Carcinoma (HCC)	Contrast-Enhanced CT	Prediction of Early Recurrence (ER) (≤2 years)	SS-3D-CNN outperforms NSS-3D- CNN in predicting ER; AUC for SS-3D-CNN in test group = 0.789, NSS- 3D-CNN = 0.560; SS-3D-CNN significantly stratifies patients into low- and high-risk groups (P < .001)

Author/Year	Study Design	Sample Size	Malignancy Type	Imaging Modality	3D CNN Application	Key Outcome Measure
18. Alotaibi et al. (2023) [10]	Retrospective Study	BUSI (780), Dataset B (162), KAIMRC (5693)	Breast Cancer (Benign, Malignant, Normal)	Ultrasound (US)	Classification using VGG19	Enhanced classification accuracy with preprocessing (denoising, ROI highlighting, RGB fusion). VGG19 outperformed EfficientNet V2 (Accuracy: 91% vs. 67%). The model was evaluated using five-fold cross-validation.
19. Zhou J et al. (2023) [11]	Retrospective Study	1537 female cases (mean age 47.5 ± 11.8 years)	Breast Cancer	3D DCE-MRI	Classification, Weakly Supervised Localization	Achieved accuracy of 83.7%, sensitivity of 90.8%, specificity of 69.3%; Comparable performance with radiologists; Lowered overdiagnosing rate by 10%; Weakly supervised localization with Dice similarity coefficient of 0.501 ± 0.274
20. Muñoz-Aseguinolaza U et al. (2023) [4]	Experimental Study	436 studies (251,135 images) from 355 patients (Lung-Pet-Ct-Dx dataset)	Lung Cancer	CT, PET	Detection, Classification	Model performance affected by class imbalance; Proposed resampling methods to improve detection
21. Chithra PL et al. (2024) [35]	Retrospective study	492 NSCLC patients (16,208 training, 1,012 validation, 1,664 test)	Non-small cell lung cancer	CT (DICOM and NIfTI formats)	Tumor segmentation	Dice coefficient, IoU, Sensitivity, F1-score, Accuracy, 10% improvement over SOTA models
22. Abinaya K et al. (2024) [21]	Deep Learning-Based Approach	917 cervical images	Cervical Cancer (Normal, Severe Dysplastic, Light Dysplastic, Carcinoma in Situ, Moderate Dysplastic)	Herlev Pap Smear Dataset	Classification	Achieved 98.6% accuracy in cervical cancer classification using 3D CNN, Vision Transformer, and KELM

Author/Year	Primary Outcome and Secondary outcome	Efficiency/ Performance	Performance Metric (Accuracy, Sensitivity, Specificity, Dice Score, AUC-ROC)	3D CNN Application (Detection, Classification, Segmentation, Prognosis)	Model Used (e.g., CNN Type, Algorithm)	Validation Method (e.g., Cross-validation, External Test Set)	Statistical Significance (e.g., p- value, CI)
Alakwaa W et al. (2017) [25]	Primary Outcome: Classification of lung cancer; Secondary Outcome: Nodule detection efficiency	Accuracy: 86.6%	Accuracy: 86.6%, Sensitivity: 85.3%, Specificity: 88.1%, AUC-ROC: 0.83	Detection and Classification	3D CNN (Vanilla CNN, GoogLeNet-based 3D CNN)	30% test set validation	Not explicitly stated
Xu S et al. (2019) [16]	Primary: Classification of oral cancer as benign/malignant; Secondary: Comparison of 3DCNN vs. 2DCNN	High 3DCNN outperformed 2DCNN by >6%	AUC: 0.801 (3DCNN), 0.754–0.796 (enhanced images); Sensitivity: 0.818; Specificity: 0.739	Classification of benign/malign ant oral cancer lesions using spatial and dynamic features from CT images for improved diagnostic accuracy.	3DCNN vs. 2DCNN	Data augmentation, cross-validation	Not explicitly stated
Trivizakis E et al. (2019) [2]	Primary: Classification of liver tumors (primary vs. metastatic) Secondary: Feature extraction for CAD system development	High 3D CNN: 85.5% accuracy 2D CNN (Original Slice): 71.9% accuracy 2D CNN (Patch-wise): 65.2% accuracy		Classification	3D CNN with voxel-wise feature extraction	Stratified random sampling, Training/Validation/ Test split (74/33/23 patients)	Improvement in classification performance was statistically significant
Zhang Z et al. (2019) [26]	Primary: Estimation of sphere viability using bright-field images. Secondary: Drug efficacy prediction (IC50 values)	93.2% accuracy for classification; IC50 estimation difference of 18%	Correlation coefficient R > 0.84 for all three drugs (Doxorubicin, Oxaliplatin, Irinotecan). Classification accuracy: 93.2% for Doxorubicin	Classification, Prediction	CNN	80% training, 20% testing	Mean difference in IC50s: 18%. R-values: 0.84 (Doxorubici n), 0.89 (Oxaliplatin), 0.91 (Irinotecan)
Wei R et al. (2019) [27]	Primary: Tumor localization accuracy; Secondary: Diaphragm height estimation	Mean localization error: 0.71 mm (X), 0.76 mm (Y); Max localization error <0.9 mm	Max error X, Y: <0.9 mm, Mean error single patient: <0.8 mm, Mean error all: 0.71 mm (X), 0.76 mm (Y), RMSE CNN: <0.4%, RMSE MM, MM-FD: >3%, Diaphragm error CNN: 0.8 mm.	Localization	Convoluti ona l Neural Network (CNN) Regression Model	Comparison with MM and MM-FD methods, Real Patient Data Testing	Mean error < 0.2 mm in 3D space. RMSE < 0.4%.

Table 2: Quantitative Outcomes Supporting the Meta-Analysis

Author/Year	Primary Outcome and Secondary outcome	Efficiency/ Performance	Performance Metric (Accuracy, Sensitivity, Specificity, Dice Score, AUC-ROC)	3D CNN Application (Detection, Classification, Segmentation, Prognosis)	Model Used (e.g., CNN Type, Algorithm)	Validation Method (e.g., Cross-validation, External Test Set)	Statistical Significance (e.g., p- value, CI)
Alakwaa W et al. (2017) [25]	Primary Outcome: Classification of lung cancer; Secondary Outcome: Nodule detection efficiency	Accuracy: 86.6%	Accuracy: 86.6%, Sensitivity: 85.3%, Specificity: 88.1%, AUC-ROC: 0.83	Detection and Classification	3D CNN (Vanilla CNN, GoogLeNet-based 3D CNN)	30% test set validation	Not explicitly stated
Xu S et al. (2019) [16]	Primary: Classification of oral cancer as benign/malignant; Secondary: Comparison of 3DCNN vs. 2DCNN	High 3DCNN outperforme d 2DCNN by >6%	AUC: 0.801 (3DCNN), 0.754–0.796 (enhanced images); Sensitivity: 0.818; Specificity: 0.739	Classification of benign/malign ant oral cancer lesions using spatial and dynamic features from CT images for improved diagnostic accuracy	3DCNN vs. 2DCNN	Data augmentation, cross-validation	Not explicitly stated
Trivizakis E et al. (2019) [2]	Primary: Classification of liver tumors (primary vs. metastatic) Secondary: Feature extraction for CAD system development	High 3D CNN: 85.5% accuracy 2D CNN (Original Slice): 71.9% accuracy 2D CNN (Patch-wise): 65.2% accuracy		Classification	3D CNN with voxel-wise feature extraction	Stratified random sampling, Training/Validation/ Test split (74/33/23 patients)	Improvement in classification performance was statistically significant
Zhang Z et al. (2019) [26]	Primary: Estimation of sphere viability using bright-field images. Secondary: Drug efficacy prediction (IC50 values)	93.2% accuracy for classificatio n; IC50 estimation difference of 18%	Correlation coefficient R > 0.84 for all three drugs (Doxorubicin, Oxaliplatin, Irinotecan). Classification accuracy: 93.2% for Doxorubicin	Classification, Prediction	CNN	80% training, 20% testing	Mean difference in IC50s: 18%. R-values: 0.84 (Doxorubicin) , 0.89 (Oxaliplatin), 0.91 (Irinotecan)
Wei R et al. (2019) [27]	Primary: Tumor localization accuracy; Secondary: Diaphragm height estimation	Mean localization error: 0.71 mm (X), 0.76 mm (Y); Max localization error <0.9 mm	Max error X, Y: <0.9 mm, Mean error single patient: <0.8 mm, Mean error all: 0.71 mm (X), 0.76 mm (Y), RMSE CNN: <0.4%, RMSE MM, MM-FD:>3%, Diaphragm error CNN: 0.8 mm.	Localization	Convolutiona l Neural Network (CNN) Regression Model	Comparison with MM and MM-FD methods, Real Patient Data Testing	Mean error < 0.2 mm in 3D space. RMSE < 0.4%.
Perez G et al. (2020) [18]	Nodule Detection, Cancer Classification	High Sensitivity and Precision	AUC: 0.913	Classification	5-Way 3D CNN with Multi-Path Network	External Test Set (ISBI 2018), Internal Validation on LIDC-IDRI & Kaggle DSB 2017	Ranked 1st in ISBI 2018 challenge

