Deep learning models for CT image classification: a comprehensive literature review

Isah Salim Ahmad[1,2], Jingjing Dai[1,2], Yaoqin Xie[1,2], Xiaokun Liang[1,2]

[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China; [2]University of Chinese Academy of Sciences, Beijing, China

Contributions: (I) Conception and design: All authors; (II) Administrative support: Y Xie, X Liang; (III) Provision of study materials or patients: Y Xie, X Liang; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiaokun Liang, PhD. Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, No. 1068 Xueyuan Avenue, Shenzhen University Town, Xili, Nanshan District, Shenzhen 518055, China; University of Chinese Academy of Sciences, Beijing, China. Email: xk.liang@siat.ac.cn.

Background and Objective: Computed tomography (CT) imaging plays a crucial role in the early detection and diagnosis of life-threatening diseases, particularly in respiratory illnesses and oncology. The rapid advancement of deep learning (DL) has revolutionized CT image analysis, enhancing diagnostic accuracy and efficiency. This review explores the impact of advanced DL methodologies in CT imaging, with a particular focus on their applications in coronavirus disease 2019 (COVID-19) detection and lung nodule classification.

Methods: A comprehensive literature search was conducted, examining the evolution of DL architectures in medical imaging from conventional convolutional neural networks (CNNs) to sophisticated foundational models (FMs). We reviewed publications from major databases, focusing on developments in CT image analysis using DL from 2013 to 2023. Our search criteria included all types of articles, with a focus on peer-reviewed research papers and review articles in English.

Key Content and Findings: The review reveals that DL, particularly advanced architectures like FMs, has transformed CT image analysis by streamlining interpretation processes and enhancing diagnostic capabilities. We found significant advancements in addressing global health challenges, especially during the COVID-19 pandemic, and in ongoing efforts for lung cancer screening. The review also addresses technical challenges in CT image analysis, including data variability, the need for large high-quality datasets, and computational demands. Innovative strategies such as transfer learning, data augmentation, and distributed computing are explored as solutions to these challenges.

Conclusions: This review underscores the pivotal role of DL in advancing CT image analysis, particularly for COVID-19 and lung nodule detection. The integration of DL models into clinical workflows shows promising potential to enhance diagnostic accuracy and efficiency. However, challenges remain in areas of interpretability, validation, and regulatory compliance. The review advocates for continued research, interdisciplinary collaboration, and ethical considerations as DL technologies become integral to clinical practice. While traditional imaging techniques remain vital, the integration of DL represents a significant advancement in medical diagnostics, with far-reaching implications for future research, clinical practice, and healthcare policy.

Keywords: Computed tomography (CT); deep learning (DL); foundation models; coronavirus disease 2019 (COVID-19); nodule detection

Introduction

Computed tomography (CT) imaging has transformed the landscape of medical diagnostics by offering detailed and high-resolution cross-sectional images of the human body. These images are indispensable for the detection and diagnosis of a myriad of medical conditions, including tumors, fractures, and internal injuries. However, interpreting CT images is inherently challenging and time-consuming, demanding significant expertise and experience from radiologists. With the proliferation of CT scanners and the increasing volume of medical imaging data, there is an urgent need for automated and efficient methods to analyze and interpret these images (1,2). CT imaging has a rich history dating back to the early 1970s. Developed by Sir Godfrey Hounsfield and Allan Cormack, who were awarded the Nobel Prize in Physiology or Medicine in 1979 for their work (3), CT revolutionized medical imaging by providing cross-sectional views of the body. The first clinical CT scan was performed in 1971, and since then, the technology has undergone significant advancements (4), evolving from single-slice to multi-slice scanners capable of producing high-resolution 3D images in seconds.

Lung cancer is a serious condition characterized by the abnormal growth of cells in the lungs. It is not related to coronavirus disease 2019 (COVID-19), which is a respiratory illness caused by a virus. The global healthcare landscape has been significantly impacted by the COVID-19 pandemic and the ongoing challenge of early lung cancer detection. The World Health Organization (WHO) has emphasized the critical role of rapid and accurate diagnosis in managing the spread of COVID-19, stating that "early identification, isolation, and care of COVID-19 cases are essential to limiting the spread of the virus" (5). Similarly, for lung nodules, which can be indicative of early-stage lung cancer, the WHO notes that "early detection of lung cancer is crucial for improving survival rates" (WHO, 2021). CT imaging has emerged as a powerful tool in addressing both these challenges. For COVID-19, CT scans can reveal characteristic patterns of ground-glass opacities and consolidations in the lungs, often before the onset of severe symptoms or in cases where real-time reverse transcription polymerase chain reaction (RT-PCR) tests may yield false negatives. In the context of lung nodule detection, CT imaging allows for the identification of small, potentially cancerous lesions that might be missed on conventional chest X-rays (CXRs). The application of advanced deep learning (DL) techniques to CT image analysis has shown promising results in enhancing the speed and accuracy of both COVID-19 diagnosis and lung nodule detection (6). These AI-driven approaches not only aid in rapid triage and diagnosis but also have the potential to alleviate the burden on healthcare systems by automating initial screening processes. As such, the development and refinement of DL models for CT image classification represent a critical area of research with significant implications for global public health (7).

CT imaging has revolutionized medical diagnostics since its introduction, offering unparalleled insights into the human body's internal structures. This technology works by combining a series of X-ray images taken from different angles around the body and using computer processing to create cross-sectional images (slices) of the bones, blood vessels, and soft tissues. The significance of CT imaging in modern medicine is multifaceted: high spatial resolution, three-dimensional views, quick and non-invasive, versatile applications, cost-effective diagnostics, and emergency and trauma assessment. Radiological CT image acquisition involves taking multiple X-ray images from different angles around the body, which are then processed to create detailed cross-sectional images. The axial chest CT image specifically provides valuable information about the lungs and surrounding structures, aiding in the diagnosis of various conditions. This technique allows for precise visualization of abnormalities, such as tumors or infections, enhancing clinical decision-making. The significance of CT imaging in modern healthcare cannot be overstated. Its ability to provide detailed, non-invasive insights into the human body has transformed diagnostic capabilities, improved treatment outcomes, and advanced medical research. As technology continues to evolve, CT imaging is likely to become even more powerful and indispensable in the medical field (7). Despite the substantial progress in DL for CT image classification, several challenges remain. One major challenge is data scarcity, quality and annotating large datasets, model interpretability, complexity and volume, subtle abnormalities and variability, multimodal integration, radiation exposure and generalization and robustness. Despite the efficiency, comfort, and overall safety of CT scans, patients should be aware that there is a slight possibility of experiencing an allergic reaction to the contrast medium utilized during the examination.

DL, a subset of machine learning (ML) that utilizes artificial neural networks to model complex patterns in data, has demonstrated tremendous potential in medical image analysis (8). These algorithms can automatically learn and extract features from large datasets, making them particularly suitable for tasks such as image classification,

segmentation, and detection. DL models can be trained to recognize specific anatomical structures, detect abnormalities, and classify various pathologies with high accuracy and efficiency (9). One of the most significant advancements in DL for CT image classification is the development of convolutional neural networks (CNNs). CNNs are specially designed to process visual data and consist of multiple layers of convolutional and pooling operations, followed by fully connected layers for classification (6). These networks have been successfully applied to a range of medical imaging tasks, including tumor detection, organ segmentation, and disease classification. By leveraging the hierarchical features learned by CNNs, researchers have achieved state-of-the-art performance in CT image analysis. In addition to CNNs, recurrent neural networks (RNNs) have been explored for CT image classification, especially in tasks involving sequential data or time-series analysis. RNNs are well-suited for tracking changes in image features over time, predicting future states based on past observations, and handling variable-length input sequences. By combining the strengths of CNNs and RNNs, researchers have developed hybrid architectures that effectively analyze dynamic and complex patterns in CT images (10).

Comparison of imaging techniques

Medical imaging techniques are indispensable tools in the diagnosis and treatment of various medical conditions. Each technique offers unique advantages and is suited to different types of clinical applications. A brief comparison of commonly used imaging modalities, including CT, positron emission tomography-CT (PET-CT), X-ray, ultrasound, magnetic resonance imaging (MRI), positron emission tomography (PET), single photon emission CT (SPECT), and optical coherence tomography (OCT) (11) are discussed in *Table 1*.

Understanding these modalities' strengths and weaknesses helps clinicians choose the appropriate imaging technique for accurate diagnosis and effective patient care.

Paper contribution

This review paper makes several significant contributions to the field of CT image classification using DL and foundational model (FM):

❖ Comprehensive synthesis: we provide a thorough and up-to-date synthesis of state-of-the-art

DL techniques and FMs applied to CT image classification. This consolidation of knowledge bridges gaps between computer science, medical imaging, and clinical practice.

❖ Critical analysis of methodologies: our paper offers a critical analysis of various DL architectures, including CNNs, RNNs, GANs, and hybrid models, evaluating their efficacy and limitations in CT image analysis. We also provide an in-depth examination of emerging FMs like BERT, GPT, CLIP, and ViT, elucidating their potential to enhance classification accuracy.

❖ Benchmark dataset evaluation: we present a comprehensive assessment of benchmark datasets used in CT image classification, highlighting their strengths, limitations, and potential biases. This evaluation is crucial for understanding the generalizability and robustness of current models.

❖ Novel taxonomy of challenges: our review introduces a novel taxonomy of challenges in CT image classification, encompassing issues such as data scarcity, model generalization, clinical integration, and the complexity of CT imaging protocols. This structured approach provides a clear roadmap for future research directions.

❖ Interdisciplinary perspective: by integrating insights from radiology, computer science, and data science, we offer a unique interdisciplinary perspective on the current state and future potential of AI in CT image analysis.

❖ Future research directions: we identify and discuss key areas for future research, including innovations in model design, multi-modal learning approaches, and strategies for real-world clinical implementation. These insights are valuable for guiding future studies and funding priorities in the field.

❖ Clinical relevance: our paper bridges the gap between technical advancements and clinical applications, providing clinicians and healthcare professionals with a clear understanding of the potential impact of these technologies on diagnostic accuracy and efficiency.

❖ Ethical and regulatory considerations: we address important ethical and regulatory challenges associated with the implementation of AI in medical imaging, contributing to the ongoing dialogue about responsible AI development in healthcare.

The paper is structured to offer a thorough exploration

Table 1 Comparison of imaging techniques

Imaging method	Applications	Pros	Cons
CT	Screening for lung cancer, diagnosing lung diseases, assessing the extent of lung cancer, and detecting pulmonary embolisms	Identifies small or early-stage lung tumors with exceptional resolution and sensitivity, aiding in the examination of lung nodules	Elevated radiation exposure, possible requirement for contrast material, and significant expenses
PET-CT	Evaluating lung cancer, determining the stage of lung cancer, monitoring treatment outcomes, and detecting cancer relapses	Excellent sensitivity in cancer detection, early identification of cancer, and provision of both anatomical and functional information	False alarms due to inflammation or infection, elevated radiation exposure, cost implications, and potential need for fasting before the scan
X-ray	Identifying rib fractures, detecting pneumonia, and screening for lung cancer	Fast, cost-effective, and widely available	Restricted sensitivity and specificity, potentially leading to the oversight of early-stage lung cancer
Ultrasound	Detecting pleural effusions, guiding thoracentesis procedures, and evaluating diaphragm function	Non-invasive, free of radiation, and applicable at the patient's bedside	Depending on the operator, restricted ability to scan lung tissue; potential obstacles from gas or bone
MRI	Assessment of lung cancer invasion, pulmonary embolism diagnosis, and lung function evaluation	Clear soft tissue contrast, minimal radiation exposure, and capability to evaluate lung function	Extended scanning durations, limited accessibility, elevated expenses, and possible requirements for contrast agents
PET	Cancer detection, brain disorders, cardiac imaging	Functional imaging detects metabolic activity	Exposure to radioactive tracers, expensive, limited availability
SPECT	Cardiac imaging, bone scans, detecting infections	Functional imaging, less expensive than PET	Lower resolution than PET, exposure to radiation
OCT	High-resolution imaging of the retina and other tissues	High-resolution, non-invasive, real-time imaging	Limited to optically accessible tissues, requires specialized equipment

CT, computed tomography; PET-CT, positron emission tomography-computed tomography; MRI, magnetic resonance imaging; PET, positron emission tomography; SPECT, single photon emission computed tomography; OCT, optical coherence tomography.

of this rapidly evolving field. We begin with an introduction that sets the context for our study. We provide an overview of DL methods, followed by an examination of FMs. We then discuss benchmark datasets crucial for model development and evaluation. *Figure 1* show an overview of the review paper. A detailed literature review synthesizes current research findings, methodologies, and applications of DL in CT image analysis. To ensure a rigorous and comprehensive review, we conducted a systematic literature search, the details of which are summarized in *Table 2*. This table outlines our search methodology, including the databases searched, search terms used, timeframe considered, and our inclusion and exclusion criteria. We also cover evaluation metrics, which are essential for assessing model performance, and address challenges and future directions, offering insights into potential research avenues. Finally, we present our conclusions. We present

this article in accordance with the Narrative Review reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-24-1400/rc).

Methods

A comprehensive literature search was conducted using major databases, including Web of Science, PubMed, IEEE Xplore, and Scopus, to identify relevant studies on DL methodologies in CT image analysis. The search encompassed publications from January 1, 2013, to December 31, 2023, focusing on English-language articles to ensure accessibility and relevance. Keywords such as "deep learning", "neural networks", "computed tomography", "image classification", "COVID-19", and "lung nodules" were employed to refine the search results. These terms were combined using Boolean operators

966

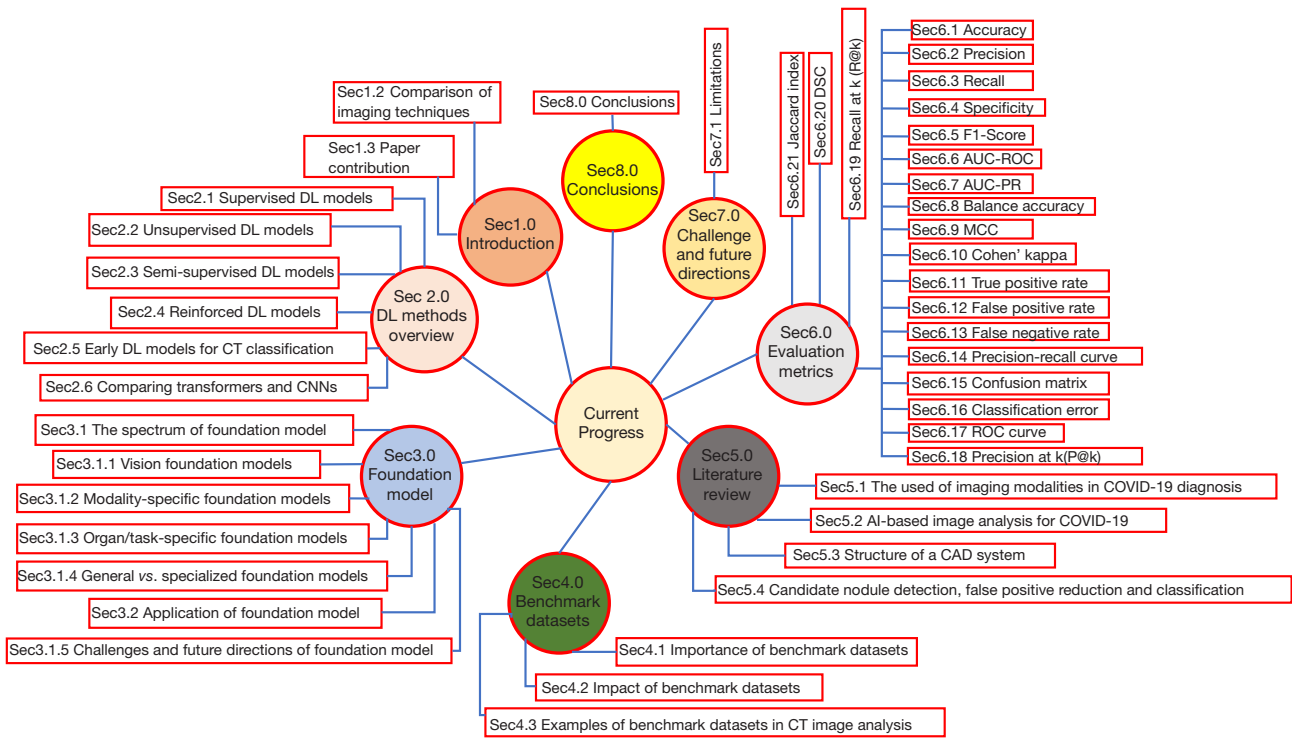Ahmad et al. Deep learning for CT image classification

Figure 1 An overview of the review paper. DL, deep learning; CT, computed tomography; CNNs, convolutional neural networks; DSC, Dice similarity coefficient; AUC-ROC, area under the receiver operating characteristic curve; AUC-PR, area under the precision-recall curve; MCC, Matthews correlation coefficient; CAD, computer-aided diagnosis; COVID-19, coronavirus disease 2019.

to create comprehensive search strings tailored to each database's specific syntax. Inclusion criteria comprised peer-reviewed articles, systematic reviews, clinical studies, and conference proceedings that addressed DL applications in CT imaging, while opinion pieces, non-English articles, and studies focusing solely on other imaging modalities were excluded.

Two independent reviewers screened the titles and abstracts of the initial search results, followed by full-text assessment of potentially relevant studies. Data extraction was performed using a standardized form, capturing information such as study design, DL architecture, CT imaging context, performance metrics, and key findings. The quality of included studies was assessed using appropriate tools for diagnostic accuracy studies. Due to the heterogeneous nature of the studies, a narrative synthesis approach was adopted, organizing the findings thematically. The search strategy aimed to capture a wide range of methodologies and outcomes to provide a thorough overview of the current landscape in this rapidly evolving field. The complete search strategy, including specific search strings

used for each database, is summarized in *Table 2*.

DL methods overview

Acquiring and interpreting medical images accurately is essential for the correct identification and diagnosis of malignant diseases. Various high-resolution imaging devices, such as CT, MRI, and X-ray scans, are available for this purpose. Following pre-processing, the medical image analysis system extracts relevant information from these images to train its models, which can then be utilized to identify diseases in unknown medical images. Traditional ML methods often struggle to provide reliable results due to the significant variations in medical images among individuals. In contrast, DL techniques have emerged as effective tools for analyzing medical images, particularly in the context of disease detection, such as cancer (11). DL methods, a subset of ML techniques, leverage neural networks with multiple layers (input, hidden, and output layers) to achieve more precise model training. These DL models can be categorized into four groups based on their

Table 2 The search strategy summary

Items	Specifications
Date of search	June 15, 2023
Databases and other sources searched	Web of Science, PubMed, arXiv, IEEE Xplore, Scopus
Search terms used	"DL", "FM" AND ("CT imaging" OR "computed tomography") AND ("COVID-19" OR "lung nodule" OR "respiratory illness") AND "Foundation model"
Timeframe	2013–2023
Inclusion criteria	Original research articles, review papers, English language publications, studies focusing on deep learning applications in CT imaging, studies related to COVID-19 detection or lung nodule classification
Exclusion criteria	Non-English publications, studies not related to CT imaging or DL
Selection process	The selection process was conducted by two independent reviewers who screened the titles and abstracts of identified studies. They used predefined inclusion and exclusion criteria to ensure consistency. Any discrepancies between the reviewers were discussed and resolved through consensus, involving a third reviewer when necessary to achieve agreement

DL, deep learning; FM, foundation model; CT, computed tomography.

learning approaches; reinforced learning, unsupervised learning, semi-supervised learning, and supervised learning models (9).

Before the advent of DL, CT image classification heavily relied on traditional methods, each with its own set of limitations. Manual segmentation was a common approach where experts would delineate regions of interest (ROIs) within CT images. Although this method could be accurate, it was extremely time-consuming and subjective, making it impractical for large datasets, especially in scenarios requiring rapid analysis, such as COVID-19 diagnosis or nodule detection. Feature extraction was another critical step in traditional CT image classification. This process involved identifying and quantifying relevant image features, such as texture, shape, and intensity. However, this method depended heavily on domain knowledge and the manual selection of features, which could lead to inconsistencies and limited scalability. For example, in COVID-19 CT classification, where rapid and consistent identification of patterns in lung tissues is crucial, manual feature extraction could introduce delays and variability in diagnosis. Traditional ML algorithms, such as SVM, decision trees, and k-NN, were widely used for classification tasks. These algorithms relied on handcrafted features extracted from CT images and predefined rules for classification. While these methods provided a foundation for early image processing, they struggled with complex tasks like distinguishing between subtle variations in COVID-19 infections or accurately identifying small lung nodules, where the boundary between normal and abnormal

tissue is often indistinct.

The introduction of DL revolutionized CT image classification, especially in challenging areas like COVID-19 diagnosis and nodule detection. DL models, particularly CNNs, automatically learn hierarchical feature representations from raw image data, eliminating the need for manual feature extraction. This ability to learn directly from the data has significantly improved the accuracy and efficiency of CT classification tasks. For instance, in COVID-19 CT classification, DL models have demonstrated remarkable performance in identifying infection patterns, even in early stages, by leveraging large datasets and learning from subtle differences in the images. Similarly, in lung nodule detection, DL models have shown superior capability in distinguishing between benign and malignant nodules by analyzing complex patterns that traditional methods might miss. These models can process vast amounts of data quickly, making them ideal for large-scale screening and early diagnosis, which is crucial in reducing mortality rates for diseases like lung cancer. Furthermore, the evolution of image processing algorithms, as illustrated in *Figure 2*, reflects this shift from traditional ML to more advanced DL techniques. The integration of DL into CT classification has not only enhanced diagnostic accuracy but also opened new avenues for real-time, automated, and scalable solutions in medical imaging. While traditional methods laid the groundwork for CT image classification, the advent of DL has significantly advanced the field (12). These technologies have overcome the limitations of manual segmentation, feature extraction,
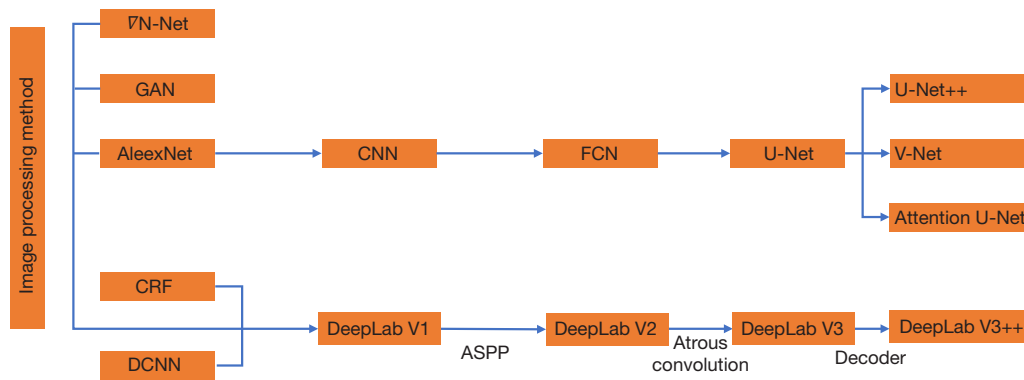
968

Ahmad et al. Deep learning for CT image classification

Figure 2 The evolutional structure of machine learning-based image processing algorithms. GAN, generative adversarial network; CRF, conditional random field; CNN, convolutional neural network; FCN, fully convolutional network.
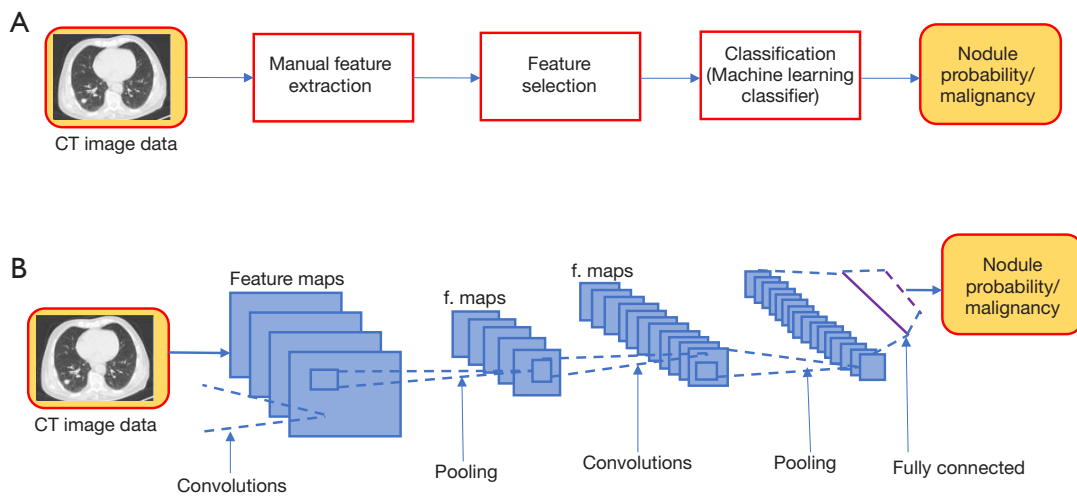

Figure 3 Comparison of traditional methods and typical CNNs workflows in image analysis. (A) Traditional methods workflow. (B) Typical CNN workflow. CT, computed tomography; CNN, convolutional neural network.

and traditional ML algorithms, making them indispensable tools in modern medical imaging, particularly in the context of COVID-19 and lung nodule detection. As DL models continue to evolve, their impact on healthcare will likely expand, offering even greater accuracy and efficiency in diagnostic imaging. Accurate acquisition and interpretation of medical images are crucial for the correct identification and diagnosis of diseases, particularly malignancies. High-resolution imaging modalities such as CT, MRI, and X-ray scans provide the raw data for analysis. The medical image analysis system processes these images, extracts relevant information, and uses it to train models capable of identifying diseases in unknown medical images (9-11).

Rule-based systems: these encoded domain-specific knowledge and rules to make diagnostic decisions, often used for simple classification tasks or decision support systems in medical imaging. *Figure 3* contrasts the workflow of traditional methods (A) with that of typical CNNs (B) the traditional workflow *vs.* typical CNNs work flow. The transition to DL methods has addressed many limitations of traditional approaches, offering improved accuracy, scalability, and the ability to capture complex patterns in medical imaging data without extensive manual feature engineering. This shift has significantly enhanced the field's capability to analyze and interpret medical images, particularly in the context of disease detection and classification. *Figure 4* illustrates the general overview of lung cancer detection-based CT.
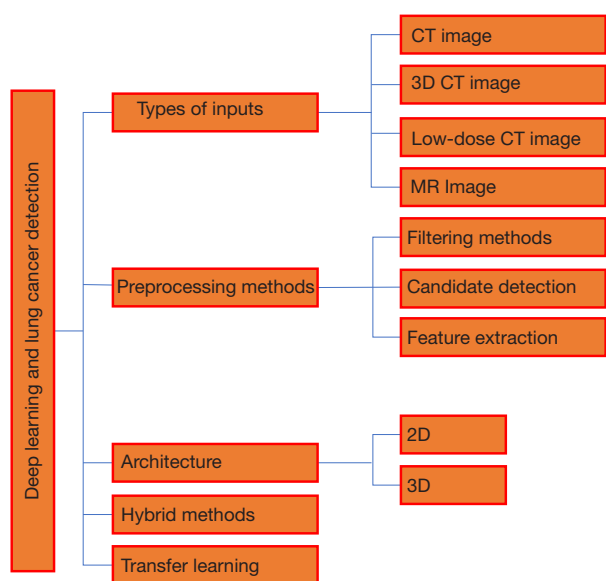
Figure 4 The general overview of lung cancer detection-based CT. CT, computed tomography.

Supervised DL models

Supervised DL models have dramatically improved CT image classification, particularly in detecting COVID-19 and lung nodules, which are essential for early diagnosis and treatment. These models, especially CNNs, are trained on annotated datasets where input images are paired with labels, enabling them to learn intricate patterns in medical images (13). In the context of COVID-19, CNNs have demonstrated high sensitivity and specificity, outperforming traditional diagnostic methods by effectively identifying subtle lung changes like ground-glass opacities. Their ability to generalize across different datasets is crucial for widespread deployment in diverse healthcare settings, though the need for large annotated datasets and variations in imaging protocols present challenges. For lung nodule detection, CNNs have shown significant promise in accurately identifying and classifying nodules, which is vital for early lung cancer detection. Advanced models like Faster R-CNN and U-Net have improved nodule localization and reduced false positives, though variability in nodule appearance and the risk of false positives remain challenges. Techniques such as data augmentation and transfer learning have been employed to enhance model robustness and generalization. Despite their success, DL models face several challenges, including the need for extensive annotated datasets, variability in CT imaging protocols,

and the "black box" nature of CNNs, which can limit their clinical adoption. Future directions include integrating DL with radiomics for more comprehensive analysis, developing federated learning frameworks to utilize decentralized data, and leveraging image improvement solutions like GE True Fidelity to standardize imaging protocols and enhance model performance (14).

Unsupervised DL models

Unsupervised DL models represent a powerful approach in CT image classification, particularly in scenarios like COVID-19 detection and lung nodule analysis where labeled data may be limited or costly to obtain. Unlike supervised models, unsupervised learning does not rely on annotated datasets. Instead, it identifies patterns and structures within the data itself, making it highly valuable for exploratory data analysis and feature extraction. In the context of COVID-19, unsupervised models such as autoencoders and GANs have been employed to detect anomalies in CT scans by learning the underlying distribution of normal lung images. These models can flag deviations from the norm, which may correspond to COVID-19-related abnormalities. This capability is particularly useful in early screening and in situations where labeled COVID-19 datasets are scarce. By focusing on anomaly detection, these models provide a flexible and scalable approach to identifying COVID-19 in diverse populations and imaging conditions. For lung nodule detection, unsupervised models can cluster CT images based on learned features, potentially distinguishing between benign and malignant nodules without prior labeling. Techniques like self-organizing maps (SOMs) and clustering algorithms can help in identifying patterns in nodule characteristics that are not immediately apparent through manual analysis. These models can also be used to pre-process data, reducing noise and enhancing important features, which can then be further analyzed using supervised methods. Despite their promise, unsupervised models face significant challenges. The interpretability of the learned features remains a key issue, as these models often produce results that are difficult to translate into clinical practice. Moreover, the lack of labels can lead to the identification of features that are not clinically relevant. However, by combining unsupervised learning with semi-supervised or supervised approaches, it is possible to refine the feature space and improve the clinical utility of these models. Future directions for unsupervised learning in CT

970

Ahmad et al. Deep learning for CT image classification

image classification include the development of hybrid models that integrate both supervised and unsupervised techniques, improving the balance between accuracy and data efficiency. Additionally, the use of unsupervised models for data augmentation, anomaly detection, and feature extraction can provide valuable insights that enhance the performance of supervised models, particularly in complex tasks like COVID-19 detection and lung nodule classification.

Semi-supervised DL models

A semi-supervised DL model leverages a combination of labeled and unlabeled data for training purposes. In the context of CT imaging, this approach allows the model to learn from a mix of data with and without predefined classifications, enhancing its ability to generalize and make accurate predictions. Commonly utilized DL models in this category for analyzing CT images include RNN, LSTM, gated recurrent unit (GRU), and GAN. These models are specifically tailored to handle the complexities of medical image analysis, enabling them to effectively extract meaningful features and patterns from both labeled and unlabeled CT data (13). Their versatility and adaptability make them valuable tools in advancing the accuracy and efficiency of disease diagnosis and treatment planning in CT imaging. Semi-supervised DL models offer a middle ground between supervised and unsupervised approaches, utilizing both labeled and unlabeled data to enhance the training process. This method is particularly valuable in medical imaging fields like COVID-19 detection and lung nodule analysis, where obtaining large amounts of labeled data is often challenging and expensive. Semi-supervised models can significantly improve the performance of CT image classification by leveraging the vast amount of available unlabeled data alongside a smaller set of labeled examples. This approach is especially useful during the early stages of a pandemic when labeled data might be limited due to the novelty of the disease. Techniques such as consistency regularization and pseudo-labeling allow the model to make use of unlabeled CT scans by predicting labels for these scans and incorporating them into the training process. This method not only enhances the model's ability to generalize from limited labeled data but also accelerates the development of robust diagnostic tools for COVID-19. For lung nodule detection, semi-supervised learning enables the model to learn from a mix of labeled CT scans and a large corpus of unlabeled scans,

which are common in clinical settings. This approach is beneficial in improving the detection and classification of nodules, particularly in distinguishing between benign and malignant cases. Semi-supervised techniques such as graph-based methods and ladder networks can propagate labels from a few labeled examples to a broader set of unlabeled data, thereby improving the model's accuracy in identifying clinically significant nodules. This is crucial for early diagnosis and treatment planning in lung cancer, where the differentiation between benign and malignant nodules can significantly impact patient outcomes.

The success of semi-supervised learning in medical imaging hinges on effectively balancing the contribution of labeled and unlabeled data. One challenge is ensuring that the model does not become overconfident in its predictions of unlabeled data, which could lead to the propagation of errors. However, with carefully designed training protocols, such as enforcing consistency across different augmentations of the same image or using uncertainty-aware models, these risks can be mitigated. The integration of semi-supervised models with other DL techniques, such as transfer learning, could further enhance their effectiveness. For example, pre-trained models on large, labeled datasets from other medical imaging tasks could be fine-tuned using semi-supervised learning on CT scans specific to COVID-19 or lung nodule detection the general overview of lung cancer detection-based CT is shown in *Figure 4*. Additionally, developing models that can efficiently handle the inherent variability in CT scan protocols and patient populations will be key to making semi-supervised learning a standard tool in clinical practice.