Author/Year	Primary Outcome and Secondary outcome	Efficiency/ Performance	Performance Metric (Accuracy, Sensitivity, Specificity, Dice Score, AUC-ROC)	3D CNN Application (Detection, Classification, Segmentation, Prognosis)	Model Used (e.g., CNN Type, Algorithm)	Validation Method (e.g., Cross-validation, External Test Set)	Statistical Significance (e.g., p- value, CI)
Mzoughi et al. (2020) [13]	Classifying NSCLC into adenocarcinoma and squamous cell carcinoma		Accuracy: 98.38%, Sensitivity: 98.11%, Specificity: 98.75%, Precision: 98.77%, F1 Score: 98.44%, AUC: 0.996	Classification	3D-CNN	Training, Validation, and Testing Sets	$p < 0.05$
Starke S et al. (2020) [28]	Predicting loco-regional tumor control (LRC) in HNSCC		C-index: 0.31 (95% CI: 0.22-0.39)	Prognosis	Ensemble of 3D-CNNs	10-fold cross-validation repeated three times, Independent Validation Cohort	p=0.001 (log-rank test for Kaplan–Meier curves)
Camalan S et al. (2021) [29]	Primary: Classification accuracy; Secondary: Heat map generation for interpretability	73.6% (±19%) and 90.9% (±12%) accuracy	Accuracy: 73.6%, 90.9%; F1-score: 97.9%, 87.2%; Precision: 95.4%, 99.3%; Recall: 100.0%, 81.1%	Classification	Inception-ResNet-v2, ResNet-101, VGG-16	10-fold Cross-validation, Leave- one-patient-out Cross-validation	Not explicitly stated
Yanagawa M et al. (2021) [30]	Primary: Diagnostic accuracy for IVA; Secondary: Sensitivity and specificity	Improved diagnostic accuracy, especially for less- experienced radiologists	Accuracy: R1 with 3D-CNN: 72.2%, R2: 74.4%, R3: 74.4%; Sensitivity: R1 with 3D-CNN: 77.1%, R2: 85.4%, R3: 93.8%; Specificity: R1 with 3D-CNN: 66.7%, R2: 61.9%, R3: 52.4%; AUC: R1 with 3D-CNN: 0.72, R2: 0.74, R3: 0.73	Classification	3D-CNN with seven convolution-pooling layers, two max-pooling layers, batch normalizatio n, residual connection, and global average pooling	Training and test set split (123 training, 82 test) Nested 10-fold cross-validation	Accuracy improvemen t for R1 and R2 with 3D- CNN support was statistically significant ($p < 0.01$); no significant difference in AUC for each radiologist with and without 3D-CNN ($p > 0.88$)
Zhang Y et al. (2021) [31]	Preoperative MVI prediction in HCC	Fusion model highest performance	AUC 0.81 (training), 0.72 (validation); Sensitivity: 69% (training), 55% (validation); Specificity: 79% (training), 81% (validation)	Detection, Classification	3D CNN with multiparamet ric MRI (T2WI, T2- SPIR, PVP)	Cross-validation with training and validation sets	Not reported
Tan X et al. (2021) [32]	Cervical cancer screening, lesion severity classification (ASCUS, LSIL, HSIL)	Sensitivity: 99.4%, Specificity: 34.8%	Sensitivity: ASCUS - 89.3%, LSIL - 71.5%, HSIL - 73.9%; AUC-ROC: 0.67	Classification & Detection	Faster R- CNN	External Test Set (290 slides, 408,030 images)	Not reported
Huang X et al. (2022) [17]	Primary: EGFR mutation prediction Secondary: Comparison of deep learning and radiomic models and computer vision models	High predictive accuracy	AUC(ModelCNN+radiomic+clinical): -AUC(ModelCNN):0.738 -(Modelradiomic+clinical)	Classification	3D ResNet-101 Deep learning and radiomics-based predictive models	Random train-test split (770 training, 304 test)	Delong test (p=0.0067 for Model CNN vs. Model radiomic + clinical)

Author/Year	Primary Outcome and Secondary outcome	Efficiency/ Performance	Performance Metric (Accuracy, Sensitivity, Specificity, Dice Score, AUC-ROC)	3D CNN Application (Detection, Classification, Segmentation, Prognosis)	Model Used (e.g., CNN Type, Algorithm)	Validation Method (e.g., Cross-validation, External Test Set)	Statistical Significance (e.g., p- value, CI)
Ding Y et al. (2022) [19]	Segmentation of cervical cancer CTV and OARs	V-Net performed better than U-Net	**CTV:** Dice Score: 0.85 (V-Net) vs. 0.83 (U-Net); **OARs:** Sigmoid: 0.80 (V-Net) vs. 0.72 (U-Net), Small bowel: 0.79 (V-Net) vs. 0.74 (U-Net), Bladder: 0.94 (V-Net) vs. 0.93 (U-Net), etc.	Segmentation	V-Net, U-Net (comparison)	10% of the dataset used as validation set	Colon p = 0.046, Small bowel p = 0.048
Jung Y et al. (2022) [34]	Classification of normal vs. ovarian tumors, classification of benign vs. malignant tumors	High diagnostic accuracy	- Normal vs. others: Accuracy 97.22%, Sensitivity 97.22%, AUC-ROC 0.9936 (DenseNet121) - Malignancy detection: Accuracy 90.12%, Sensitivity 86.67%, AUC-ROC 0.9406 (DenseNet161) - Cystadenoma: Sensitivity 82.18%, AUC-ROC 0.9394 - Mature Cystic Teratoma: Sensitivity 80.82%, AUC-ROC 0.9414 Endometrioma: Sensitivity 73.33%, AUC-ROC 0.9248	Classification of ovarian tumors	DenseNet121, DenseNet161, DenseNet201	DenseNet121, DenseNet161, DenseNet201	95% CI reported for all performance metrics
Cui H et al. (2022) [3]	Prediction of Early Recurrence (≤2 years); Comparison of SS vs. NSS	SS-3D-CNN outperforms NSS-3D- CNN	SS-3D-CNN: AUC = 0.789 (95% CI: 0.637-0.941), Sensitivity = 61.9%, Specificity = 90.5%, Accuracy = 0.762; NSS-3D-CNN: AUC = 0.560 (95% CI: 0.378-0.742), Sensitivity = 52.4%, - Specificity = 61.9%, Accuracy = 0.571	Prognosis	3D-CNN using Keras, Python	5-fold Cross-validation, External Test Set	SS-3D- CNN: P < .001 (significant), NSS-3D- CNN: P = .2204 (not significant)
Alotaibi et al. (2023) [10]	Primary: Classification accuracy; Secondary: Recall, Precision, Specificity, F1-Score, AUC	High Performance	**BUSI Dataset:** Accuracy: 87.8%, Sensitivity: 83.8%, Specificity: 89.8%, AUC-ROC: 0.9463 **KAIMRC Dataset:** Accuracy: 85.2%, Sensitivity: 76.4%, Specificity: 89.4%, AUC-ROC: 0.900	Classification	VGG19 (with preprocessing: denoising, ROI highlighting, RGB fusion)	Five-fold cross-validation	Not explicitly reported

Author/Year	Primary Outcome and Secondary outcome	Efficiency/ Performance	Performance Metric (Accuracy, Sensitivity, Specificity, Dice Score, AUC-ROC)	3D CNN Application (Detection, Classification, Segmentation, Prognosis)	Model Used (e.g., CNN Type, Algorithm)	Validation Method (e.g., Cross-validation, External Test Set)	Statistical Significance (e.g., p- value, CI)
Zhou J et al. (2023) [11]	Primary: Breast Cancer Diagnosis Secondary: Lesion Localization	High 83.7% accuracy	Accuracy: 83.7%, Sensitivity: 90.8%, Specificity: 69.3%, AUC-ROC: 0.859, Dice Score: 0.501 ± 0.274(model ensemble), 90.8% sensitivity, 69.3% specificity	Classification and Localization Model Ensemble	3D DenseNet with GAP, GMP, and	Training (1073), Validation (157), Testing (307) subsets	P = 0.039 (compared to GAP model), P = 0.146 (compared to GMP model), 95% CI for accuracy: [79.1%, 87.4%]
Muñoz-Aseguinolaz a U et al. (2023) [4]	Improved nodule detection, reducing false negatives	High 76.44% (3D classification n), 66.73% (2D classification	Accuracy: 76.44% (Case 3.1), 62.79% (Case 3.2); Loss: Decreasing trend	Detection, Classification	3D-CNN with Conv3D layers, ReLU activation, Max Pooling, Global Average Pooling	Training-validation split	Not reported
Chithra PL et al. (2024) [35]	Accurate lung tumor segmentation Detection of tiny tumors (<14mm)	High performance in segmentatio n 98% accuracy	Dice Score: 0.921, Sensitivity: 0.869, F1 Score: 0.815, Mean IoU: 0.842	Segmentation	3D Multi- Layer Convolutiona l Neural Network (3D- MLCNN)	Student's t-test with a 0.05 significance level 80% training, 20% validation	p < 0.05 (vs. MSDS- UNet, 3D U-Net, etc.)
Abinaya K et al. (2024) [21]	Primary: Cervical cancer classification Secondary: Performance comparison with other models	98.6% accuracy	Accuracy: 98.6% Sensitivity: 98.1% Specificity: 98.2% Precision: 97.5% F1 Score: 98.4% AUC-ROC: 0.987	Classification	3D CNN + Vision Transformer + KELM (Kernel Extreme Learning Machine)	Five-fold Cross-validation	

Table 3: Quality assessment done for the retrospective studies using the Newcastle Ottawa Scale (NOS)

No.	Study	Selection S1	S2	S3	S4	Comparability	Outcome O1	O2	O3	Total Score (Max 9 Stars)
1	Alakwaa W et al. (2017) [25]	*	*	*	-	*	*	*	*	7 High- quality studies
2	Trivizakis E et al. (2019) [2]	*	*	*	-	**	*	*	*	8 High- quality studies
3	Perez G et al. (2020) [18]	*	*	*	*	**	*	*	*	9 High- quality studies
4	Mzoughi et al. (2020) [13]	*	*	*	*	*	*	*	*	8 High- quality studies
5	Starke S et al. (2020) [28]	*	*	*	*	**	-	*	*	8 High- quality studies
6	Camalan S et al. (2021) [29]	*	*	*	-	**	*	*	*	8 High- quality studies
7	Yanagawa M et al. (2021) [30]	*	*	*	-	*	*	*	*	7 High- quality studies
8	Zhang Y et al. (2021) [31]	*	*	*	*	**	*	*	*	9 High- quality studies
9	Tan X et al. (2021) [32]	*	*	*	-	*	*	*	*	7 High- quality studies
10	Huang X et al. (2022) [17]	*	*	*	*	*	*	*	*	8 High- quality studies
11	Tao J et al. (2022) [33]	*	*	*	*	*	*	*	*	8 High- quality studies
12	Ding Y et al. (2022) [19]	*	*	*	*	**	*	*	*	9 High- quality studies
13	Jung Y et al. (2022) [34]	*	*	*	*	**	*	*	*	9 High- quality studies
14	Cui H et al. (2022) [3]	*	*	*	*	**	*	*	*	9 High- quality studies
15	Alotaibi et al. (2023) [10]	*	*	*	*	*	*	*	*	8 High-quality studies
16	Zhou J et al. (2023) [11]	*	*	*	*	**	-	*	*	8 High- quality studies
17	Chithra PL et al. (2024) [35]	*	*	*	*	**	*	*	*	9 High- quality studies
18.	Abinaya K et al. (2024) [21]	*	*	*	*	**	*	*	*	9 High- quality studies

Table 4: Quality assessment done for the experimental studies using the SYRCLE's Risk of Bias Tool

No	Author (Year)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total Score (Out of 10)	Risk Level
1	Xu S et al. (2019) [16]	1	0.5	1	1	0.5	1	0.5	0.5	1	1	7.5/10	Moderate Risk
2	Zhang Z et al. (2019) [26]	1	0.5	1	1	0	1	1	0.5	1	1	8/10	Low Risk
3	Wei R et al. (2019) [27]	1	1	1	1	1	1	1	1	1	1	10/10	Low Risk
4	Muñoz-Aseguinolaza U et al. (2023) [4]	1	0	1	1	1	1	0	1	1	1	8/10	Low Risk