Reinforced DL models

Reinforced DL models, which combine the principles of RL with DL, offer a novel approach to medical image analysis, particularly in the classification of CT images for COVID-19 detection and lung nodule identification. These models are designed to optimize decision-making processes by learning policies that maximize cumulative rewards through interactions with an environment, making them well-suited for tasks that involve sequential decision-making, such as detecting abnormalities in medical images. Reinforced DL models can be applied to dynamically adjust the analysis of CT images, focusing on ROIs that are more likely to contain relevant pathological features. For example, a reinforced model can learn to prioritize the examination of lung areas most susceptible to COVID-

19-related changes, such as ground-glass opacities or consolidations, by receiving feedback (rewards) based on the accuracy of its predictions. This dynamic and adaptive approach enables the model to concentrate computational resources on critical areas, improving both the efficiency and accuracy of COVID-19 diagnosis from CT scans. In lung nodule detection, reinforced DL models can similarly enhance the detection process by learning to navigate through 3D CT scans more effectively. These models can be trained to identify and focus on potential nodule regions by simulating a radiologist's decision-making process. For instance, the model can be rewarded for correctly identifying and classifying nodules as benign or malignant, with penalties for false positives and negatives. Over time, the reinforced model refines its policy, improving its ability to distinguish between various types of nodules and reducing the likelihood of misdiagnosis (12).

The application of reinforced learning in these models also allows for more nuanced control over the trade-offs between sensitivity and specificity, which are crucial in medical diagnostics. For example, in COVID-19 detection, a reinforced DL model could be fine-tuned to minimize the risk of false negatives, ensuring that cases are not missed, while in nodule detection, it could be adjusted to reduce false positives, thereby minimizing unnecessary biopsies or interventions. One of the significant advantages of reinforced DL models is their ability to learn from both successes and mistakes, continually improving their performance as they process more data. However, the challenge lies in defining appropriate reward structures that align with clinical goals and ensuring that the model's learning process is stable and converges to optimal policies. Additionally, these models require substantial computational resources and time for training, as they need to simulate many interactions with the environment to learn effective policies, integrating reinforced DL models with other AI techniques, such as supervised and semi-supervised learning, could lead to even more powerful diagnostic tools. For instance, a hybrid approach could involve using reinforced learning to guide the model's attention to specific regions of a CT scan, followed by supervised classification of those regions. Additionally, combining reinforced DL models with domain knowledge from radiologists could enhance the interpretability of the model's decisions, making them more reliable and trustworthy in clinical settings. Reinforced DL models hold significant potential for advancing CT image classification in COVID-19 detection and lung nodule analysis. By learning to optimize decision-making

processes through interaction with the data, these models can improve diagnostic accuracy and efficiency. However, their successful application requires careful consideration of reward structures, computational resources, and integration with other AI methods to maximize their clinical utility.

Early DL models for CT image classification

In the early days of DL for medical imaging, researchers explored various neural network architectures and training strategies to leverage the power of DL for CT image classification (15). These early models laid the foundation for the development of more advanced and specialized DL models tailored to medical imaging tasks. In this section, we will discuss some of the key early DL models that have had a high impact on CT image classification and have been influential in shaping the field of medical imaging. As illustrated in *Figure 4* the general overview of lung cancer detection based CT. However, CNNs have been a cornerstone in DL for image analysis, including CT image classification. CNNs are specifically designed to process visual data such as images by leveraging convolutional layers to extract hierarchical features from raw pixel data (9). CNNs have been used for tasks such as tumor detection, organ segmentation, and disease classification. Early CNN models proposed by Krizhevsky *et al.* (16), demonstrated the power of DL in image classification tasks and laid the groundwork for more advanced architectures.

The success of CNNs in CT image classification can be attributed to their ability to automatically learn and extract hierarchical features from raw pixel data. CNNs consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers, that work together to capture spatial relationships and patterns in images. Convolutional layers apply filters to the input image to extract local features while pooling layers downsample the feature maps to reduce computational complexity. Fully connected layers combine the extracted features to make classification decisions based on learned representations. One of the key strengths of CNNs is their capability to learn complex and abstract features directly from the data, eliminating the need for manual feature engineering. This end-to-end learning approach enables CNNs to capture intricate patterns and relationships in CT images, leading to improved accuracy and generalization performance. By leveraging large-scale annotated datasets, CNNs can be trained to recognize subtle differences in image features and make accurate predictions, aiding radiologists in diagnosing
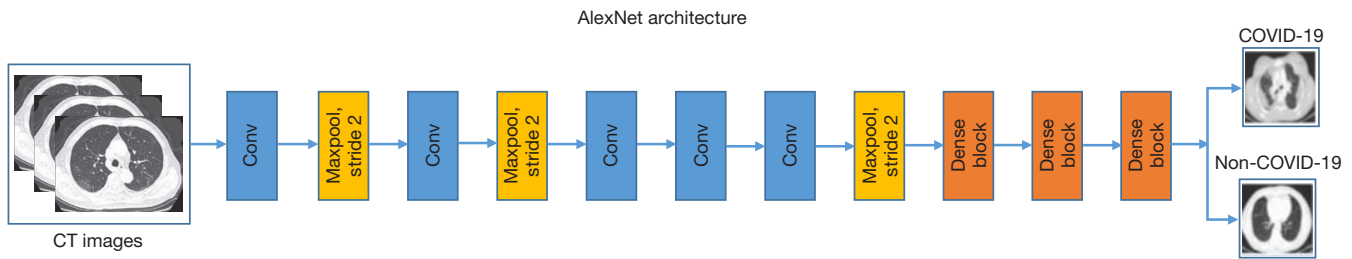
AlexNet architecture

Figure 5 AlexNet model for detecting COVID-19 based on CT images CT, computed tomography; COVID-19, coronavirus disease 2019.

and interpreting medical images (17). CNNs have been applied to a wide range of clinical tasks, including but not limited to:

- ❖ Tumor detection: CNNs have been used to automatically detect and classify tumors in CT scans, enabling early diagnosis and treatment planning for cancer patients. By learning from labeled examples of tumor and non-tumor regions, CNNs can accurately localize and segment tumors in CT images, assisting radiologists in identifying suspicious lesions.

- ❖ Organ segmentation: CNNs have been employed for segmenting and delineating anatomical structures in CT images, such as the brain, lungs, liver, and heart. By training on annotated datasets of organ boundaries, CNNs can generate precise segmentation masks that facilitate quantitative analysis and volumetric measurements of organs for clinical assessment.

- ❖ Disease classification: CNNs have demonstrated high performance in classifying different types of diseases and abnormalities in CT images, such as pneumonia, fractures, and pulmonary embolism. By learning from diverse examples of diseased and healthy tissues, CNNs can differentiate between different pathologies and provide diagnostic support to healthcare providers.

- ❖ The landscape of DL architectures for CT image classification has expanded to include various specialized models: advanced CNNs: VGG, ResNet, and DenseNet have further improved feature extraction and classification capabilities. U-Net: particularly effective for medical image segmentation tasks. 3D CNNs: developed to capture spatial and temporal information in volumetric CT data. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks: these architectures excel in analyzing

sequential CT scans, enabling disease progression tracking and treatment response monitoring. Generative adversarial networks (GANs): employed for data augmentation and synthetic CT image generation, addressing data scarcity issues and improving model generalization.

The integration of CNNs into CT image classification workflows has significantly advanced the field of medical imaging, enabling automated and efficient analysis of complex and large-scale imaging data. By leveraging the capabilities of CNNs to learn from vast amounts of labeled CT images, researchers and clinicians can improve the accuracy, speed, and consistency of image interpretation, ultimately enhancing patient care and outcomes. In this review paper, we will delve into the recent advancements and applications of CNNs in CT image classification, highlighting their impact on the field of medical imaging and discussing future directions for research and clinical implementation (18). These early DL models, including CNNs, AlexNet, VGGNet, GoogLeNet, and ResNet, have had a significant impact on the field of CT image classification and have paved the way for more advanced and specialized DL architectures tailored to medical imaging tasks (19). *Figure 5* shown an AlexNet model for detecting COVID-19 based on CT images proposed by Cortés and Sánchez (20). *Figure 6* shows the ML-based diagnosis models and their evolutional structure *Figure 7* shows the VGG19 architecture for COVID-19 detection proposed by Zouch *et al.* (21) and *Figure 8* shows a GAN model proposed by Hage Chehade *et al.* (22). These models have demonstrated the power of DL in capturing intricate patterns and features in CT images, leading to improved accuracy and efficiency in automated image analysis (9).

Hybrid models in CT image classification combine CNNs with RNNs or Transformers, offering enhanced performance in analyzing medical imaging data. These models leverage CNNs' strength in spatial feature extraction
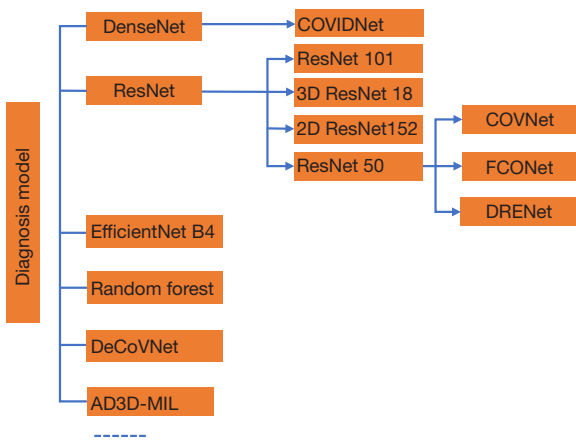
Figure 6 The ML-based diagnosis models and their evolutional structure. DRENet, Dense Residual Ensemble Network; ML, machine learning.

and RNNs' or Transformers' ability to capture temporal patterns or long-range dependencies. The integration of these architectures enables more comprehensive analysis of CT scans, improving accuracy in tasks such as tumor detection, organ segmentation, and disease classification. Hybrid models can effectively capture both spatial and temporal information in sequential CT scans, leading to more robust and reliable predictions.

Transformers, originally developed for natural language processing, have recently made significant strides in CT image classification *Figure 9* shows a standard transformer block. Vision transformers (ViTs) adapt the transformer architecture to image analysis by treating images as sequences of patches. *Figure 10* shows an overview of ViTs (left) and an illustration of the Transformer encoder (middle), and CNN block (right). The method for
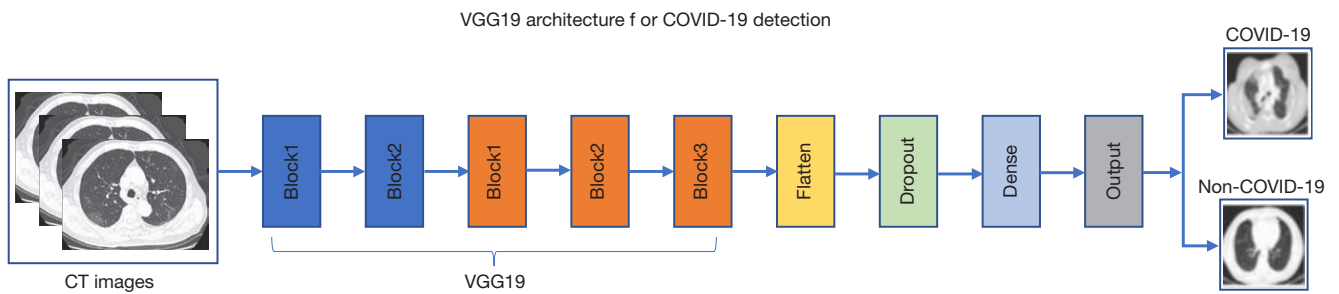

Figure 7 VGG19 architecture f or COVID-19 detection. CT, computed tomography; VGG19, Visual Geometry Group Network with 19 layers; COVID-19, coronavirus disease 2019.
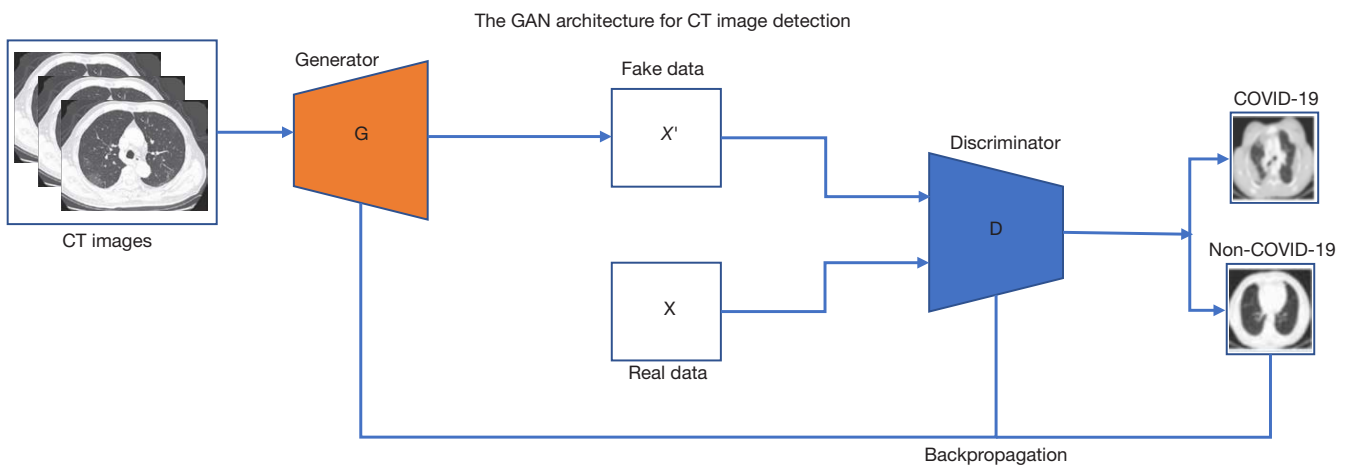

Figure 8 The GAN architecture for CT image detection. CT, computed tomography; COVID-19, coronavirus disease 2019; GAN, generative adversarial network.
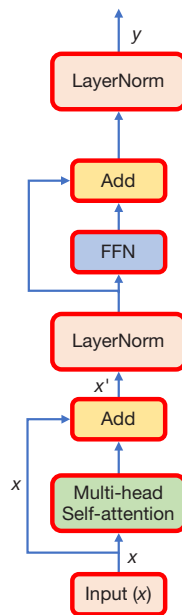
Figure 9 Standard transformer block typically consists of multiple layers of self-attention mechanisms and feedforward neural networks. The self-attention mechanism allows the model to weigh the importance of different input tokens when making predictions, while the feedforward neural network processes the information learned from the attention mechanism. FFN, feedforward neural network.

processing an image involves splitting it into multiple fixed-size patches, which are then handled as sequences using an effective Transformer approach derived from NLP (23). Key advantages of transformers in CT image classification include their ability to capture long-range dependencies, scalability to large datasets, and reduced inductive bias compared to CNNs. Recent advancements include the development of hybrid models combining CNNs and transformers, the use of pre-trained models for transfer learning, and the creation of more efficient transformer architectures. These innovations have led to improved performance and generalization in CT image classification tasks, pushing the boundaries of medical imaging technology and enhancing patient care through more accurate and efficient diagnosis.

Comparing transformers and CNNs for image analysis

Table 3 compares the transformer with CNN, the CNN model has traditionally shown strong performance in image analysis, but ViTs have demonstrated comparable or even superior results, particularly when pre-trained or when scaled datasets are available (Dosovitskiy *et al.*, 2020) (24). This raises the question of how Transformers and CNNs differ in their approach to understanding images.

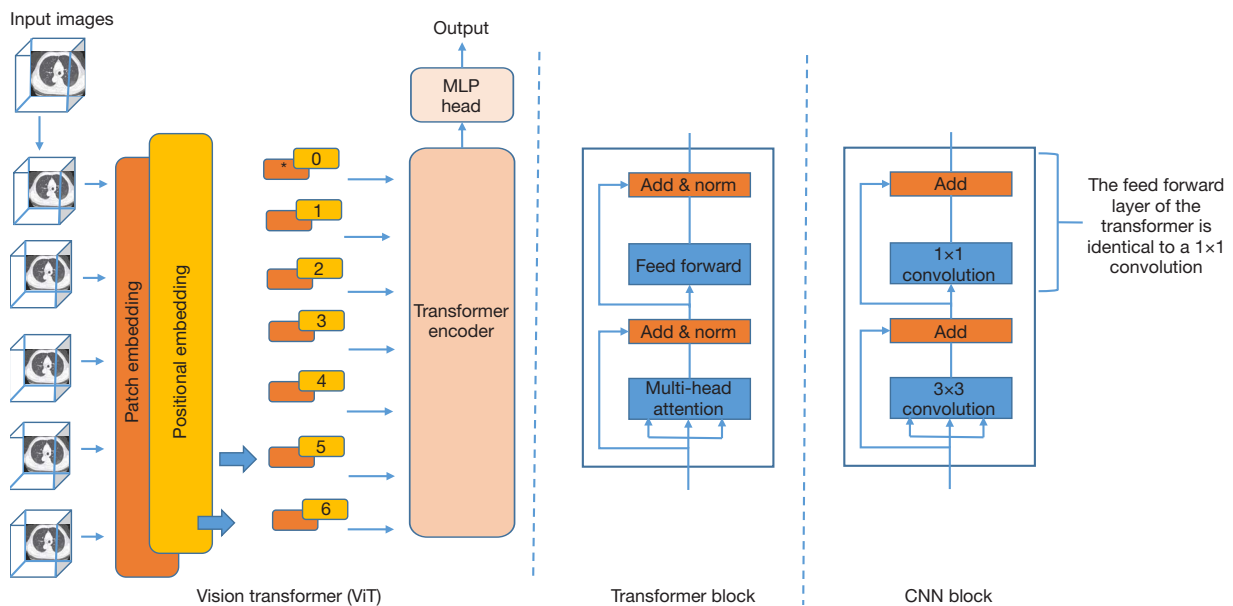**Figure 10** Overview of vision transformer (left), illustration of the Transformer block (middle), and the CNN block (right). The method for processing an image involves splitting it into multiple fixed-size patches, which are then handled as sequences using an effective Transformer approach derived from NLP. MLP, multi-layer perceptron; CNN, convolutional neural network; NLP, natural language processing.

Table 3 Comparing transformers and CNNs for image analysis

Feature	CNNs	ViTs
Architecture	Convolutional layers, pooling layers, fully connected layers	Self-attention mechanism, feed-forward layers
Receptive field	Gradually expands with depth	Global receptive field from the lowest layer
Feature extraction	Layer-by-layer hierarchical extraction	Global feature extraction using self-attention
Locality	Excellent at capturing local structures	Can capture global and local structures simultaneously
Weight sharing	Convolutional layers share weights	No weight sharing; each position computes attention independently
Handling long-range dependencies	Limited due to local receptive fields	Excellent due to global self-attention
Pooling operations	Uses pooling to reduce spatial dimensions	No explicit pooling; positional encoding is used instead
Scalability	Scales well with moderate datasets	Requires large-scale datasets for optimal performance
Training data requirements	Performs well with smaller datasets	Performs best with large pre-trained models and datasets
Computational efficiency	More efficient for smaller and moderate-sized data	Computationally intensive, especially for large models
Parameter efficiency	Efficient use of parameters due to weight sharing	Requires more parameters due to lack of weight sharing
Interpretability	Easier to interpret layer outputs	Harder to interpret due to complex attention mechanisms
Strengths	Effective at local pattern recognition, robust with limited data	Excels at capturing global context, superior for long-range dependencies
Limitations	Struggles with capturing global context and long-range dependencies	Requires large datasets and computational resources, harder to interpret

CNNs, convolutional neural networks; ViTs, vision transformers.

CNNs and Transformers both have unique strengths and weaknesses, making them suitable for different aspects of image analysis. CNNs excel at local pattern recognition with efficient use of parameters and computational resources, making them robust for tasks with limited data. On the other hand, Transformers, with their global receptive field and self-attention mechanisms, excel in capturing long-range dependencies and complex global patterns, although they require larger datasets and computational power. Hybrid models that combine the strengths of both architectures are also emerging, providing enhanced performance for various image analysis tasks, including CT image classification. These hybrid approaches demonstrate the potential to harness the benefits of both CNNs and Transformers, leading to improved performance in various image analysis tasks, including CT image classification, the *Figure 11* illustrates the taxonomy of typical approaches in combining CNNs and Transformers.

FMs

FM in CT image classification leverages pre-trained DL models and large-scale datasets to enhance performance in medical imaging tasks. These models, which include transfer learning approaches and pre-training on extensive datasets, serve as a basis for developing more specialized and task-specific models. The concept draws inspiration from the evolution of language models like the generative pre-trained transformer (GPT) series, which have shown remarkable advancements in natural language processing. From GPT-1's introduction in June 2018 to the release of GPT-4-turbo in November 2023 the history of FM is shown in *Figure 12*. The FMs in these models have demonstrated the power of pre-training on large datasets followed by fine-tuning for specific tasks. FMs aim to apply similar principles, utilizing the knowledge gained from vast amounts of medical imaging data to improve accuracy,
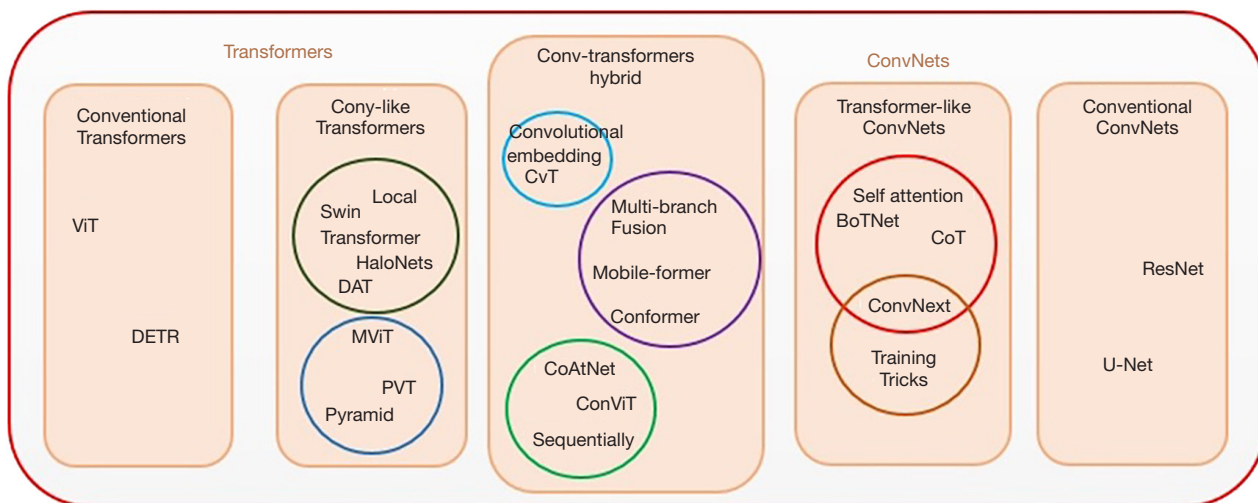
Figure 11 Taxonomy of combining CNNs and Transformer. ViT, vision transformer; DETR, detection transformer; DAT, dynamic attention transformer; MViT, multiscale vision transformer; PVT, pyramid vision transformer; CvT, convolutional vision transformer; CoT, chain of thought; CNN, convolutional neural network.
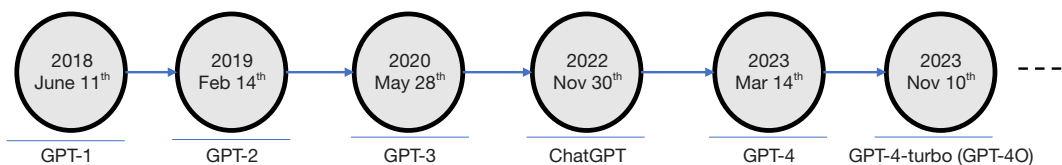

Figure 12 History of foundation models. GPT, generative pre-trained transformer.

efficiency, and generalization in various diagnostic tasks. This approach has the potential to significantly advance the field of medical image analysis, enabling more sophisticated and versatile AI applications in healthcare.

FM originally developed for natural language processing and computer vision, such as BERT, GPT, CLIP, and ViT, have been successfully adapted for medical imaging tasks, including CT image classification. These models leverage their pre-trained capabilities in language understanding, generative tasks, multimodal fusion, and attention mechanisms to enhance the performance of DL algorithms in medical image analysis. By fine-tuning these models on medical imaging datasets, researchers have improved various aspects of CT image classification, including feature extraction, image-text fusion, report generation, and anomaly detection. The adaptation of these FMs has significantly advanced the field of medical imaging, offering improved accuracy, efficiency, and interpretability in diagnostic tasks and clinical decision-making *Figure 13* shows GPT from training to output of CT images.

The spectrum of FMs

Vision FMs

FMs trained on natural images can be adapted for medical tasks using specialized algorithms (25). However, the lack of high-quality annotations in medical imaging has hindered the development of large-scale DL models for clinical use (26). While medical professionals can provide a few sample cases, extensive hand-labeling is challenging. Vision FM, trained on diverse visual data, offer a starting point for medical applications. Yet, medical images have unique characteristics that differ from natural images, requiring tailored approaches to adapt these models effectively. Strategies like fine-tuning, adapters, prompting, and architectural modifications are crucial for optimal performance in medical contexts (26).

The Segmentation Anything Model (SAM) (27), despite its success with natural images, shows limitations in complex medical tasks. To improve its performance, researchers can employ fine-tuning (28), specialized
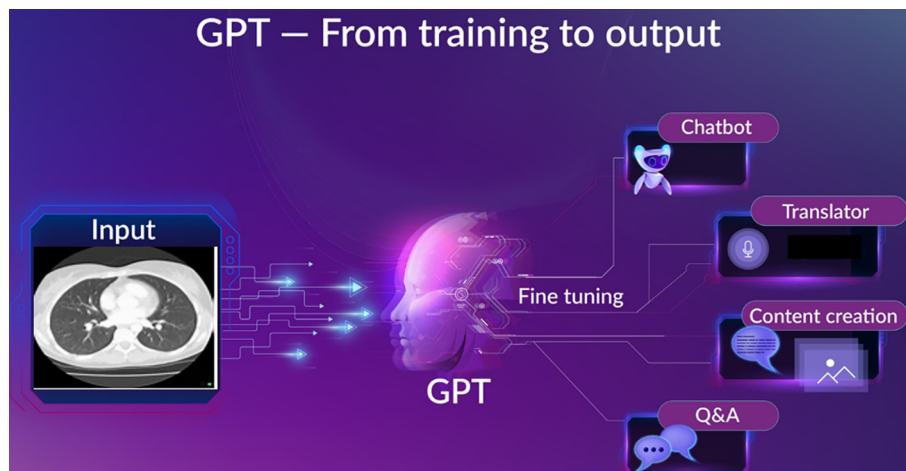
Figure 13 GPT from training to output of CT images. GPT, generative pre-trained transformer; CT, computed tomography.

adapters, or effective prompting strategies (27). Combining SAM with other algorithms or integrating it into medical imaging software could enhance its utility in medical applications (29). A single, universal FM approach may not achieve top performance across all medical image analysis tasks due to the wide variety in anatomical structures, textures, and imaging modalities. Researchers are exploring efficient methods to adapt vision foundation models to the challenges posed by diverse medical data (25-30).

Modality-specific FMs

Medical professionals utilize a variety of imaging techniques for diagnosis and treatment, each suited to different medical conditions. These include X-rays, CT (29), MRI, Ultrasound, and PET. FMs can be developed to specialize in specific imaging modalities or groups of related modalities. For instance, a radiology-focused FM might encompass X-ray, CT, MRI, and ultrasound, while a 3D imaging FM could handle volumetric data from CT and MRI scans. Alternatively, FMs can be designed for a single modality, such as CT imaging. CT-specific models are engineered to extract and interpret features unique to CT scans. These models excel at analyzing: density variations, contrast enhancement, bone and calcification detection, lung and airway assessment, vascular imaging, multi-planar reconstruction (31).

However, general vision FMs trained on natural images provide a broad foundation for medical image analysis, modality-specific FMs like those focused on CT can leverage the unique attributes of CT imaging. This specialization often results in superior accuracy and

efficiency for CT-specific tasks, such as automated lesion detection, organ segmentation, or quantitative analysis of tissue characteristics (31).

Organ/task-specific FMs

FMs can be customized for specific medical organs or diagnostic tasks, like segmentation (31,32), to address the challenges posed by varying organ appearances across different medical imaging modalities and the wide array of clinical image analysis tasks. While gathering sufficient labeled data to train these specialized FMs is demanding, the resulting models offer improved accuracy and interpretability. A key advantage of these organ/task-specific FMs is their ability to reduce the amount of labeled data needed for new, related tasks. This is because they have already learned relevant features during their initial training. By focusing on particular organs or tasks, these models are optimized for their specific applications, leading to enhanced effectiveness and reliability in clinical settings (31). *Figure 14* that demonstrates the significant visual differences between various imaging modalities. These differences present a substantial challenge in developing a single, comprehensive FM that can effectively handle all types of medical images.

General *vs.* specialized FMs

In the field of expertise, specialists have deep knowledge in a specific area, while generalists have a broader understanding across one or multiple fields. In medical image analysis, a general AI system is a versatile platform that can perform various tasks like classification, detection, segmentation, and registration on different types of medical images across
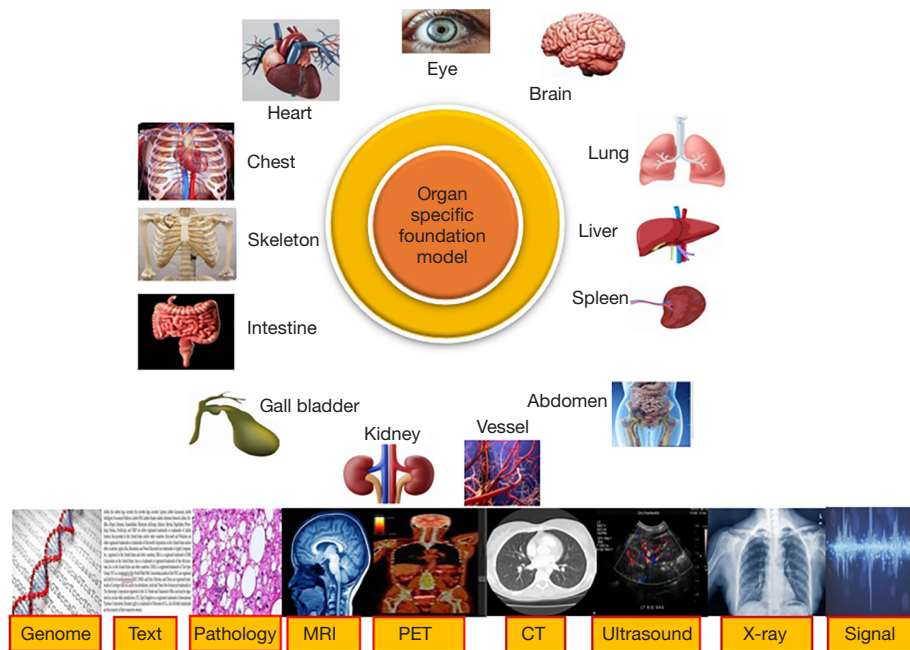
Figure 14 It illustrates that various image modalities exhibit substantial image-level differences, posing challenges during the training of a cohesive foundational model. Each modality, influenced by unique imaging characteristics, yields images with notable variations in organ appearance and associated structures, thereby influencing the choice of modality based on the target organ pathology. By leveraging fine-grained models to understand organ appearance and pathology across modalities, essential clinical techniques and tools, including reliable CADs and surgical planning, can be enhanced. MRI, magnetic resonance imaging; PET, positron emission tomography; CT, computed tomography; CADs, computer-aided diagnosis.

organs and diseases, using a single set of model parameters. On the other hand, specialized AI systems are designed for particular clinical tasks, such as detecting lung nodules or diagnosing liver cancer. These systems usually focus on a specific organ and imaging modality, placing them towards the more specialized end of the FM spectrum (33).

The computer science community is showing increased interest in developing general AI frameworks, driven by advancements in large, multimodal generative models capable of processing diverse medical data. However, most research in academia, medical institutions, and industry still concentrates on developing specialized AI systems. This focus is due to several reasons: Most current state-of-the-art medical image analysis systems use a single imaging modality and are trained for a specific task. AI primarily serves as an assistant to medical professionals who need targeted support aligned with their expertise. Specialized systems often perform better and more accurately on specific tasks. General AI systems typically require significantly more computational resources and may lack the necessary accuracy for specific medical tasks (34). Both specialized

and general AI systems have their own advantages and are suitable for different applications. Therefore, a thorough exploration of the entire FM spectrum is recommended to find the optimal balance between development efforts and practical effectiveness in medical image analysis.

Application of FMs

FMs in CT image analysis offer numerous applications and benefits, addressing key challenges in medical imaging. They excel in handling imbalanced datasets and improving rare disease detection through few-shot learning and data augmentation techniques. FMs enhance model interpretability and transferability, crucial for clinical trust and generalization across diverse medical settings, *Figure 15* shows various applications of FMs. They also support privacy-preserving methods, enabling knowledge sharing without compromising patient data security. The integration of FMs with large language models opens up possibilities for advanced vision-language applications in healthcare, such as automated report generation and improved decision
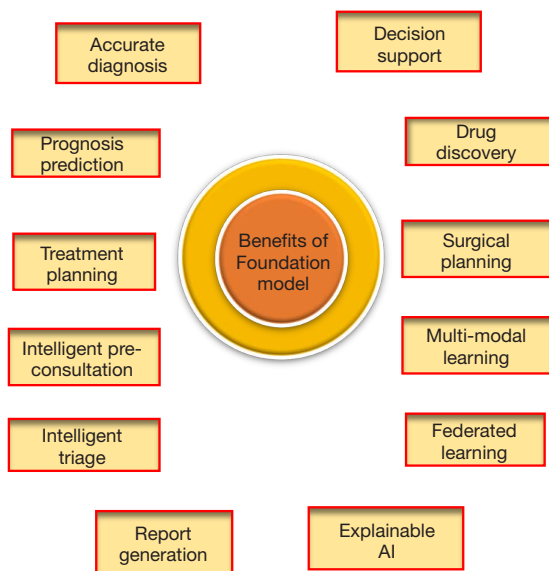
Figure 15 FMs trained on multimodality datasets will lead to comprehensive clinical solutions. AI, artificial intelligence; FMs, foundational models.

support systems. Specific applications include nodule detection and classification, organ and tissue segmentation, lesion characterization, disease progression tracking, and radiomics feature extraction. These capabilities lead to more accurate diagnoses, personalized treatment plans, and improved patient outcomes. Additionally, FMs facilitate multimodal data integration, combining CT images with other clinical data for comprehensive patient assessments. Overall, FMs significantly advance medical imaging analysis, offering the potential for more efficient, accurate, and personalized healthcare solutions.