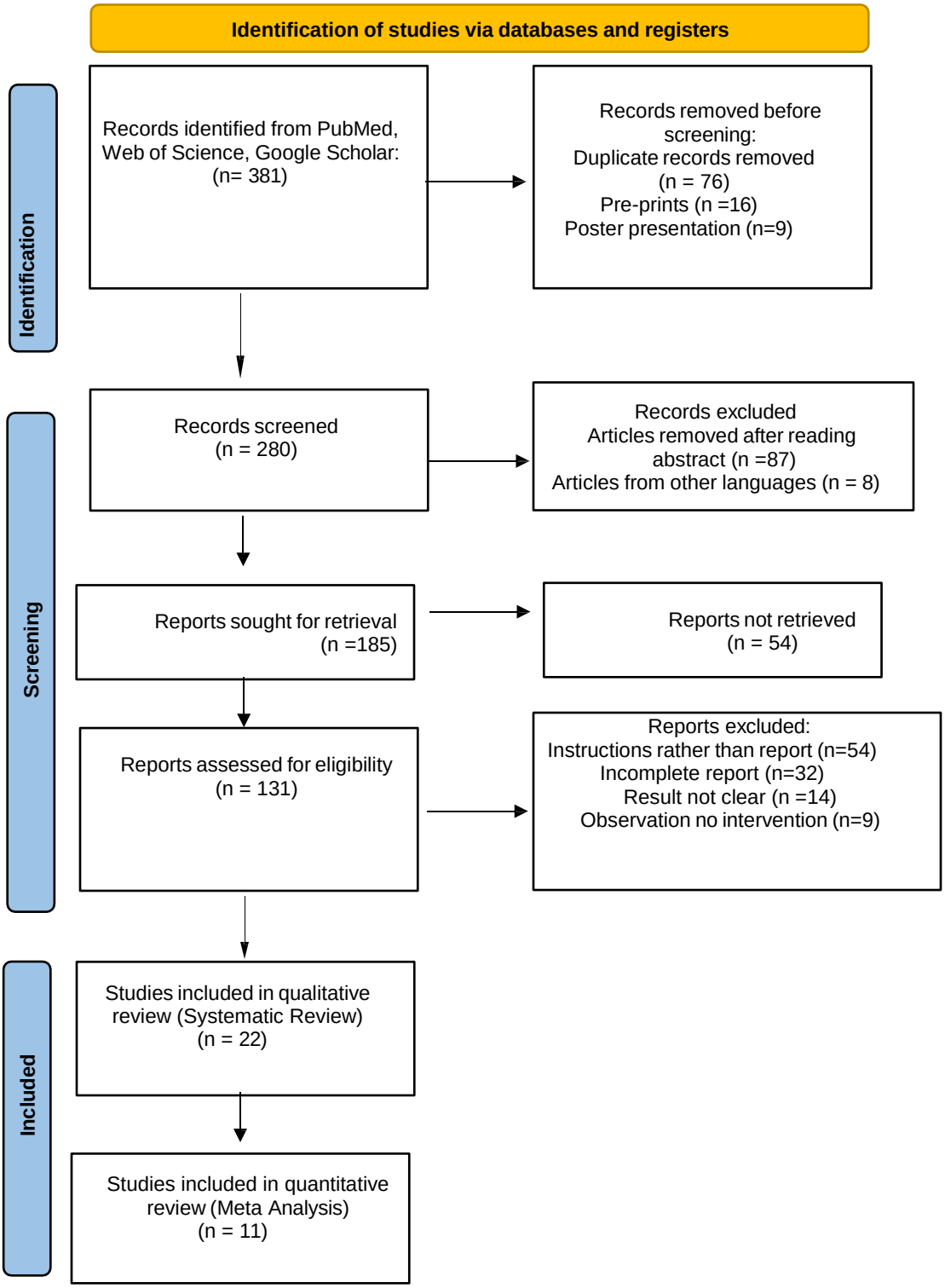

Figure 1: PRISMA 2020 Flowchart for the Review

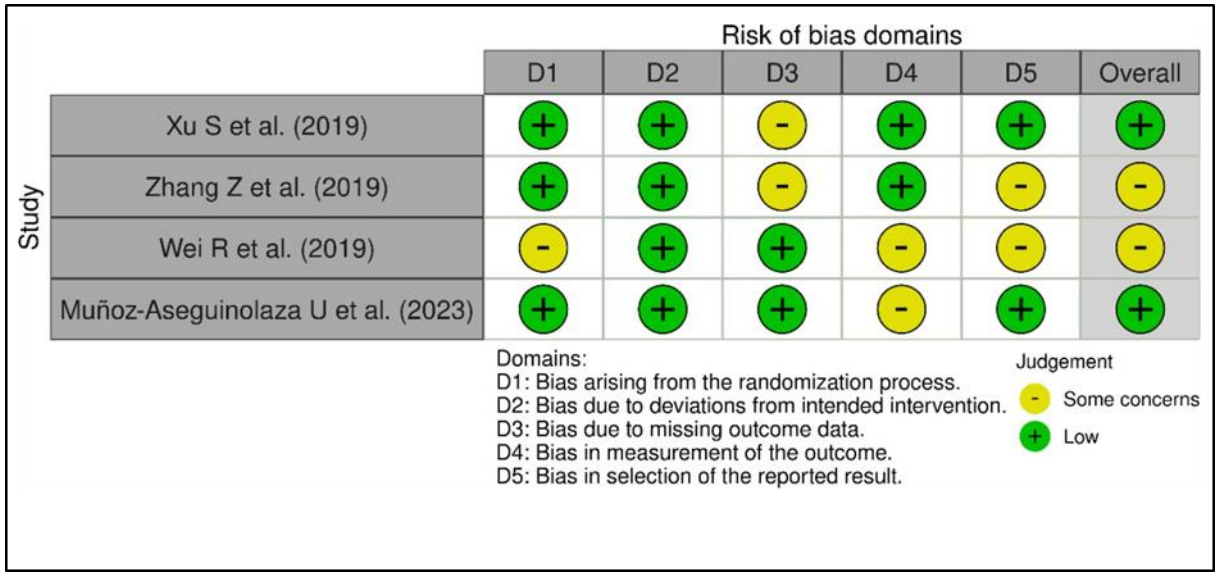

Figure 2. Individual studies - Risk of bias assessment for Experimental studies done using the ROB2 Tool

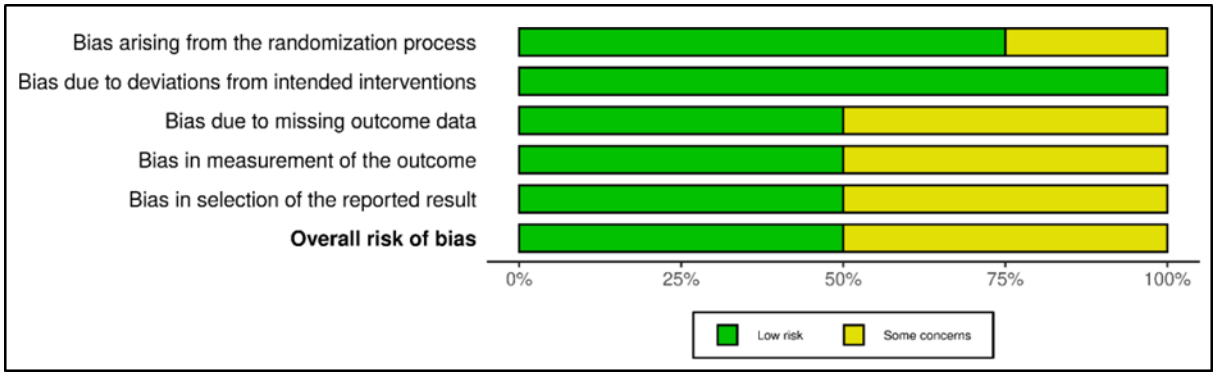

Figure 3. Overall Risk of bias assessment for Experimental studies done using the ROB2T

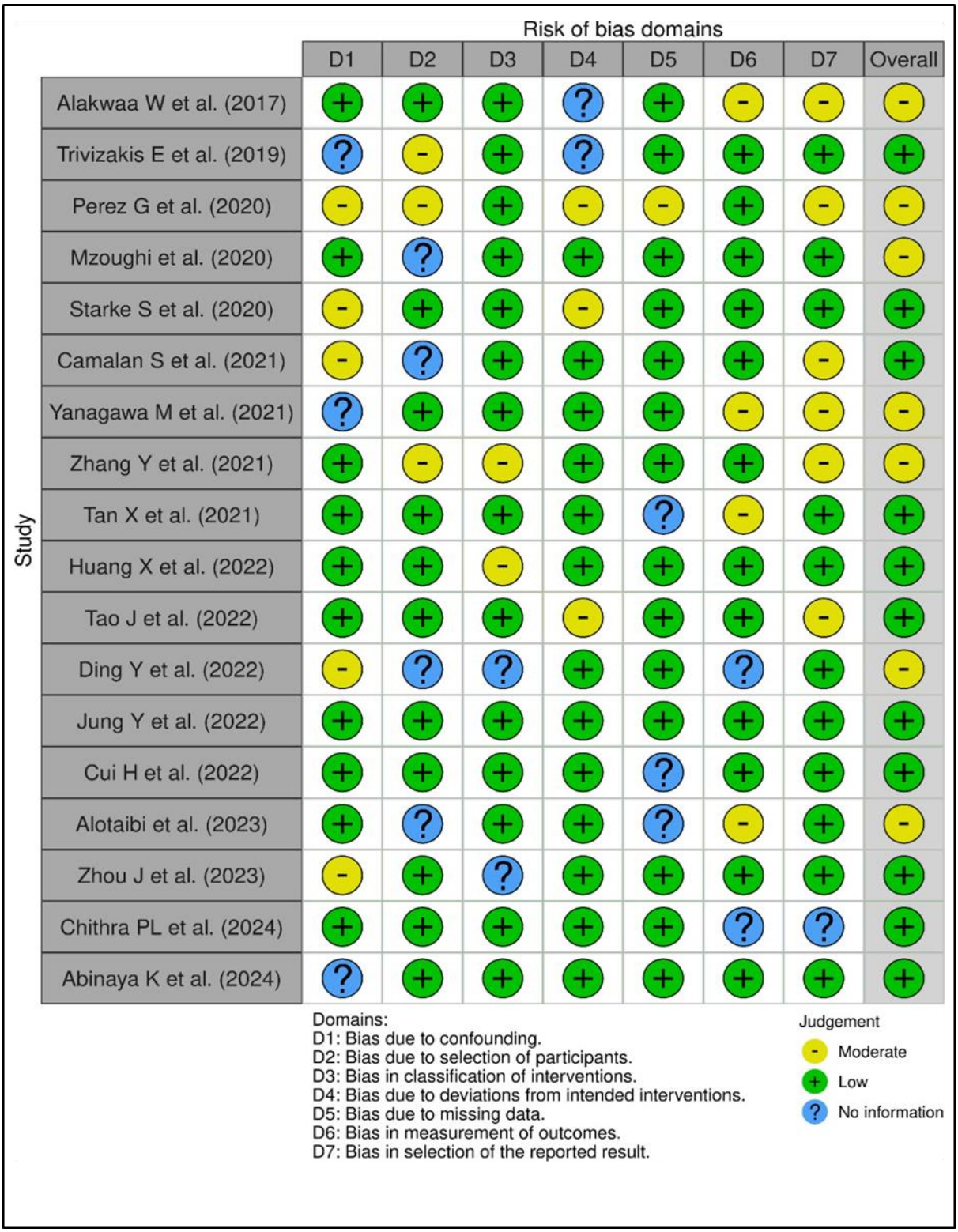

Figure 4. Individual studies - Risk of bias assessment for Retrospective studies done using the ROBINS - I Tool