Challenges and future directions FMs

FM-based CT image classification faces several challenges despite its significant potential. The primary issues include the need for substantial computational resources, which can limit accessibility and raise environmental concerns. Data scarcity and quality remain critical, as high-quality annotated medical imaging datasets are essential for effective training. Bias and fairness are ongoing concerns, as these models may inadvertently perpetuate biases present in their training data. Interpretability is another major challenge, as the complexity of these models often makes it difficult to understand their decision-making processes, which is crucial in medical applications. Additionally, the

field grapples with issues related to privacy preservation, especially when dealing with sensitive medical data. Future directions should focus on developing more efficient training methods, improving data collection and annotation processes, implementing robust bias detection and mitigation strategies, enhancing model interpretability, and strengthening privacy-preserving techniques. Addressing these challenges will be crucial for the widespread adoption and ethical use of FM in CT image classification and broader medical imaging applications.

Benchmark datasets and their impact in CT image analysis

The automation of lung abnormality detection and classification, including both cancer nodules and COVID-19-related findings, heavily relies on the availability of comprehensive datasets. These datasets are fundamental to achieving reliable performance results using computational techniques, serving as the foundation for developing and validating algorithms. Publicly accessible datasets for the detection, identification, and classification of lung abnormalities have become invaluable resources in the field. It is important to distinguish between the tasks involved: lung cancer detection focuses on differentiating between nodules and non-nodules within the lungs, while classification involves distinguishing between benign and malignant nodules. Similarly, COVID-19 detection aims to identify characteristic patterns associated with the viral infection. To provide a clear overview of the available resources, *Table 3* presents a comprehensive summary of lung imaging datasets, including both cancer and COVID-19-related collections. This table includes crucial information such as the dataset name, release date, number of samples, dataset size in gigabytes, total image count, imaging modality, image dimensions, file format, and the availability of ground truth annotations. The datasets are arranged chronologically based on their release dates, offering insights into the evolution of data resources in this field.

The emergence of the COVID-19 pandemic has led to the rapid development and release of specific datasets focused on CT and X-ray images of COVID-19 patients (5). These datasets have become crucial for the development of AI-driven diagnostic tools and for understanding the radiological manifestations of the disease. Notable COVID-19 datasets include the COVID-CT Dataset, which contains CT scans from COVID-19 patients and

980

Ahmad et al. Deep learning for CT image classification

normal controls, and the COVIDx CT Dataset, which offers a large-scale dataset of CT images for COVID-19 detection. The COVID-19 Image Data Collection, while primarily focused on X-rays, also includes some CT images and has been widely used in research. These datasets have unique characteristics compared to lung cancer datasets, such as the rapidity of their collection and release, the evolving nature of the disease understanding reflected in the annotations, and the global collaborative efforts in their creation. The integration of COVID-19 datasets into lung imaging research has not only advanced our ability to detect and manage the disease but has also pushed the boundaries of rapid dataset creation and AI model development in response to global health crises. Understanding these trends across both cancer and COVID-19 datasets is crucial for researchers aiming to benchmark their algorithms against established standards or to identify gaps in current datasets that might inspire the creation of new, more comprehensive resources. The rapid development and utilization of COVID-19 datasets alongside established cancer imaging datasets demonstrate the adaptability and responsiveness of the medical imaging research community to emerging health challenges.

Importance of benchmark datasets

Benchmark datasets serve as critical reference points for evaluating the performance of various algorithms and models in CT image analysis. These curated collections of images, accompanied by expert-validated ground truth annotations, provide a standardized foundation for assessing the efficacy of computational methods across a range of tasks. In the realm of CT imaging, these tasks encompass critical areas such as tumor detection, organ segmentation, disease classification, and more recently, COVID-19 identification. By offering a common set of images and annotations, benchmark datasets enable researchers and developers to conduct fair and meaningful comparisons of their methods against existing state-of-the-art techniques. This standardization is particularly crucial in the medical imaging field, where the stakes of accuracy and reliability are exceptionally high. Furthermore, benchmark datasets play a vital role in driving innovation by highlighting areas where current methods fall short, thereby guiding future research directions and fostering healthy competition within the scientific community.

The significance of benchmark datasets extends beyond mere performance comparison. They facilitate the development and adoption of standardized evaluation metrics and protocols, ensuring objective and consistent assessments of algorithm performance across different studies and institutions. This standardization is essential for building trust in AI-driven medical imaging solutions and for translating research findings into clinical practice. Moreover, benchmark datasets contribute to the reproducibility of research results, a cornerstone of scientific progress. They allow other researchers to verify claims, build upon existing work, and adapt methods to new contexts. In the rapidly evolving field of CT image analysis, where new techniques and models are constantly emerging, benchmark datasets provide a stable reference point for measuring progress over time. They also play a crucial role in addressing challenges such as data scarcity and bias in medical imaging, offering researchers access to diverse, high-quality data that might otherwise be difficult to obtain due to privacy concerns or resource limitations. As the field continues to advance, the development and maintenance of comprehensive, up-to-date benchmark datasets remain paramount in pushing the boundaries of what is possible in CT image analysis and, ultimately, in improving patient care through more accurate and efficient diagnostic tools.

Impact of benchmark datasets

Benchmark datasets have profoundly influenced the field of CT image analysis, serving as catalysts for innovation, collaboration, and scientific progress. These standardized data collections have significantly advanced algorithm development by providing researchers with high-quality, annotated data for training and testing, enabling the creation of more accurate and robust models. They facilitate comparative studies, allowing researchers to evaluate different techniques on a level playing field, leading to valuable insights into best practices and areas for improvement. Benchmark datasets foster collaboration and reproducibility in the scientific community by offering a common reference point for evaluating and sharing results, promoting transparency, and enabling researchers to build upon existing work. Perhaps most importantly, these datasets drive innovation and breakthroughs by challenging researchers to push the boundaries of what is possible in CT image analysis. By setting high-performance standards on benchmark tasks, they have inspired the development of novel algorithms, architectures, and techniques that have significantly advanced the state-of-the-art in medical imaging. This continuous cycle of improvement, facilitated

Table 4 A variety of public datasets containing a substantial volume of lung CT scans

Database	Year	Sample number	Patients number	Modality	Data access	Collection status	Annotation
VIA/IELCAP (35)	2003	N/A	50	CT	Available	Complete	Nodule position and type
NELSON (36)	2003	15,822	N/A	CT	N/A	Complete	N/A
QIN LUNG CT (37)	2015	47	47	CT	Available	Complete	N/A
LungCT-Diagnosis (38)	2015	61	61	CT	Available	complete	Image position
ACRIN-NSCLC-FDG-PET (39)	2013	242	3,377	PT, CT, MR, CR, DX, SC, NM	Available	Complete	Clinical data
LIDC-IDRI (40)	2011	N/A	1,018	CT, CR, DX	Available	Complete	Nodule characteristics, type, and position
NSCLC-Radiomics (41)	2014	422	1,265	CT, RTSTRUCT, SEG	Available	Complete	Participant characteristic and diagnostics information
Mosmed COVID-19 CT Scans (42)	2020	1,000	–	CT	Available	Complete	N/A
COVID-19 CT Lung and Infection Segmentation Dataset COVID-19-CT-CXR (43)	2020	1,327	263	CT	Available	Complete	N/A
COVID-19 CT segmentation Dataset (44)	2020	100	40	CT	–	–	–

CT, computed tomography; N/A, not available; PT, pre-training; MR, medical records; CR, computed radiography; DX, digital radiography; SC, secondary capture; NM, nuclear medicine; RTSTRUCT, radiotherapy structure set; SEG, segmentation; COVID-19, coronavirus disease 2019.

by benchmark datasets, has led to tangible advancements in diagnostic accuracy, efficiency, and ultimately, patient care in the field of radiology and medical imaging. *Table 4* displays a variety of public datasets containing a substantial volume of lung CT scans.

Examples of benchmark datasets in CT image analysis

Several benchmark datasets have been established for CT image analysis, each focusing on specific clinical tasks and challenges. Some notable examples include:

(I) The Cancer Imaging Archive (TCIA)-Lung CT Segmentation Challenge dataset. Link: https://wiki.cancerimagingarchive.net/display/Public/Lung+CT+Segmentation+Challenge; description: provides CT images of the lungs with manual segmentations for lung segmentation tasks.

(II) Medical Decathlon Dataset-Brain Tumor Segmentation. Link: http://medicaldecathlon.com/; description: contains CT images of brain tumors for tumor segmentation research.

(III) Radiological Society of North America (RSNA) Pneumonia Detection Challenge Dataset. Link: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data; description: includes chest CT images for pneumonia detection tasks.

(IV) The Cancer Genome Atlas (TCGA) - Pancreatic Cancer Dataset. Link: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga; description: provides CT images of pancreatic cancer for research and analysis.

(V) The National Institutes of Health (NIH) Chest X-ray Dataset. Link: https://nihcc.app.box.com/v/ChestXray-NIHCC; description: large collection of chest radiographs, including CT images, for chest pathology analysis.

(VI) The Medical Segmentation Decathlon - Liver Tumor Segmentation. Link: http://medicaldecathlon.com/; description: dataset for liver tumor segmentation tasks in CT images.

(VII) The Multimodal Brain Tumor Segmentation Challenge (BRATS) Dataset. Link: https://www.med.upenn.edu/sbia/brats2019.html;

description: contains CT images of brain tumors for segmentation and classification tasks.

(VIII) The RSNA Intracranial Hemorrhage Detection Dataset. Link: https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection/data; description: provides CT images for detecting intracranial hemorrhage in the brain.

(IX) The NIH DeepLesion Dataset. Link: https://nihcc.app.box.com/v/DeepLesion; description: large-scale dataset of CT images with annotations for lesion detection and classification.

(X) The Ischemic Stroke Lesion Segmentation (ISLES) Challenge Dataset. Link: http://www.isles-challenge.org/; description: dataset for ischemic stroke lesion segmentation in CT images.

(XI) The CQ500 Dataset. Link: https://headctstudy.qure.ai/; description: contains head CT images for brain pathology detection and classification.

(XII) The Japanese Society of Radiological Technology (JSRT) Database. Link: https://www.jsrt.or.jp/jsrt-db/eng.php; description: dataset of chest CT images for lung nodule detection and classification.

(XIII) The Liver Tumor Segmentation (LiTS) Challenge Dataset. Link: https://competitions.codalab.org/competitions/17094; description: provides CT images of liver tumors for segmentation and analysis.

(XIV) The Musculoskeletal Radiographs (MURA) Dataset. Link: https://stanfordmlgroup.github.io/competitions/mura/; description: includes CT images for musculoskeletal pathology detection and analysis.

(XV) The International Skin Imaging Collaboration (ISIC) Melanoma Dataset. Link: https://www.isic-archive.com/; description: contains CT images for melanoma detection and classification tasks.

(XVI) The ChestX-ray8 Dataset. Link: https://stanfordmlgroup.github.io/competitions/chexpert/; description: large-scale dataset of chest radiographs, including CT images, for chest pathology analysis.

(XVII) The CheXpert Dataset. Link: https://stanfordmlgroup.github.io/competitions/chexpert/; description: dataset of chest radiographs, including CT images, for disease classification and detection.

(XVIII) The MIMIC-CXR Dataset. Link: https://physionet.org/content/mimic-cxr-jpg/2.0.0/; description: contains chest radiographs, including CT images, for disease detection and analysis.

(XIX) The NIH DeepCT Dataset. Link: https://deepct.nih.gov/; description: large-scale dataset of CT images for various clinical tasks in medical imaging.

(XX) The SIIM-ACR Pneumothorax Segmentation Dataset. Link: https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation/data; description: provides CT images for pneumothorax segmentation tasks in medical imaging.

The benchmark datasets play a critical role in advancing research and innovation in CT image analysis by providing standardized data for algorithm development, evaluation, and comparison. Researchers can accelerate progress in CT image analysis and contribute to the development of more effective and reliable algorithms for clinical practice.

Literature review

The use of imaging modalities in COVID-19 diagnosis

Medical imaging, particularly CT scans, plays a crucial role in diagnosing COVID-19, complementing RT-PCR testing. COVID-19 (19) patients exhibit typical imaging characteristics, including ground-glass opacities (GGO), consolidation (45), bilateral patchy shadowing, and crazy-paving patterns (46). The American College of Radiology (ACR) and the Radiological Society of North America (RSNA) (47) have approved a classification system for identifying COVID-19 and pneumonia on CT scans. CT imaging is valuable for screening asymptomatic or atypical patients, as it can detect typical radiographic findings even in cases with negative RT-PCR results. COVID-19 pneumonia exhibits distinct radiological features compared to other types of pneumonia, including bilateral involvement, peripheral distribution, and fine reticular opacity. The progression of COVID-19 pneumonia is characterized by rapid infiltration in lung lobes, with late-stage patients showing spider webs and crazy-paving patterns on CT scans. *Table 5* shows the classification system for identifying COVID-19 and pneumonia. These imaging modalities provide valuable insights into the disease's characteristic features and aid in differentiating COVID-19

Table 5 Classification system for identifying COVID-19 and pneumonia

COVID-19 pneumonia imaging classification	CT findings	Suggested reporting language
Typical appearance	Reveals peripheral, bilateral, multilobar ground-glass opacities, possibly with consolidation or "crazy-paving" pattern	CT findings are suggestive of COVID-19 pneumonia, but differential diagnoses include other viral pneumonia, organizing pneumonia, toxicity, and connective tissue diseases
Indeterminate appearance	Absence of typical COVID-19 features, with possible reverse halo sign or other organizing pneumonia characteristics	COVID-19 pneumonia may exhibit these features, but they are non-specific and can occur in various infectious and non-infectious conditions
Atypical appearance	Atypical CT findings for COVID-19 include unilateral, non-rounded, or non-peripheral ground-glass opacities, few ground-glass opacities, lobar consolidations, centrilobular nodules, and cavitation	Atypical features suggest alternative diagnoses to COVID-19 pneumonia
Negative findings	No CT features indicative of pneumonia are present	No CT features indicative of pneumonia are present

COVID-19, coronavirus disease 2019; CT, computed tomography.

from other respiratory conditions shown in *Figure 16*.

AI-based image analysis for COVID-19

Due to the rapid spread of the COVID-19 pandemic, there was a strain on medical resources in various regions. The utilization of AI for supporting the management of COVID-19 has become increasingly crucial. Manual diagnosis through CT scanning involves significant manual labor and time consumption (11). To alleviate the workload on radiologists, CAD tools have been developed based on DL or ML technologies *Figure 17* shows the main tasks of DL applied to CT images for COVID-19 detection. These tools have demonstrated the potential to enhance diagnostic efficiency and alleviate the pressure on radiologists (19).

Recent studies have highlighted that COVID-19 typically manifests as GGO or lesions in CT images. Therefore, the identification of abnormal areas like GGO or lesions in CT images plays a vital role in the diagnosis of COVID-19 by radiologists. The automated detection of GGO or nodules in CT images can assist in reducing human effort. For instance, Chen *et al.* (48) utilized U-Net++ to delineate abnormal lung areas in CT images. The model effectively segments areas with lesions in CT image slices and generates the bounding box of the segmented lesion. In their study, 2D CT image slices from 106 patients were utilized for training and internal validation. The model exhibited high accuracy in diagnosing patients, with a per-patient sensitivity of 100%, specificity of 93.55%, and accuracy of 95.24%. Additionally, a per-image sensitivity

of 94.34%, specificity of 99.16%, and accuracy of 98.85% were achieved. Notably, the prospective validation set further confirmed the model's performance, demonstrating comparable results to those of expert radiologists and significantly reducing radiologists' reading time by 65%.

Furthermore, in addition to detecting infectious areas, there is a growing interest in developing AI models capable of directly diagnosing COVID-19. Fang *et al.* (49) employed a radiomics analysis method for COVID-19 diagnosis, where radiomic features were extracted from manually delineated ROI. An unsupervised consensus clustering approach was used to select significant features associated with COVID-19, followed by the application of an SVM classifier for COVID-19 classification. The study achieved an AUC of 0.826 in the testing set. Wang *et al.* (50) designed a DL model to differentiate COVID-19 from typical viral pneumonia. The model, trained on annotated infectious areas as ROIs, leveraged a modified ResNet34 for feature extraction and a combination of decision tree and AdaBoost for classification. The model demonstrated an accuracy of 73.1% at the ROI-level in 99 patients. Additionally, Xu *et al.* (51) proposed a DL model for automatic infectious area detection in CT images, followed by the use of 3DResNet to identify COVID-19 presence. The model achieved an accuracy of 86.7% in the testing set.

Moreover, Jin *et al.* (52) developed an AI system for COVID-19 lesion segmentation and classification, involving lung region extraction, lesion segmentation, and lesion classification using a CNN-based classifier. The system, trained on a large dataset from multiple hospitals, exhibited

984

Ahmad et al. Deep learning for CT image classification
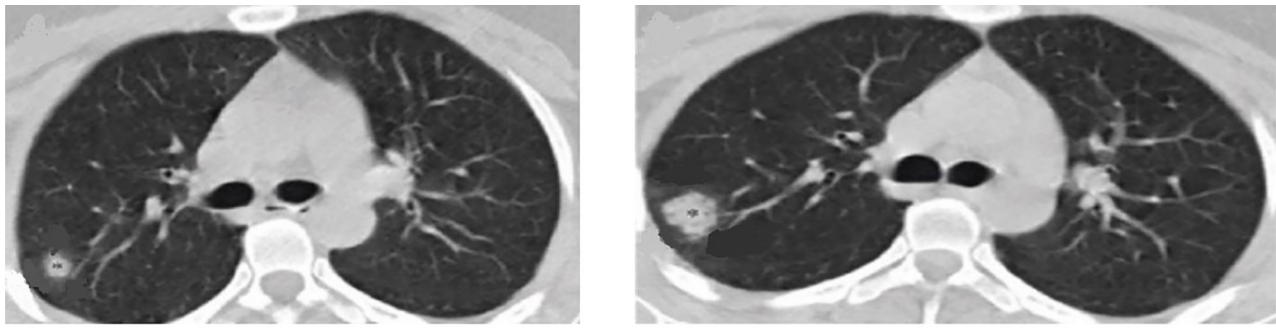
Bottom: sequential transverse thin-section CT scans of a 77-year-old male taken following the onset of symptoms

Day 5   Day 15   Day 20

Figure 16 CT images sourced illustrating CT characteristics associated with COVID-19, including ground glass opacities (bottom) and ground glass halo (top). CT, computed tomography; COVID-19, coronavirus disease 2019.
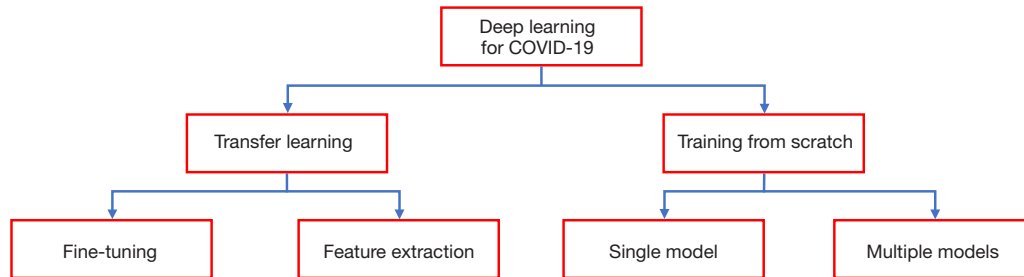

Figure 17 The main tasks of DL applied to CT images for COVID-19 detection. COVID-19, coronavirus disease 2019; DL, deep learning; CT, computed tomography.

high sensitivity and specificity in the testing set, meeting clinical application requirements. Additionally, Song *et al.* (53) focused on classifying the entire lung area, utilizing a details relation extraction neural network (DRE-Net) for feature extraction and patient-level diagnoses. The model achieved an AUC of 0.95 and a sensitivity of 0.96 in the testing set.

Various researchers have explored different approaches to detect COVID-19 using CT-scan images as presented in *Table 5*. Kaur *et al.* (54) combined classifiers with a pre-trained ResNet50 model, achieving 98.35% accuracy and

98.02% precision. Mishra *et al.* (55) used decision fusion to combine predictions from multiple DL models, with DenseNet121 performing best at 88.34% accuracy and an F1-score of 0.86. Li *et al.* (56) introduced COVNet, a novel convolutional model, attaining 90% sensitivity and 96% specificity. Wang *et al.* (57) modified InceptionNet using transfer learning, reporting 89.5% accuracy and 87.0% sensitivity. Gaur *et al.* (58) employed Extreme Learning Trees with DenseNet121 and transfer learning, achieving 85.5% accuracy and an F1-score of 0.85. Soares *et al.* (59) developed a DL model with 88.6% accuracy and 89.7%

precision. Goel *et al.* (60) proposed an optimized GAN, attaining 97.78% specificity and an F1-score of 0.98. Lu *et al.* (61) introduced CGENet, a graph theory-based approach, with 97.78% accuracy. Basu *et al.* (62) applied feature selection techniques, reporting 97.78% accuracy and 92.88% precision. These studies demonstrate the variety of approaches and DL architectures being explored to improve COVID-19 screening using CT-scan data, with different models excelling in various performance metrics.

Wu *et al.* (63) developed a DL-based coronavirus screening system that uses multi-view fusion. The system employs ResNet50, a variant of CNN, and was trained on 495 images from two Chinese hospitals, including 368 COVID-19 cases and 127 other pneumonia cases. The dataset was split into 80% training, 10% testing, and 10% validation sets, with images resized to 256×256. During testing, the system achieved 76% accuracy, 81.1% sensitivity, 61.5% specificity, and an AUC of 81.9%. The multi-view fusion model outperformed single-view models. In a separate study, Li *et al.* (56) introduced COVNet, an automated system for diagnosing coronavirus from CT images using ResNet50. Their dataset consisted of 4,536 chest CT samples, including 1,296 COVID-19, 1,735 community-acquired pneumonia, and 1,325 non-pneumonia cases. The dataset was split 90% for training and 10% for testing. COVNet demonstrated 90% sensitivity, 96% specificity, and a 96% AUC for identifying COVID-19 cases.

Yousefzadeh *et al.* (64) developed ai-corona, a DL framework for COVID-19 diagnosis using CT images. The system employs multiple CNN variants like DenseNet, ResNet, Xception, and EfficientNetB0. Using a dataset of 2,124 CT slices (1,418 non-COVID-19 and 706 COVID-19), split 80-20 for training and validation, the system achieved 96.4% accuracy, 92.4% sensitivity, 98.3% specificity, 95.3% F1-score, and 98.9% AUC. Jin *et al.* (52) created an AI-based coronavirus diagnostic system using ResNet152, a CNN variant with 152 layers. Their dataset included 1,881 cases (496 COVID-19 positive, 1,385 negative) from Chinese hospitals and public databases. The system demonstrated 94.98% accuracy, 94.06% sensitivity, 95.47% specificity, 91.53% precision, 92.78% F1-score, and 97.91% AUC. Xu *et al.* (51) designed a system to differentiate between healthy individuals, COVID-19 pneumonia, and influenza-A viral pneumonia cases using Resnet18. Their dataset comprised 618 CT images from Chinese hospitals (219 COVID-19, 224 influenza-A, 175 healthy). Using 85.4% for training and the rest for testing, the system achieved 86.7% accuracy, 81.5% sensitivity, 80.8% precision, and 81.1% F1-score.

Wang *et al.* (65) developed a COVID-19 medical screening system using DL techniques. The system utilized various pre-trained CNN models, including DPN-92, Inception-v3, ResNet-50, and Attention ResNet-50 with 3D U-Net++. Their dataset, sourced from five Chinese hospitals, contained 1,391 samples (850 COVID-19 cases, 541 negative cases). The dataset was randomly split for training and testing. The 3D U-Net++-ResNet-50 model performed best, achieving 97.4% sensitivity, 92.2% specificity, and 99.1% AUC. Javaheri *et al.* (66) introduced CovidCTNet, a DL approach for detecting coronavirus infection in CT images. The system used BCDU-Net architecture, a U-Net derivative, to distinguish COVID-19 from community-acquired pneumonia (CAP) and other lung conditions. Their extensive dataset comprised 89,145 CT images (32,230 COVID-19, 25,699 CAP, and 31,216 healthy or other disorders). Using a 90-10 split for training and testing, the system achieved 91.66% accuracy, 87.5% sensitivity, 94% specificity, and 95% AUC, demonstrating its effectiveness in COVID-19 detection.

Ardakani *et al.* (67) conducted a study on COVID-19 detection using various CNN techniques in CT images. They evaluated ten popular CNN variants, including AlexNet, VGG-16, VGG-19, SqueezeNet, GoogleNet, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and Xception. Their dataset consisted of 1,020 CT samples from both COVID-19 and non-COVID-19 cases, with an 80-20 split for training and validation. ResNet-101 and Xception emerged as top performers. ResNet-101 achieved 99.51% accuracy, 100% sensitivity, 99.4% AUC, and 99.02% specificity. Xception demonstrated 99.02% accuracy, 98.04% sensitivity, 87.3% AUC, and 100% specificity. Chen *et al.* (48) introduced a DL approach using the pre-trained U-Net++ model for COVID-19 detection in high-resolution CT images. U-Net++ was initially used to extract valid regions within the CT images. Their dataset included 46,096 images from a hospital (51 COVID-19 cases, 55 other diseases). After filtering, 35,355 images were selected and split into training and testing sets. The system achieved 94.34% sensitivity, 99.16% specificity, 98.85% accuracy, 88.37% precision, and 99.61% negative predictive value (NPV).

Cifci *et al.* (68) developed an early coronavirus detection method using deep transfer learning with AlexNet and Inception-V4 models. Using a dataset of 5,800 CT images (80% training, 20% testing), AlexNet outperformed Inception-V4 with 94.74% accuracy, 87.37% sensitivity,

and 87.45% specificity. Elghamrawy and Hassanien (69) combined CNN with the Whale Optimization Algorithm (WOA) for COVID-19 diagnosis, achieving 96.40% accuracy, sensitivity, and precision. He *et al.* (70) introduced CRNet for COVID-19 detection in CT images, attaining 86% accuracy and 94% AUC. Wang *et al.* (50) presented modified-Inception for COVID-19 diagnosis, achieving 79.3% accuracy and 83% sensitivity. Liu *et al.* (71) designed an automated system using modified DenseNet-264, reaching 94.3% accuracy and 98.6% AUC. Song *et al.* (53) introduced DeepPneumonia for COVID-19 diagnosis, achieving 94% accuracy and 99% AUC. Zheng *et al.* (72) proposed DeCoVNet, a 3D DCNN using U-Net architecture, achieving 90.1% accuracy with 630 CT samples. Hasan *et al.* (73) developed a hybrid system using Q-deformed entropy and DL features, achieving 99.68% accuracy in differentiating COVID-19 from pneumonia and healthy cases. Amyar *et al.* (74) used a DL method for COVID-19 diagnosis, achieving 86% accuracy with a dataset of 1,044 cases.

Table 6 provides a comprehensive overview of the DL-based COVID-19 diagnosis systems utilizing CT samples with pre-trained models and deep transfer learning. The table outlines key factors including data sources, image quantities and classes, data partitioning methods, diagnostic techniques employed, and the performance metrics achieved by these systems. In the analysis of the results, it is evident that various DL models and techniques have been explored for COVID-19 screening using CT imaging. The high accuracies and specificities achieved by these models indicate their potential for accurate and reliable diagnosis of COVID-19 (11). The utilization of transfer learning, decision fusion, and feature selection techniques has contributed to the performance improvements of the models. However, further studies and validations are necessary to assess the generalizability and scalability of these models across diverse patient populations and imaging conditions. Additionally, the integration of these models into clinical practice and their performance in real-world scenarios warrant further investigation to ensure their efficacy and impact on patient care.

In conclusion, the application of AI in CT image analysis has shown promising results in the rapid and accurate diagnosis of COVID-19. By leveraging AI technologies, clinicians can efficiently diagnose COVID-19, predict disease severity, and tailor treatment strategies (11), ultimately improving patient outcomes. Additionally, the integration of AI with CT imaging offers a cost-effective and efficient approach to diagnosing COVID-19 while ensuring the safety of healthcare professionals and patients. To perform COVID-19 detection various methods have been proposed recently. However, not all the methods are reliable and efficient enough to deploy in real-time applications. Some of the deep methods have resulted in promising performances in terms of accuracy, MCC, ROC, sensitivity, and so on. The performances that are high amongst the DL methods are ensemble CNN and DL-based segmentation algorithms. The most popular CNN methods resulting in higher performance include a combination of two or more conventional CNN models. We can conclude that ensemble methods provide better results as they combine the advantages of two or more networks to learn input features in a better way.

Structure of a CAD system

Over the past decades, numerous studies have been conducted to enhance the efficiency of lung cancer diagnosis using CAD systems. As shown in *Figure 18*, the complete workflow of a CAD system utilizes thin cross-section images to supplement radiologists in identifying pulmonary nodules (110). These systems generally include three main components: preprocessing, nodule detection (encompassing candidate nodule detection and false positive reduction), and nodule classification. Preprocessing is a crucial initial step that aims to improve image quality by reducing unwanted distortions or enhancing necessary features for further processing, which is vital for achieving higher accuracy in models. For instance, applying contrast limited adaptive histogram equalization (CLAHE) to CT images enhances visual characteristics, facilitating better model interpretation, as shown in *Figure 19* below. Removing distracting elements like chest tissues and artifacts, and enhancing relevant information, particularly within the lung volume (ROI), can prevent 5–17% of nodule detection misses. The performance of CAD systems varies significantly due to differences in CT inputs, nodule characteristics, and particularly the diversity of algorithms used. Most studies focus on improving both sensitivity and specificity by reducing false positives and enhancing nodule classification using the same datasets (17), as illustrated in *Figure 20*. Common segmentation methods based on Hounsfield unit (HU) contrast between the lung and surrounding tissue are categorized into rule-based and data-based approaches, combining techniques like thresholding, component analysis, region growing, morphological

Table 6 Analysis of COVID-19 using CT images

Author	Year	Method	Performance
Chen *et al.* (48)	2020	U-Net++	ACC =0.95
Fang *et al.* (49)	2020	Randiomic feature, consensus clustering	AUC =0.826
Wang *et al.* (50)	2020	ResNet34, decision tree	AUC =0.78
Xu *et al.* (51)	2020	3D-CNN, 3DResNet	ACC =0.86
Alaiad *et al.* (75)	2023	DL	ACC =0.995
Song *et al.* (53)	2021	OpenCV, DRE-Net	AUC =0.95
Ullah *et al.* (76)	2023	CNN	ACC =99.8%
Zheng *et al.* (72)	2020	U-Net, 3DResNet	AUC =0.98
Silva *et al.* (77)	2020	DL	ACC =0.8768
Shi *et al.* (78)	2020	VBNet, Hand-crafted feature, random forest	ACC =0.88
Wang *et al.* (57)	2020	FPN, DenseNet	AUC =0.87, AUC =0.88
Shi *et al.* (79)	2020	V-Net, LASSO, logistic regression	AUC =0.89
Kaur *et al.* (54)	2022	Classifier fusion with ResNet50	ACC =98.35%; PRC =98.02%
Mishra *et al.* (55)	2020	Decision fusion with DenseNet121	ACC =88.34%; F1-Score =0.86
Li *et al.* (56)	2020	COVNet	SEN =90.0%; SPE =96.0%
Afif *et al.* (80)	2023	DL	ACC =96.23%
Gaur *et al.* (58)	2022	EWT with DenseNet121	ACC =85.5%; F1-Score =0.85
Soares *et al.* (59)	2020	DLM	ACC =88.6%; PRC =89.7%
Goel *et al.* (60)	2021	Optimized GAN	SPE =97.78%; F1-Score =0.98
Lu *et al.* (61)	2022	CGENet	ACC =97.78%
Basu *et al.* (62)	2023	Feature selection technique	ACC =97.78%; PRC =92.88%
Gupta *et al.* (11)	2023	DarkNet19 with repeated holdout 10FCV	ACC =98.91%; SEN =98.96%; SPE =98.86%; PRC =98.88%; F1-Score =0.99
Khan *et al.* (81)	2023	CNN-based STM	ACC =98.01%
Sharma *et al.* (82)	2020	ML	ACC =91.0%
Kathamuthu *et al.* (83)	2022	CNN	ACC =98.0%
Motwani *et al.* (84)	2023	CNN	ACC =93.78%
Gozes *et al.* (85)	2020	2D DCNN-based ResNet-50	AUC =99.6%; SEN =92.2%
Shan *et al.* (86)	2020	VB-Net	DICE =91.6%
Jin *et al.* (52)	2020	2D CNN	ACC =94.98% AUC =97.91%
Sahoo *et al.* (87)	2022	ViT	ACC =98.39%; F1-Score =98.49%
Panwar *et al.* (88)	2020	VGG19	95%
Dosovitskiy *et al.* (89)	2020	DenseNet201	ACC =96.25%
Devlin *et al.* (90)	2018	DenseNet101	ACC =97.4%
Tekade *et al.* (91)	2018	CNN	ACC =95.66%