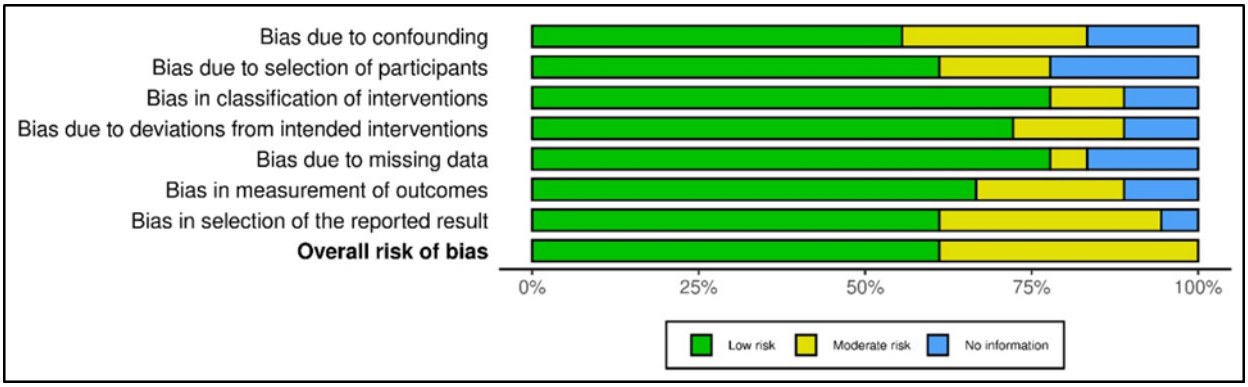

Figure 5. Overall Risk of bias assessment for observational studies done using the ROBINS - I Tool

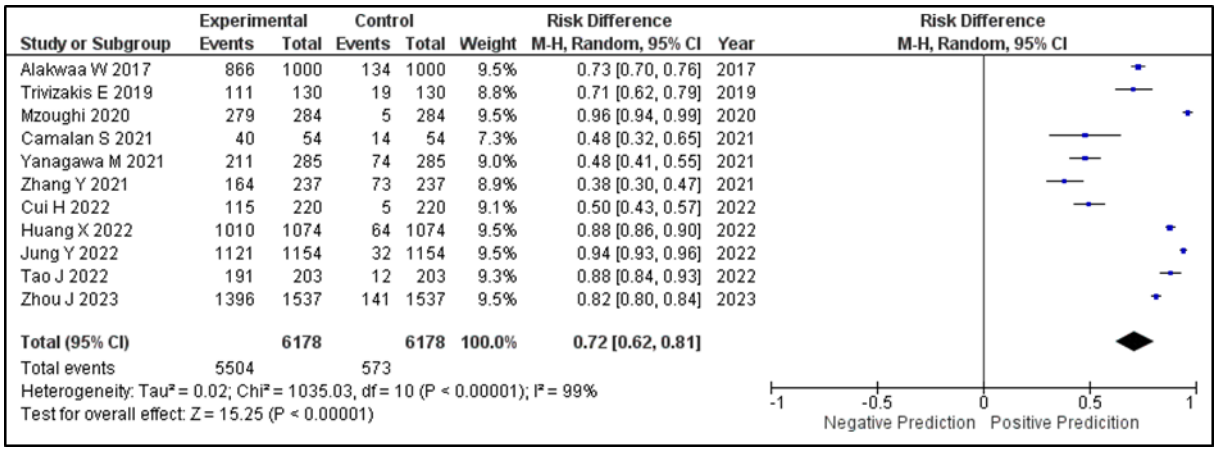

Figure 6. Forest plot overall 3D CNN in image processing and recognition in oncology – sensitivity

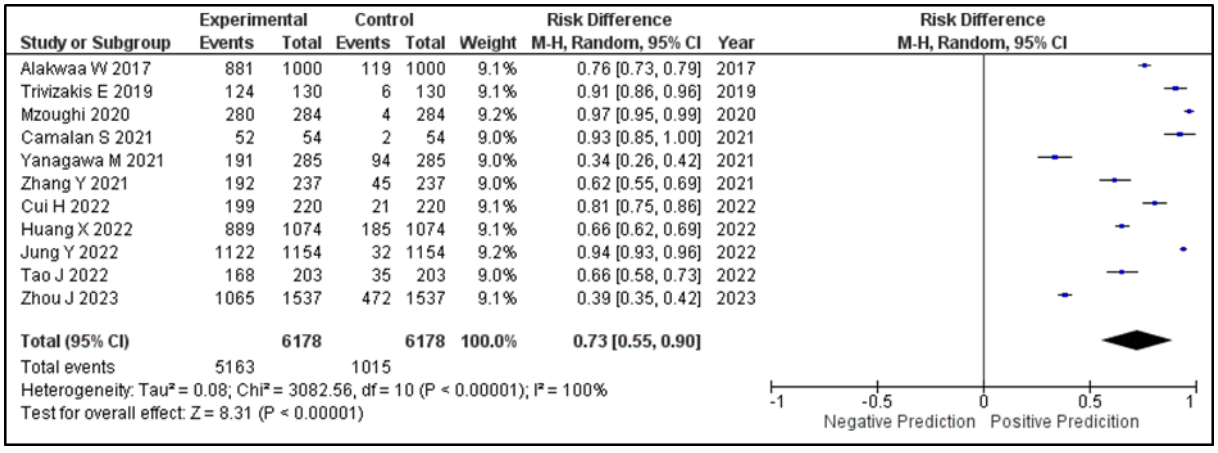

Figure 7. Forest plot overall 3D CNN in image processing and recognition in oncology – specificity

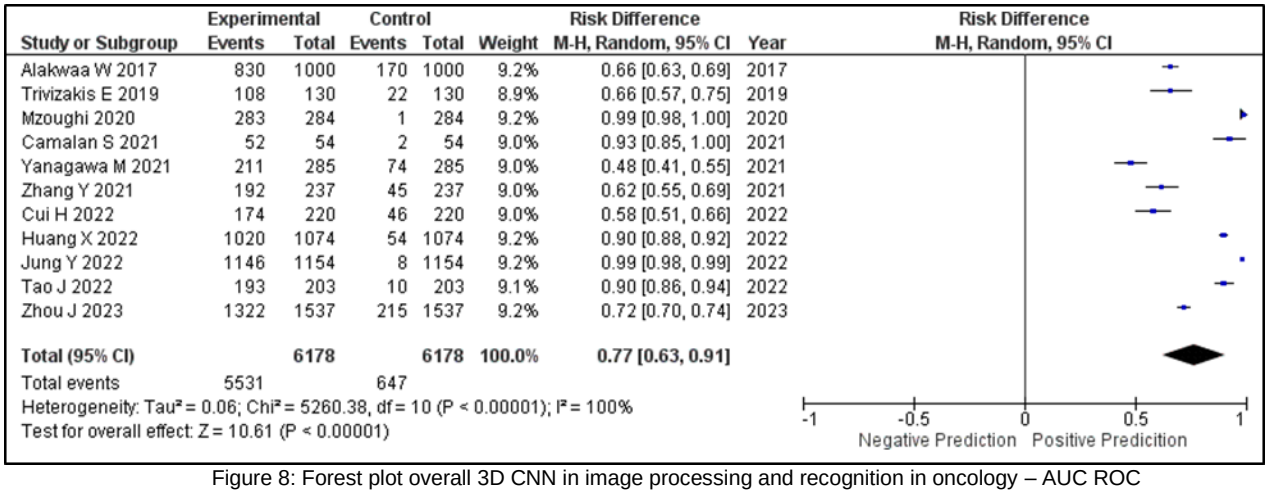

Figure 8: Forest plot overall 3D CNN in image processing and recognition in oncology – AUC ROC

Figure 5 highlights key concerns in observational studies, including confounding and missing data, emphasizing the need for better participant selection, data handling, and reporting transparency to enhance study validity. Using the Newcastle-Ottawa Scale (NOS), Table 3 rates the quality of retrospective investigations; all studies receive excellent quality ratings (scores 7–9). Notably, the studies with the highest scores (9/9) Some studies showed strong selection, comparability, and output evaluation. Using SYRCLE's Risk of Bias Tool, experimental investigations are evaluated in Table 4, with Wei R et al. (2019)[27] exhibiting the best quality (10/10, low risk). Xu S et al. (2019)[16] had a moderate risk (7.5/10), however most experimental studies had minimal risk. These assessments strengthen the reliability of the results and support the solid technique of the included study.

Figure 6 illustrates the pooled sensitivity of 3D convolutional neural network (CNN) models in oncology image processing and recognition. The pooled sensitivity estimate is 0.72 (95% CI: 0.62–0.81), indicating moderate-to-high accuracy in correctly identifying true positive cases.

However, substantial variability exists across studies, as reflected in the high heterogeneity (I^2 = 99%).

Among the included studies, Mzoughi et al. (2020)[13] reported the highest sensitivity at 0.96 (95% CI: 0.94–0.99), demonstrating excellent performance in detecting positive cases. In contrast, Zhang Y et al. (2021)[31] documented the lowest sensitivity at 0.38 (95% CI: 0.30–0.47), indicating challenges in identifying true positives within that dataset. Differences in datasets, imaging techniques, and model training methodologies likely contribute to the observed variations. Despite this heterogeneity, the overall effect remains statistically significant (Z = 15.25, P < 0.00001), supporting the effectiveness of 3D CNN models in oncology image analysis.

Figure 7 presents the pooled specificity of 3D CNN models in distinguishing true negative cases. The pooled specificity is 0.73 (95% CI: 0.55–0.90), suggesting reasonable accuracy in ruling out non-cancerous cases.

However, specificity varies significantly across studies, as indicated by the high heterogeneity (I^2 = 100%). Trivizakis E et al. (2019) [2] reported the highest specificity at 0.91 (95% CI: 0.86–0.96), demonstrating strong performance in correctly classifying negative cases. Conversely, Zhou J et al. (2023)[11] reported the lowest specificity at 0.39 (95% CI: 0.35– 0.42), suggesting a weaker ability to differentiate between cancerous and non-cancerous samples. Differences in imaging modalities, dataset characteristics, and classification thresholds likely contribute to this variability. Despite these inconsistencies, the overall effect is statistically significant (Z = 8.31, P < 0.00001), reinforcing the reliability of 3D CNN models for specificity, though further improvements are needed. Figure 8 evaluates the overall diagnostic performance of 3D CNN models in oncology imaging using the area under the receiver operating characteristic curve (AUC-ROC). The pooled AUC of 0.77 (95% CI: 0.63–

0.91) shows that it can discriminate between positive and negative situations with excellent accuracy. Extreme heterogeneity (I2 = 100%), however, indicates considerable variation between trials. Mzoughi et al. (2020) [13] and Jung Y et al. (2022) [34] recorded the greatest AUC values, both at 0.99 (95% CI: 0.98–1.00), indicating nearly flawless classification. The lowest AUC, 0.62 (95% CI: 0.55–0.69), was found by Zhang Y et al. (2021) [31], in contrast, suggesting worse predictive ability. These variations likely stem from dataset quality, imaging protocols, and neural network architectures. However, the overall effect is statistically significant ($Z = 10.61$, $P < 0.00001$), confirming the potential of 3D CNN models in medical image analysis. Therefore, 3D CNN models show great promise for image processing and recognition in cancer. AUC, sensitivity, and specificity results all support their diagnostic skills, although substantial heterogeneity emphasizes the necessity of standardization.

DISCUSSION

Impact of Imaging Protocols, Scanner Types, and Annotation Standards on Generalizability

A critical factor influencing the performance and generalizability of 3D CNN models is the variability introduced by differing imaging protocols, scanner hardware, and annotation standards across institutions. Inconsistent image acquisition parameters—such as slice thickness, contrast settings, and resolution—can significantly alter the visual characteristics of tumors, affecting model training and inference. For example, models trained on high-resolution MRI datasets may underperform when applied to lower-resolution or differently parameterized images from another center. Scanner variability is another concern. Variations in manufacturers (e.g., GE, Siemens, Philips) and hardware configurations can lead to domain shifts, which reduce the transferability of models across clinical settings. Additionally, differences in image preprocessing pipelines and noise reduction techniques may introduce artifacts or alter feature representation.

Annotation standards also play a crucial role. Many studies rely on manual segmentations or labels from expert radiologists, yet inter-observer variability remains high. Some datasets lack clearly defined annotation protocols, leading to inconsistencies that compromise model reliability. Given these limitations, the development and public sharing of multi-center, large-scale annotated datasets is essential. Datasets spanning diverse scanners, imaging settings, and clinical populations provide a more representative training base, improving model robustness and reducing bias. Harmonizing annotation standards and promoting inter-rater consensus for ground truth labeling can also enhance reproducibility.

Future research should prioritize collaborative data sharing initiatives and domain adaptation techniques to bridge inter-institutional differences. Incorporating federated learning and synthetic data augmentation may further bolster model generalizability across heterogeneous clinical environments.

3D convolutional neural networks (3D CNNs) have revolutionized oncology imaging, namely in cancer detection, classification, segmentation, and prognosis, according to the systematic review and meta-analysis. 3D CNNs have better diagnostic accuracy, sensitivity, and specificity than conventional imaging methods and classic two-dimensional (2D) CNNs [16]. Their capacity to handle volumetric data improves the extraction of spatial features, which in turn improves the diagnostic reliability of cervical cancer segmentation[32], breast cancer classification[31], and lung cancer screening[25]. Deep learning applications in tumor identification and analysis in the present review supports the body of data publishes in previous literature that demonstrates the benefits of 3D CNNs in cancer imaging [36,37,38].

In treatment planning, tumor segmentation is essential, especially for surgical and radiation procedures. This review discovered that 3D CNN-based models greatly improve segmentation accuracy when compared to traditional techniques. Ding et al., for example, used a 3D deep neural network for automated cervical cancer delineation in CT image planning, to achieve accurate tumor boundary recognition[19]. Similarly, Meng et al. used dual-scale 3D CNNs to segment liver tumors, showing that they were better than conventional

methods [12]. The integration of long-range 2D context within 3D CNN architectures, as shown by Mlynarski et al., further refines spatial representations, improving segmentation outcomes[14]. These advancements emphasize the potential of 3D CNNs in optimizing clinical workflows and minimizing interobserver variability [1].

One important conclusion of this study is that 3D CNNs are better than 2D CNNs in capturing spatial contextual information. In oncological imaging, volumetric evaluation is necessary for tumor characterisation instead of slice-by-slice examination. Research has indicated that 3D CNNs boost diagnostic performance by improving tumor identification and segmentation in imaging modalities such MRI, CT, and PET scans[17,39]. The results of this research support these findings, showing that 3D CNNs have greater tumor classification sensitivity and specificity than traditional techniques[33,35]. Tumor form, size, and texture changes make automated tumor segmentation a difficult process. High accuracy segmentation of breast cancer lesions, lung nodules, and brain tumors has been demonstrated with the use of 3D CNN architectures like U-Net and V-Net[15,40]. These findings are supported by the results of the present review, which showed that the 3D CNN models that were used performed well in segmentation, as evidenced by their high Dice similarity coefficients[2]. As previously shown in the literature, the inclusion of spatial dependencies in three dimensions improves model resilience and decreases segmentation mistakes [30].

3D CNNs perform better than conventional radiomics-based methods for tumor classification because they can directly extract hierarchical characteristics from imaging data. The results of the study indicate that when it comes to differentiating between benign and malignant lesions, deep learning-based feature extraction techniques outperform handmade features [16,41]. These findings are in line with earlier research that shown how effectively 3D CNNs trained on sizable datasets extend to situations that have not been observed, increasing classification accuracy [42]. The capacity of these models to learn spatial hierarchies of tumor morphology a crucial component of accurate diagnosis and prognosis estimation is what gives them their resilience[37]. Multi-modal fusion for cancer imaging is another important benefit of 3D CNNs that this study found. More accurate tumor characterization results from the integration of multi-parametric imaging modalities like PET-CT and MRI-PET, which improve the complementary information gleaned from various sources [18,28]. This strategy is supported by the review's findings, which showed that 3D CNNs performed better in diagnostic tasks when trained on multi-modal datasets as opposed to single-modality inputs[13]. Combining structural and functional imaging data has been shown in other studies to improve tumor grading and treatment response evaluation [27].