Table 6 *(continued)*

Table 6 *(continued)*

Author	Year	Method	Performance
Wang *et al.* (92)	2020	CNN	PRE =87.87%
Chen *et al.* (93)	2021	CNN	PRE =80.10%
Humayun *et al.* (94)	2022	CNN	ACC =89.68%
Al-Yasriy *et al.* (95)	2020	CNN	FM =96.40%
Al-Huseiny *et al.* (96)	2021	CNN	ACC =94.38%
Raza *et al.* (97)	2023	CNN	ACC =99.1%
Mohammed *et al.* (98)	2023	CNN	ACC =73.8%
Gulsoy *et al.* (99)	2023	SwinT	ACC =99.69%
Sahin *et al.* (100)	2023	R-CNN	ACC =93.86%
Gunraj *et al.* (101)	2020	COVIDNer-CT	ACC =94.9%
Shi *et al.* (102)	2021	Teacher-student attention	96.4%
Yu *et al.* (103)	2021	ResGNet	93.9%
Mondal *et al.* (104)	2021	XViTCOS-CT	ACC =98.1%
Harmon *et al.* (105)	2020	DenseNet-121 and AH-Net segmentation	ACC =90.8%
Ouyang *et al.* (106)	2020	Dual sampling, attention network with ResNet-34	ACC =87.5%
Wu *et al.* (63)	2020	Multiview fusion model using ResNet-50	ACC =81.1%
Ardakani *et al.* (67)	2020	ResNet-101	ACC =99.51%
Sun *et al.* (107)	2020	Adaptive feature selection-guided deep forest—SVM	ACC =91.79%
Wang *et al.* (108)	2020	Prior-attention residual model 3D ResNets	ACC =93.3%
Hasan *et al.* (73)	2020	LSTM using Qdeformed entropy and deep features	ACC =99.8%
Butt *et al.* (109)	2020	3D ResNets with location attention mechanism	ACC =86.7%

COVID-19, coronavirus disease 2019; CNN, convolutional neural network; LSTM, long short-term memory; SVM, support vector machine; GAN, generative adversarial network; DLM, deep learning model; 3D-CNN, three-dimensional convolutional neural network; ResNet50, residual network with 50 layers; U-Net, U-shaped network; VBNet, V-Net with Bottleneck; FPN, feature pyramid network; DenseNet, Densely Connected Convolutional Network; CGENet, Conditional Graph Ensemble Network; 2D DCNN, two-dimensional deep convolutional neural network; ViT, vision transformer; VGG19, Visual Geometry Group Network with 19 layers; SwinT, Swin Transformer; XViTCOS-CT, Crossover Vision Transformer with Conditional Optimization Strategy for Computed Tomography; DRENet, Dense Residual Ensemble Network.

operations, and filtering for effective preprocessing. The selection of preprocessing methods, such as noise reduction filters or histogram equalization, depends on specific research objectives and CT image characteristics. The list of preprocessing methods used for analyzing CT images is shown in *Table 7*. Nodule detection typically involves two stages: candidate detection and false positive reduction. While the former aims for high sensitivity by identifying potential nodules, the latter focuses on distinguishing true nodules from false positives using advanced algorithms like CNNs and SVMs. The final stage, nodule classification, determines the probability of malignancy by extracting features such as shape, texture, and intensity from nodule images and applying classifiers like CNNs, SVMs, and random forests. Advanced DL techniques have shown promising results in improving nodule classification accuracy and robustness, leading to better clinical decision-making and treatment planning (131).

Candidate nodule detection, false positive reduction, and classification

Before delving into the specific studies and methodologies, it's important to note that in the field of lung nodule
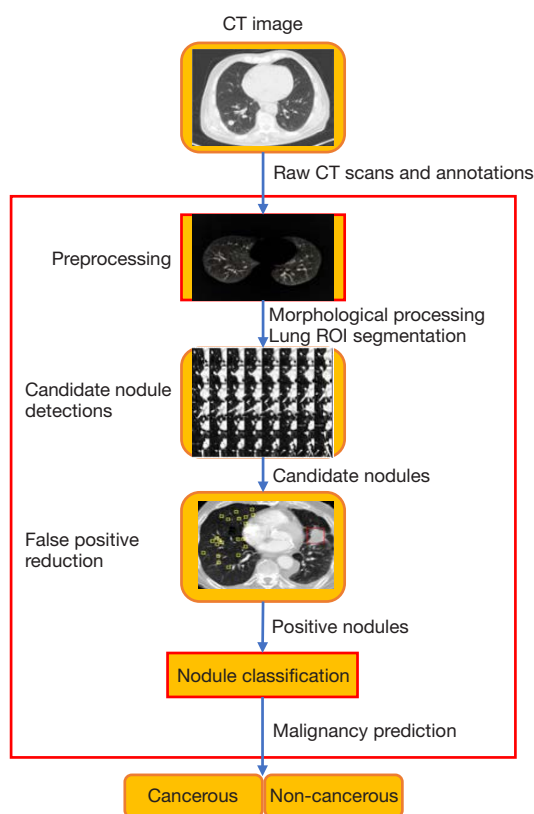
Figure 18 Complete workflow of a CAD system. CT, computed tomography; ROI, region of interest; CAD, computer-aided diagnosis.

detection and classification, the methodologies of both ML and DL is shown in *Figure 20* below have seen significant advancements in recent years, driven largely by the rapid progress in DL and computer vision techniques. This section provides an overview of key research efforts that have shaped our current understanding and capabilities in this critical area of medical image analysis (124). The studies presented here represent a diverse range of approaches, from traditional ML methods to state-of-the-art DL architectures, each contributing unique insights and innovations to the field as shown in *Table 8* the related worked for nodule classification. While not exhaustive, this selection of related work highlights the evolution of techniques, the challenges addressed, and the progressive improvements in accuracy and efficiency in nodule detection and classification. As we review these studies, we'll see a clear trend towards more sophisticated, multi-stage approaches that aim to mimic and augment the diagnostic process of expert radiologists.

Tajbakhsh *et al.* [2019] (180) used novel vessel-oriented image representation (VOIR) that can improve the machine perception of pulmonary embolism (PE) through a consistent, compact, and discriminative image in pulmonary nodules. Helm *et al.* [2009] (181) found that radiologists achieved an 80% sensitivity rate for detecting lung nodules measuring 4 mm or larger, averaging 0.9 false positives per study. However, detection performance significantly declined for nodules smaller than 4 mm. In a related study by Armato *et al.* [2007] (182), four radiologists independently reviewed 30 CT images in a two-phase annotation trial, revealing agreement on nodules larger than 3 mm but substantial variability in annotations. Smaller nodules exhibited similar challenges, making precise detection of these nodules arduous for radiologists. This difficulty underscores the importance of CAD systems to improve the identification of small lung nodules and streamline the annotation process for creating reliable ground truth data for training. To enhance manual annotation accuracy, multiple experienced radiologists can review images collectively, particularly for contentious cases. However, discrepancies between annotating groups persist, suggesting that variability may limit the effectiveness of computer-assisted detection. For classification tasks, integrating additional diagnostic information, such as pathological results, can help CAD systems extract features that are challenging for human eyes to identify, thereby improving classification performance. Enhanced visualization techniques can also aid in distinguishing nodules from small vessels. Martini *et al.* [2020] (183) demonstrated that vessel suppression notably improved nodule detection rates and inter-reader agreement while reducing reading time. Agam *et al.* [2005] (184) applied correlation-based enhancement filters and fuzzy shape representation to reduce false positives by accurately depicting vessel trees. Variability in manual delineation can affect nodule detection and size estimation. Meyer *et al.* [2006] (185) highlighted this issue, showing significant variability in measurements of 23 lung nodules among six radiologists, which was more pronounced than discrepancies associated with drawing tools. To address these challenges, the Lung Imaging Database Consortium (LIDC) was established to provide multiple annotations for validation. Ross *et al.* [2007] (186) created three-dimensional models of nodule segmentations from various radiologists, noting that while discrepancies in bounding boxes could be tolerated, nodule measurements were often derived from the separately detectable long and short axes. Interestingly, variability
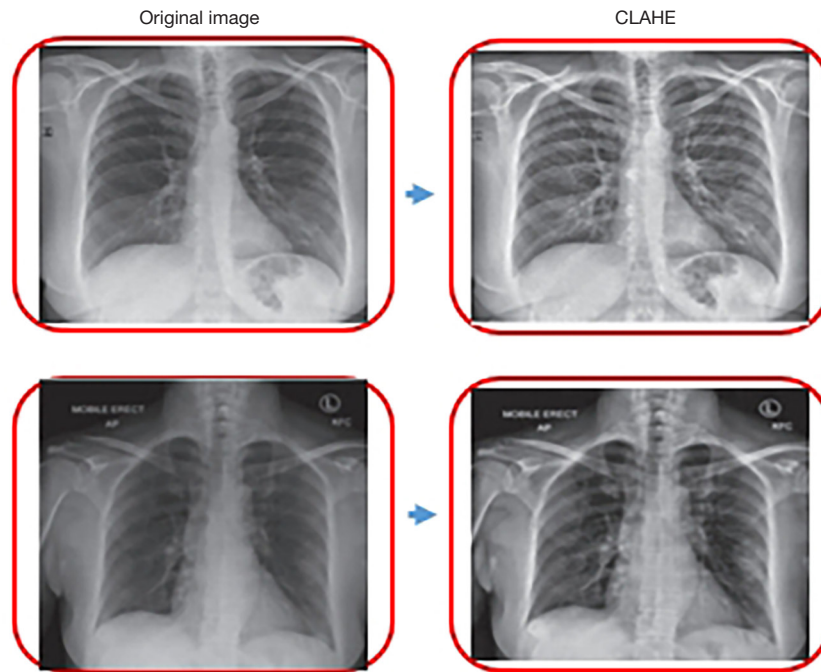
Figure 19 Original CT image compared to CT image processed with CLAHE (111). CT, computed tomography; CLAHE, contrast limited adaptive histogram equalization.
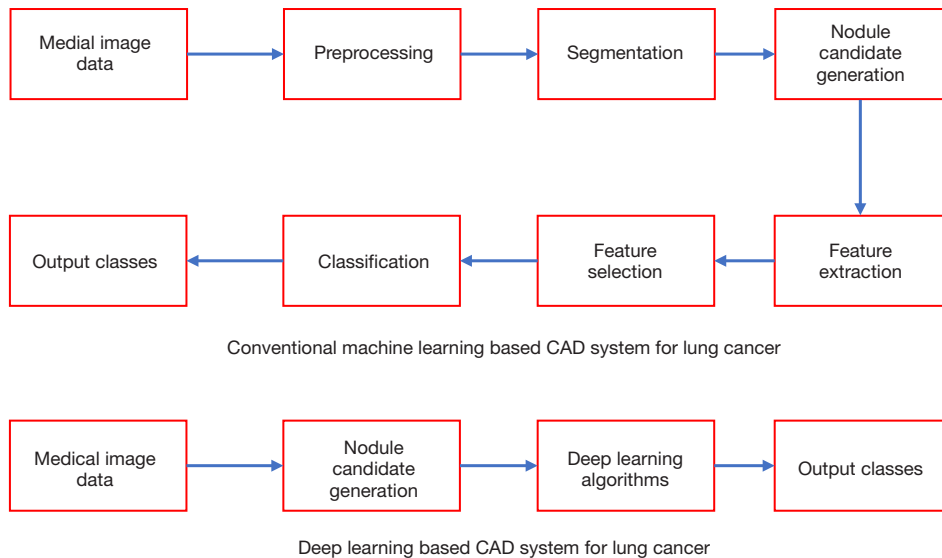
Figure 20 DL-based CAD system for lung cancer. CAD, computer-aided diagnosis; DL, deep learning.

in diagnoses and patient management recommendations based on CT images appears lower than that for detection and measurement tasks. Nair *et al.* [2018] (187) assessed evaluations of 69 nodules by 107 radiologists across 25 countries and found satisfactory overall agreement for

nodule composition (Fleiss' kappa =0.65) and management strategies (kappa =0.63–0.73). However, the agreement on morphological variables and diameter measurements was relatively low, indicating that while guideline-based management and nodule composition analysis showed good

Table 7 List of preprocessing methods

Reference	Methods	Uses
Veldhuizen and Jernigan (112)	Wiener filter	Produces an estimate of a desired or target random process
Tian *et al.* (113)	Binarization	Transforms data features of any entity into vectors of binary numbers
Yadav *et al.* (111)	CLAHE	Works on small regions called tiles
Prabha and Kumar (114)	Smoothing filter	Utilized in blurring regions
Kociołek *et al.* (115)	Normalization	Scaling pixel values to a standard range for consistency
Lehmann *et al.* (116)	Interpolation	Best estimation of a pixel's color and intensity in context to the values at neighboring pixels
Pizer *et al.* (117)	Adaptive histogram equalization	Improving image contrast for better analysis
Gungor (118)	Wavelet transform	Decomposes special patterns hidden in the mass of data
Zhu *et al.* (119)	Image rescaling	Adjusting image size for consistency
Luo *et al.* (120)	Noise reduction filters	Enhancing image quality by reducing noise
Jung *et al.* (121)	ROI selection	Focusing analysis on specific areas of interest
Zhang *et al.* (122)	Image registration	Aligning images for comparison and analysis
He *et al.* (123)	Segmentation techniques	Identifying and separating different structures in the image
Ahmad *et al.* (5)	Data augmentation	Increasing dataset size for training DL models
Li *et al.* (124)	Artifact removal	Eliminating unwanted artifacts in the images
John and Mini (125)	Median filter	The median filter is a non-linear digital filtering technique, often used to remove noise from an image or signal
Ayshath Thabsheera *et al.* (126)	Guided filtering	The guided filtering technique is a method that performs image smoothing by using the content of a second image and also preserves edges in the image
Javaid *et al.* (127)	Histogram equalization	The histogram equalization is an image processing that enhances the contrast of the image
Elavarasu *et al.* (128)	Mean filter	Mean filtering is a technique where the intensity deviation of one pixel and its successor pixel is decreased using arithmetic mean
Vignesh *et al.* (129)	Gaussian filter	Gaussian filter is a type of linear smoothing filter, in which the weights of the filter are chosen based on the Gaussian function
Fortin *et al.* (130)	Laplacian of Gaussian	Laplacian of Gaussian filter is used for detecting edges in the image and also removing noise by smoothening the image using the Gaussian filter

ROI, region of interest; CLAHE, contrast limited adaptive histogram equalization.

consensus, detailed measurements lacked the same level of agreement.

The integration of advanced ML and DP techniques, particularly CNNs, has significantly improved the performance and accuracy of these systems. Methods such as multi-stream frameworks, transfer learning, and hybrid models have shown promising results, addressing challenges in sensitivity, specificity, and robustness. The challenges of building DL models from CT images, it is crucial to acknowledge the significant impact of image complexity and the scarcity of standardized datasets. One of the primary obstacles is the diversity of CT scan protocols, which vary widely depending on the equipment, settings, and clinical purpose. This variability can introduce inconsistencies in the images, making it challenging to create a robust DL model. Re-evaluating

Table 8 Related worked for nodule classification

Author	Year	Feature/method	Performance
Ozdemir *et al.* (132)	2020	3D segmentation network based on V-net architecture	ACC =0.921
Li *et al.* (133)	2019	3D CNN	ACC =0.912
Masood *et al.* (134)	2020	VGG-16, 3D-CNN	ACC =0.946 LUNA16
			SEN =0.976 LIDC-IDRI
			SEN =0.988 ANODE09
El-Regaily *et al.* (135)	2019	2D CNN	SEN =0.853
Wang *et al.* (136)	2019	CNN	ACC =0.903
Zheng *et al.* (137)	2020	2D U-net, 3D-CNN	ACC =0.955
Ali *et al.* (138)	2020	CNN	ACC =0.9669
Zuo *et al.* (139)	2020	3D-CNN	ACC =0.83
Zuo *et al.* (140)	2019	2D CNN	ACC =0.762
Zheng *et al.* (141)	2020	2D U-Net, VGG-net	SEN =0.942
Liu *et al.* (142)	2019	ResNET-18	ACC =0.957 LUNA16
Wang *et al.* (143)	2020	SSL	ACC =0.907
Zhai *et al.* (144)	2020	SF2T	ACC =0.9730
Liu *et al.* (145)	2020	3D ResU-Net, 3D Dense U-Nets	ACC =0.879
Shi *et al.* (146)	2019	VGG-16, SVM	SEN =0.872
Zhou *et al.* (147)	2019	Encoder-decoder	ACC =0.971
			AUC =0.982
Riquelme *et al.* (148)	2020	CAD	ACC =0.996
Ren *et al.* (149)	2020	MRC-DNN	SEN =0.90
Al-Shabi *et al.* (150)	2019	Gated-Dilated network	AUC =0.9514
Liu *et al.* (151)	2020	CNN	AUC =0.9797
Xie *et al.* (152)	2019	3D DNN with ResNet-50	AUC =0.957
Apostolopoulos *et al.* (153)	2020	Dual deep solitary pulmonary nodules network	ACC =0.93
Nasrullah *et al.* (154)	2019	CMixNet and gradient boosting machine	SEN =0.94
Harsono *et al.* (155)	2020	3D ConvNet with DPN	AUC =0.818
Liao *et al.* (156)	2019	3D-DNN	ACC =0.859, 0.814
Yang *et al.* (157)	2020	3D DensNet	AUC =0.932
Balagurunathan *et al.* (158)	2019	Optimal linear classifier	AUC =0.85
Hussein *et al.* (159)	2019	3D CNN	ACC =0.786
Al-Shabi *et al.* (160)	2019	Deep local-global network	ACC =0.885
Ardila *et al.* (161)	2019	3D Inception blocks	ACC =0.944
Gao *et al.* (162)	2019	DLSTM using TEM	AUC =0.8905
Chen *et al.* (163)	2019	ResNets,	ACC =0.919