Even though 3D CNNs have several benefits for cancer imaging, several obstacles prevent their broad clinical use. Since deep 3D CNN models take a significant amount of processing power and memory to train, one of their main drawbacks is their high computational demands[26]. Since this limitation is well known, researchers are looking at optimization strategies including knowledge distillation, model pruning, and hybrid architectures to increase computing efficiency[25]. The lack of annotated medical imaging datasets also poses a problem because deep learning models need a lot of labeled data to achieve good generalization. This has been addressed by methods like data augmentation and transfer learning, which allow models to use prior information and adjust to cancer imaging tasks[9,29]. The results of this work indicate that, in line with other studies, pre-training on sizable datasets and fine-tuning on domain-specific cancer pictures improve model performance[43]. Deep learning models' interpretability is still a crucial problem in clinical practice, even in the face of computational limitations. Adoption may be hampered by 3D CNNs' "black-box" aspect, which reduces clinician confidence even if they achieve excellent diagnostic accuracy [44].

Class activation maps (CAMs) and gradient-based visualization approaches are examples of explainability strategies that have been investigated to increase the transparency of deep learning judgments [10]. Using these interpretability techniques can boost clinician confidence and make it easier to

integrate 3D CNNs into standard oncology procedures, according to the review's results[45]. These techniques are also being used to the detection of ovarian cancer, where regular MRI imaging shows notable improvements in lesion localization and benign-malignant tumor distinction using deep learning models [46]. The adaptability of 3D CNNs in medical imaging is further shown by recent developments that show the potential of AI-driven systems for identifying localized liver lesions in a variety of clinical contexts [47].

The fields of radiomics and radiogenomics, which utilize deep learning models to connect imaging properties with molecular and genetic tumor profiles, are another developing field where 3D CNNs show potential. 3D CNN-based radiomic signatures have been shown in recent studies to be more accurate than conventional techniques at predicting tumor aggressiveness and responsiveness to therapy[11,48]. The precision of AI-driven oncological imaging has been improved by recent developments in deep learning applications, which have further shown the technology's potential in ovarian cancer diagnosis and lesion localization, as well as in the identification and categorization of localized liver lesions [46,47,49].

The results of the present investigation support previous findings, emphasizing how deep learning and radiogenomic analysis may be used to improve individualized cancer treatment planning[50]. These developments may make it possible to better stratify patients, which might result in more individualized treatment plans based on imaging biomarkers. High I2 values for sensitivity, specificity, and AUC-ROC (I2 = 100%) in the meta-analysis demonstrated significant heterogeneity among investigations, despite encouraging results. This variation most likely results from variations in model topologies, imaging procedures, and dataset quality[36]. For example, Zhang et al. reported a significantly lower AUC of 0.62 for hepatocellular carcinoma prediction, highlighting the inconsistent efficacy of the model across different cancers[31], while Mzoughi et al. reported an AUC of 0.99 for glioma classification, demonstrating exceptional model performance[13]. These disparities show that to increase model universality and reproducibility, consistent procedures, multi-center datasets, and harmonized imaging methods are required. Most of the studies' risk of bias evaluations utilizing the ROB2 and ROBINS-I methods showed moderate to low risks. There were several issues with confounding factors, selective reporting, and missing data (Table 3, Table 4). For instance, Trivizakis et al.[2] experienced problems with participant selection, which impacted the overall dependability of findings, whereas Xu et al. [16] displayed mild bias due to missing outcome data (Figure 4).

These findings highlight how crucial thorough study design and open reporting procedures are to guaranteeing the validity of research. Strong diagnostic performance was continuously shown by high-caliber research, including those by Perez et al.[18] and Zhang et al. [26], highlighting the necessity of following methodological guidelines to enable the clinical translation of 3D CNN models (Table 3). To address the existing challenges, future research should focus on developing advanced hardware infrastructure to support computationally intensive models, improving model interpretability through explainable AI techniques, and establishing standardized imaging protocols for oncology applications. Overcoming these barriers will be crucial for enhancing the clinical utility of 3D CNNs, ultimately improving cancer diagnosis, treatment planning, and patient outcomes.

STRENGTHS AND LIMITATIONS

This study comprehensively evaluates the role of 3D CNNs in oncology imaging, demonstrating their superiority in tumor detection, classification, and segmentation. A major strength is the inclusion of diverse cancer types and a systematic meta- analysis, enhancing the reliability of findings. The study also considers key factors such as model generalizability, interpretability, and clinical applicability. However, significant heterogeneity among included studies, reflected by high I^2 values, may affect result consistency. Additionally, real-world challenges like computational costs and clinical integration are not extensively addressed. Potential publication bias is another limitation. Standardized methodologies, larger multi-center datasets, and improved model interpretability are essential for ensuring the clinical

translation of 3D CNNs and optimizing their role in precision oncology.

Future Directions

The next frontier in the application of 3D CNNs in oncology imaging involves improving transparency through explainable AI (XAI) techniques and establishing clinical workflows that incorporate AI outputs effectively. Traditional deep learning models often function as "black boxes," hindering clinical trust and interpretability. Future research should explore and integrate explainability techniques such as class activation maps (CAM), saliency maps, Grad-CAM, LIME, and SHAP values to visualize which parts of the image contributed most to the model's decision. These methods enable radiologists to better understand, validate, and interact with AI-generated predictions. To facilitate adoption, AI systems should be embedded within existing clinical workflows through intuitive interfaces and decision-support tools. For example, integration within PACS or radiology information systems (RIS) can enable real-time feedback on lesion detection or classification. User interfaces should support interactive visualization highlight areas of uncertainty, and allow manual annotation to refine model outputs. Additionally, AI models should support multidisciplinary decision-making by generating structured reports, facilitating collaborative review, and aligning outputs with standard oncologic pathways. Future clinical workflows may benefit from hybrid decision-making systems where AI assists in triage, pre-screening, and prioritization, while radiologists retain final interpretive authority. By enabling radiologists to interact with AI predictions rather than passively receive results, the partnership between human expertise and machine intelligence can be optimized for precision diagnostics.

VALIDITY OF CONCLUSION

According to this study, 3D CNNs have revolutionized cancer imaging by demonstrating their superiority in tumor identification, segmentation, classification, and prognosis. Utilizing volumetric medical data, these models improve patient outcomes by increasing the accuracy of diagnosis and facilitating more precise treatment planning. In clinical contexts, 3D CNNs can improve decision-making, expedite processes, and shorten diagnostic turnaround times, especially when resources are limited. Broader use, however, requires addressing issues such high computing costs, data restrictions, and model interpretability. The main aim of future studies should be to improve explainability, integrate multi-modal imaging, and maximize efficiency. For these models to be improved and more automated, dependable, and individualized cancer diagnosis and therapy to be possible, cooperation between AI researchers and physicians is essential.

REFERENCES

1. Singh SP, Wang L, Gupta S, et al. 3D Deep Learning on Medical Images: A Review. Sensors (Basel). 2020;20(18):5097 .

2. Trivizakis E, Manikis GC, Nikiforaki K, et al. Extending 2-D Convolutional Neural Networks to 3-D for Advancing Deep Learning Cancer Classification With Application to MRI Liver Tumor Differentiation. IEEE J Biomed Health Inform. 2019;23(3):923-930 .

3. Cui H, Wang KY, Li WY, et al. CT images-based 3D convolutional neural network to predict early recurrence of solitary hepatocellular carcinoma after radical hepatectomy. Diagn Interv Radiol. 2022;28(6):524-531 .

4. Muñoz-Aseguinolaza U, Fernandez-Iriondo I, Rodríguez-Moreno I, et al. Convolutional neural network-based classification and monitoring models for lung cancer detection: 3D perspective approach. Heliyon. 2023;9(11):e21203 .

5. Mishra S, Chaudhary NK, Asthana P, et al. Deep 3d convolutional neural network for automated lung cancer diagnosis. In Computing and Network Sustainability: Proceedings of IRSCNS 2018. Springer Singapore, India, 2019, pp 157-165.

6. Ahmed T, Parvin MS, Haque MR, et al. Lung cancer detection using CT image based on 3D convolutional neural network. J Comput Commun. 2020;8(03):35-42.

7. Jafari Z, Karami E. Breast cancer detection in mammography images: a CNN-based approach with feature selection. Inf. 2023;14(7):410.

8. Neal Joshua ES, Bhattacharyya D, Chakkravarthy M, et al. 3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-Weighted Class Activation. J Healthc Eng. 2021;2021:6695518 .

9. Yu J, Yang B, Wang J, et al. 2D CNN versus 3D CNN for false- positive reduction in lung cancer screening. J Med Imaging (Bellingham). 2020;7(5):051202 .

10. Alotaibi M, Aljouie A, Alluhaidan N, et al. Breast cancer classification based on convolutional neural network and image fusion approaches using ultrasound images. Heliyon. 2023;9(11):e22406 .
11. Zhou J, Luo LY, Dou Q, et al. Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. J Magn Reson Imaging. 2019 Oct;50(4):1144-1151 .
12. Meng L, Tian Y, Bu S. Liver tumor segmentation based on 3D convolutional neural network with dual scale. J Appl Clin Med Phys. 2020;21(1):144-157 .
13. Mzoughi H, Njeh I, Wali A, et al. Deep Multi-Scale 3D Convolutional Neural Network (CNN) for MRI Gliomas Brain Tumor Classification. J Digit Imaging. 2020;33(4):903-915 .
14. Mlynarski P, Delingette H, Criminisi A, et al. 3D convolutional neural networks for tumor segmentation using long-range 2D context. Comput Med Imaging Graph. 2019;73:60-72 .
15. Ilesanmi AE, Ilesanmi TO, Ajayi BO. Reviewing 3D convolutional neural network approaches for medical image segmentation. Heliyon. 2024;10(6):e27398 .
16. Xu S, Liu C, Zong Y, et al. An early diagnosis of oral cancer based on three-dimensional convolutional neural networks. IEEE Access. 2019;7:158603-11.
17. Huang X, Sun Y, Tan M, et al. Three-dimensional convolutional neural network-based prediction of epidermal growth factor receptor expression status in patients with non-small cell lung cancer. Front Oncol. 2022;12:772770.
18. Perez G, Arbelaez P. Automated lung cancer diagnosis using three-dimensional convolutional neural networks. Med Biol Eng Comput. 2020;58(8):1803-1815 .
19. Ding Y, Chen Z, Wang Z, et al. Three-dimensional deep neural network for automatic delineation of cervical cancer in planning computed tomography images. J Appl Clin Med Phys. 2022;23(4):e13566 .
20. Zhou C, Ding C, Wang X, et al. One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. IEEE Trans Image Process. 2020. Online ahead of print.
21. Abinaya K, Sivakumar B . A Deep Learning-Based Approach for Cervical Cancer Classification Using 3D CNN and Vision Transformer. J Imaging Inform Med. 2024;37(1):280-296 .
22. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
23. https://www.ohri.ca/programs/ clinical_epidemiology/ oxford.asp
24. Hooijmans CR, Rovers MM, De Vries RB, et al. SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol. 2014; 14: 43.
25. Alakwaa W, Nassef M, Badr A. Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). Int. J. Adv. Comput. Sci. Appl.2017;8.(٨)
26. Zhang Z, Chen L, Wang Y, et al. Label-Free Estimation of Therapeutic Efficacy on 3D Cancer Spheres Using Convolutional Neural Network Image Analysis. Anal Chem. 2019;91(21):14093-14100 .
27. Wei R, Zhou F, Liu B, et al. Convolutional neural network (CNN) based three dimensional tumor localization using single X-ray projection. IEEE Access. 2019;7:37026-38.
28. Starke S, Leger S, Zwanenburg A, et al. 2D and 3D convolutional neural networks for outcome modelling of locally advanced head and neck squamous cell carcinoma. Sci Rep. 2020;10:15625 .
29. Camalan S, Mahmood H, Binol H, et al. Convolutional Neural Network-Based Clinical Predictors of Oral Dysplasia: Class Activation Map Analysis of Deep Learning Results. Cancers. 2021;13(6):1291 .
30. Yanagawa M, Niioka H, Kusumoto M, et al. Diagnostic performance for pulmonary adenocarcinoma on CT: comparison of radiologists with and without three-dimensional convolutional neural network. Eur Radiol. 2021;31(4):1978-1986 .
31. Zhang Y, Lv X, Qiu J, et al. Deep Learning with 3D Convolutional Neural Network for Noninvasive Prediction of Microvascular Invasion in Hepatocellular Carcinoma. J Magn Reson Imaging. 2021;54(1):134-143 .
32. Tan X, Li K, Zhang J, et al. Automatic model for cervical cancer screening based on convolutional neural network: a retrospective, multicohort, multicenter study. Cancer Cell Int. 2021;21(1):35 .
33. Tao J, Liang C, Yin K, et al. 3D convolutional neural network model from contrast-enhanced CT to predict spread through air spaces in non-small cell lung cancer. Diagn Interv Imaging. 2022;103(11):535-544 .
34. Jung Y, Kim T, Han MR, et al. Ovarian tumor diagnosis using deep convolutional neural networks and a denoising convolutional autoencoder. Sci Rep. 2022;12(1):17024 .
35. Chithra PL, Bhavani P. A Novel 3D Multi-Layer Convolutional Neural Networks for Lung Cancer Segmentation in CT Images. Indian J Sci Technol. 2024;17(13):1368-80.
36. Tiwari S, Jain G, Shetty DK, et al. A Comprehensive Review on the Application of 3D Convolutional Neural Networks in Medical Imaging. Eng Proc. 2023;59(1):3.

37. Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3d shape recognition. ICCV.2015;945-953.
38. Wang TW, Hong JS, Huang JW, et al. Systematic review and meta- analysis of deep learning applications in computed tomography lung cancer segmentation. Radiother Oncol. 2024:197:110344.
39. Jin T, Cui H, Zeng S, et al. Learning deep spatial lung features by 3D convolutional neural network for early cancer detection. DICTA. 2017:1-6.
40. Ji S, Yang M, Yu K. 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell. 2013;35(1):221-31.
41. Taddese AA, Tilahun BC, Awoke T, et al. Deep-learning models for image-based gynecological cancer diagnosis: a systematic review and meta- analysis. Front Oncol. 2024;13:1216326 .
42. Leung JH, Karmakar R, Mukundan A, et al. Systematic Meta-Analysis of Computer-Aided Detection of Breast Cancer Using Hyperspectral Imaging. Bioengineering (Basel). 2024;11(11):1060 .
43. Alabi RO, Bello IO, Youssef O, et al. Utilizing Deep Machine Learning for Prognostication of Oral Squamous Cell Carcinoma-A Systematic Review. Front Oral Health. 2021;2:686863 .
44. Beristain-Colorado MDP, Castro-Gutiérrez MEM, Torres-Rosas R, et al Application of neural networks for the detection of oral cancer: A systematic review. Dent Med Probl. 2024;61(1):121-128 .
45. Xu Q, Wang X, Jiang H. Convolutional neural network for breast cancer diagnosis using diffuse optical tomography. Vis Comput Ind Biomed Art. 2019;2(1):1 .
46. Wang Y, Zhang H, Wang T, et al. Deep learning for the ovarian lesion localization and discrimination between borderline and malignant ovarian tumors based on routine MR imaging. Sci Rep. 2023;13(1):2770.
47. Ying H, Liu X, Zhang M, et al. A multicenter clinical AI system study for detection and diagnosis of focal liver lesions. Nat Commun. 2024;15(1):1131.
48. Jiang X, Hu Z, Wang S, et al. Deep Learning for Medical Image-Based Cancer Diagnosis. Cancers (Basel). 2023 Jul 13;15(14):3608 .
49. Sadeghi MH, Sina S, Omidi H, et al. Deep learning in ovarian cancer diagnosis: a comprehensive review of various imaging modalities. Pol J Radiol. 2024;89:e30-e48 .
50. Thakur A, Gupta M, Sinha DK, et al. Transformative breast Cancer diagnosis using CNNs with optimized Reduce L R On Plateau and Early stopping Enhancements. Int J Comput Intell Sys. 2024;17(1):14.