Table 8 (*continued*)

Table 8 *(continued)*

Author	Year	Feature/method	Performance
Zhang *et al.* (164)	2024	S-Ne	ACC =0.914
Yang *et al.* (165)	2019	RCNN, U-Net	ACC =88.5%
Wu *et al.* (166)	2017	Random forest	ACC =82.1%
Li *et al.* (167)	2018	AE CNN	ACC =80.3%
Shen *et al.* (168)	2017	CNN	ACC =88.5%
Guo *et al.* (169)	2016	Threshold	SEN =100%
Huang *et al.* (170)	2018	Threshold	SEN =100%
Cui *et al.* (171)	2019	Morphology	ACC =89.89%
Song *et al.* (172)	2015	DL	SES =70%
Shin *et al.* (173)	2016	DL	SEN =70%
Onishi *et al.* (174)	2020	DCNN and GAN	SPEC =0.778; SEN =0.939
Gomes Ataide *et al.* (175)	2020	ML	ACC =0.993
Ma *et al.* (176)	2020	ML	ACC =80%
Kanjanasurat *et al.* (177)	2023	DenseNet, VGG19, ResNet52,	ACC =93.37%
Celik *et al.* (178)	2023	COVIDDWNet + GB	ACC =99.84%
de Jesus Silva *et al.* (179)	2023	EnsenbleDVX	ACC =97.7%

3D-CNN, three-dimensional convolutional neural network; ACC, accuracy; VGG-16, Visual Geometry Group Network with 16 layers; SEN, sensitivity; SSL, self-supervised learning; SVM, support vector machine; CNN, convolutional neural network; 2D DCNN, two-dimensional convolutional neural network; R-CNN, region-based convolutional neural network; ResNet-50, residual network with 50 layers; DL, deep learning; SES, socioeconomic status; ML, machine learning; DLSTM, deep long short-term memory; COVIDDWNet, COVID Data Warehouse Network; GB, gradient boosting; EnsembleDVX, ensemble deep learning for vision; QuCNet, Quantum-Inspired Convolutional Neural Networks for Optimized Thyroid Nodule Classification; SF2T, leveraging swin transformer and two-stream networks for lung nodule detection; Nodule-CLIP, lung nodule classification based on multi-modal contrastive learning; AttentNet, Fully Convolutional 3D Attention for Lung Nodule Detection; AUC, area under the curve.

and standardizing these protocols is essential to ensure the development of high-quality datasets that can lead to more accurate and generalizable models. Another critical aspect is the verification of ground truth. While expert verification, particularly by radiologists, is indispensable, it should be complemented by more objective methods, such as pathological or surgical findings. This dual approach ensures that the ground truths used in model training are both accurate and reliable, enhancing the model's performance in real-world applications. Additionally, recent advancements in image improvement solutions, such as GE True Fidelity, offer promising avenues to mitigate some of these challenges. These technologies can enhance the quality of CT images, potentially making it easier to establish consistent datasets and build more effective DL models. Incorporating such solutions into

the dataset creation process could significantly improve the robustness and reliability of the models, ultimately leading to better clinical outcomes. Future research should continue to focus on enhancing the interpretability of these models and improving their continuous learning capabilities to adapt to real-time clinical environments.

The detection of pulmonary nodules is a fundamental and crucial task in medical imaging analysis. Its significance lies in its role as a precursor to more advanced procedures such as nodule classification (131). The accuracy and efficiency of these subsequent tasks are heavily dependent on the quality of the initial detection. Consequently, it is imperative to investigate the various factors that impact the generalization capabilities and robustness of pulmonary nodule detection systems. One key factor known to influence CT image characteristics is the choice

of reconstruction kernel. However, it is important to acknowledge certain limitations in the current research. Primarily, the study is constrained by the size and diversity of the available dataset (6). Additionally, an intriguing finding emerges from the application of image conversion techniques: the nodule detection system demonstrates enhanced performance on sharp kernel images compared to the baseline performance on smooth kernel images. This observation not only highlights the potential for improving detection accuracy through image processing methods but also underscores the complex relationship between image reconstruction parameters and detection performance. It suggests that further exploration of image conversion techniques and their impact on various reconstruction kernels could yield valuable insights for optimizing pulmonary nodule detection systems (188).

Evaluation metrics

There are several evaluation metrics commonly used in the classification of CT images to assess the performance of DL models.

Accuracy

Accuracy is a fundamental evaluation metric used in the context of CT image analysis to assess the performance of ML models in correctly classifying instances. Accuracy plays a crucial role in determining the reliability and effectiveness of diagnostic systems, which is calculated as the ratio of correctly classified instances (true positives and true negatives) to the total number of instances (true positives, true negatives, false positives, and false negatives) (189). The formula for accuracy is:

$$Accuarcy = \frac{TP+TN}{TP+TN+FP+FN} \qquad [1]$$

Where: *TP* is a true positive, *FN* is a false negative, *TN* is a true negative, *FR* is a false positive.

The accuracy reflects the overall correctness of the model's predictions in identifying specific features, abnormalities, or diseases within the CT images. A high accuracy score indicates that the model is effectively distinguishing between different classes or categories within the CT images, leading to more reliable and accurate diagnostic outcomes (190). However, it is essential to consider the limitations of accuracy as an evaluation metric in CT image analysis. Accuracy alone may not provide a complete picture of the

model's performance, especially in cases where the dataset is imbalanced or when certain classes are more prevalent than others. In such situations, accuracy may not accurately represent the model's ability to correctly identify rare or critical instances within the CT images.

Therefore, while accuracy is a valuable metric in evaluating the overall performance of ML models in CT image analysis, it is often recommended to complement it with other evaluation metrics such as precision, recall, F1-score, and area under the ROC curve to gain a more comprehensive understanding of the model's performance and effectiveness in diagnosing and analyzing CT images accurately.

Precision

Precision is a critical evaluation metric used in the analysis of CT images to assess the performance of ML models in correctly identifying positive instances, precision plays a crucial role in determining the accuracy and reliability of diagnostic systems (189). Precision is calculated as the ratio of true positive instances to the total number of instances predicted as positive by the model. The formula for precision is:

$$Precision = \frac{TP}{TP+FP} \qquad [2]$$

The precision reflects the model's ability to accurately identify and classify positive instances, such as detecting specific abnormalities, lesions, or diseases within the CT images. A high precision score indicates that the model has a low rate of falsely classifying negative instances as positive, leading to more reliable and accurate diagnostic results. However, it is important to consider the limitations of precision as an evaluation metric in CT image analysis. Precision focuses solely on the positive predictions made by the model and does not take into account the instances that were incorrectly classified as negative (false negatives). In scenarios where false negatives are equally important, precision alone may not provide a complete assessment of the model's performance.

Therefore, while precision is a valuable metric in evaluating the model's ability to make accurate positive predictions in CT image analysis, it is recommended to consider it in conjunction with other evaluation metrics such as recall, F1-score, and specificity to obtain a comprehensive understanding of the model's performance and effectiveness in correctly identifying and classifying positive instances within CT images.

Recall (sensitivity)

Recall, also known as sensitivity, is a crucial evaluation metric utilized in the analysis of CT images to evaluate the performance of ML models in correctly identifying positive instances. In the domain of medical imaging, recall plays a significant role in assessing the model's ability to detect relevant features, abnormalities, or diseases within the CT images accurately (189).

Recall is calculated as the ratio of true positive instances to the total number of actual positive instances in the dataset. The formula for the recall is:

$$Recall = \frac{TP}{TP + FN} \qquad [3]$$

The recall reflects the model's capability to correctly identify all actual positive instances within the CT images. A high recall score indicates that the model has a low rate of missing positive instances, leading to more comprehensive and accurate diagnostic results. However, it is essential to consider the limitations of recall as an evaluation metric in CT image analysis. Recall focuses solely on the model's ability to identify positive instances and does not account for instances that were incorrectly classified as positive (false positives). In scenarios where false positives are equally critical, recall alone may not provide a complete assessment of the model's performance. Therefore, while recall is a valuable metric in evaluating the model's ability to capture all positive instances within CT images, it is recommended to complement it with other evaluation metrics such as precision, F1-score, and specificity to obtain a holistic understanding of the model's performance and effectiveness in correctly identifying and capturing positive instances within CT images.

Specificity

Specificity is a vital evaluation metric used in the analysis of CT images to assess the performance of ML models in correctly identifying negative instances. Specificity plays a significant role in determining the model's ability to accurately classify negative instances and minimize false alarms.

Specificity is calculated as the ratio of true negative instances to the total number of actual negative instances in the dataset. The formula for specificity is:

$$Specificity = \frac{TN}{TN + FP} \qquad [4]$$

Specificity reflects the model's capability to correctly identify all actual negative instances within the CT images. A high specificity score indicates that the model has a low rate of falsely classifying positive instances as negative, leading to more reliable and accurate diagnostic outcomes. However, it is important to consider the limitations of specificity as an evaluation metric in CT image analysis. Specificity focuses solely on the model's ability to identify negative instances and does not consider instances that were incorrectly classified as negative (false negatives). In scenarios where false negatives are equally critical, specificity alone may not provide a comprehensive assessment of the model's performance.

Therefore, while specificity is a valuable metric in evaluating the model's ability to correctly identify negative instances within CT images, it is recommended to consider it in conjunction with other evaluation metrics such as sensitivity, precision, F1-score, and accuracy to obtain a comprehensive understanding of the model's performance and effectiveness in accurately classifying both positive and negative instances within CT images.

F1-score

The F1-score is a critical evaluation metric used in the analysis of CT images to provide a balanced measure of a ML model's performance in terms of both precision and recall particularly in CT imaging, the F1-score is a valuable metric for assessing the overall effectiveness and accuracy of the model in correctly identifying positive instances while minimizing false positives and false negatives (189).

The F1-score is calculated as the harmonic mean of precision and recall, providing a balanced measure that considers both metrics. The formula for the F1-score is:

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \qquad [5]$$

Where: precision is the ratio of true positive instances to the total number of instances predicted as positive by the model.

Area under the receiver operating characteristic curve (AUC-ROC)

AUC-ROC is a key evaluation metric used in the analysis of CT images to assess the performance of ML models in binary classification tasks. The AUC-ROC metric provides a comprehensive measure of the model's ability to

996

Ahmad et al. Deep learning for CT image classification

distinguish between positive and negative instances across various classification thresholds. The AUC-ROC quantifies the performance of a classification model by plotting the true positive rate (TPR) (sensitivity) against the false positive rate (FPR) (1 – specificity) at different classification thresholds. The AUC-ROC score represents the area under this curve, ranging from 0 to 1, where a score closer to 1 indicates a better-performing model.

The formula for calculating the AUC-ROC score involves integrating the ROC curve, which is a plot of sensitivity against 1 – specificity at different threshold values. The AUC-ROC score provides a single value that summarizes the model's performance across all possible classification thresholds, offering insights into the model's ability to correctly classify positive and negative instances within the CT images.

$$AUC\text{-}ROC = \int TPR(FPR)\, dFPR \tag{6}$$

A high AUC-ROC score indicates that the model effectively distinguishes between positive and negative instances, leading to accurate and reliable diagnostic outcomes. The AUC-ROC metric is particularly useful for evaluating the overall performance of a classification model in CT image analysis, especially in scenarios where the dataset is imbalanced or when different classification thresholds need to be considered. Therefore, the AUC-ROC is a valuable evaluation metric in CT image analysis, providing a comprehensive measure of the model's ability to discriminate between positive and negative instances. It is recommended to use the AUC-ROC score in conjunction with other evaluation metrics to gain a complete understanding of the model's performance and effectiveness in binary classification tasks within the context of CT imaging.

Area under the precision-recall curve (AUC-PR)

AUC-PR is a crucial evaluation metric utilized in the analysis of CT images to assess the performance of ML models in binary classification tasks. The AUC-PR metric provides a comprehensive measure of the model's precision-recall trade-off and its ability to correctly identify positive instances while minimizing false positives. The AUC-PR quantifies the model's performance by plotting precision against recall at different classification thresholds. The AUC-PR score represents the area under this curve, ranging from 0 to 1, where a score closer to 1 indicates a better-performing model.

The formula for calculating the AUC-PR score involves integrating the precision-recall curve, which is a plot of precision against recall at different threshold values. The AUC-PR score offers a single value that summarizes the model's precision-recall trade-off, providing insights into the model's ability to make accurate positive predictions while capturing all actual positive instances within the CT images.

$$AUC\text{-}PR = \int P(R)\, dR \tag{7}$$

Where P is the precision, and R is the recall.

A high AUC-PR score indicates that the model strikes a balance between precision and recall, leading to accurate and reliable diagnostic outcomes. The AUC-PR metric is particularly useful for evaluating the model's performance in scenarios where precision and recall are both critical factors in the classification task. Therefore, the AUC-PR is a valuable evaluation metric in CT image analysis, offering a comprehensive measure of the model's precision-recall trade-off and its effectiveness in correctly identifying positive instances within the CT images. It is recommended to use the AUC-PR score in conjunction with other evaluation metrics to gain a complete understanding of the model's performance and effectiveness in binary classification tasks within the context of CT imaging.

Balanced accuracy

Balanced accuracy is an important evaluation metric used in the analysis of CT images to provide a balanced measure of an ML model's performance in binary classification tasks, especially in scenarios where the dataset is imbalanced. Balanced accuracy plays a crucial role in assessing the model's ability to correctly classify both positive and negative instances while accounting for class imbalances.

Balanced accuracy is calculated as the average of sensitivity (TPR) and specificity (true negative rate). The formula for balanced accuracy is:

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \tag{8}$$

Balanced accuracy provides a holistic measure of the model's performance by considering both the model's ability to correctly identify positive instances (sensitivity) and negative instances (specificity). A high balanced accuracy score indicates that the model effectively balances its performance in classifying both positive and negative

instances, leading to robust and reliable diagnostic outcomes. It is particularly useful in scenarios where the dataset contains class imbalances, ensuring that the model's performance is not skewed towards the majority class. By incorporating both sensitivity and specificity into a single metric, balanced accuracy offers a comprehensive assessment of the model's ability to accurately classify both positive and negative instances within the CT images.

Therefore, balanced accuracy is a valuable evaluation metric in CT image analysis, providing a balanced measure of the model's performance in binary classification tasks, especially in the presence of class imbalances. It is recommended to use balanced accuracy in conjunction with other evaluation metrics to gain a complete understanding of the model's performance and effectiveness in diagnosing and analyzing CT images accurately.

Matthews correlation coefficient (MCC)

MCC is a significant evaluation metric used in the analysis of CT images to provide a balanced measure of a ML model's performance in binary classification tasks. The MCC metric plays a crucial role in assessing the model's ability to correctly classify both positive and negative instances while accounting for class imbalances.

MCC takes into account all four confusion matrix values, providing a balanced measure even when classes are imbalanced.

$$MCC = \frac{TN \times TP - FN \times FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \times 100\% \quad [9]$$

The MCC provides a balanced measure of the model's performance by taking into account all four values in the confusion matrix. A high MCC score indicates a strong correlation between the model's predictions and the actual class labels, considering both false positives and false negatives.

MCC is particularly useful in scenarios where the dataset is imbalanced or when the classes have different prevalences. By incorporating all four values of the confusion matrix into its calculation, MCC offers a comprehensive assessment of the model's performance in correctly classifying both positive and negative instances within the CT images. Therefore, the MCC is a valuable evaluation metric in CT image analysis, providing a balanced measure of the model's performance in binary classification tasks, especially in the presence of class imbalances. It is recommended to use MCC

in conjunction with other evaluation metrics to gain a complete understanding of the model's performance and effectiveness in diagnosing and analyzing CT images accurately.

Cohen's Kappa

Cohen's Kappa is a valuable evaluation metric used in the analysis of CT images to assess the agreement between the model's predictions and the actual class labels, accounting for the possibility of agreement by chance. Cohen's Kappa plays a crucial role in evaluating the model's performance beyond what would be expected by random chance.

Cohen's Kappa is calculated using the formula:

$$Cohen's\ Kappa = \frac{Observed\ Accuracy - Expected\ Accuracy}{1 - Expected\ Accuracy} \quad [10]$$

Observed accuracy is the proportion of instances that the model correctly classified.

Expected accuracy is the accuracy that would be achieved by random chance.

Cohen's Kappa provides a measure of the model's performance that considers the agreement between the model's predictions and the actual class labels while accounting for the possibility of agreement by chance. A high Kappa score indicates a strong agreement between the model's predictions and the ground truth labels, beyond what would be expected by random chance. Cohen's Kappa is particularly useful when evaluating the model's performance in scenarios where the dataset has imbalanced class distributions or when the classes have different prevalences. By considering the expected agreement by chance, Cohen's Kappa offers a robust assessment of the model's performance in correctly classifying both positive and negative instances within the CT images.

Therefore, Cohen's Kappa is a valuable evaluation metric in CT image analysis, providing insights into the agreement between the model's predictions and the actual class labels, while considering the possibility of agreement by chance. It is recommended to use Cohen's Kappa in conjunction with other evaluation metrics to gain a comprehensive understanding of the model's performance and effectiveness in diagnosing and analyzing CT images accurately.

TPR (sensitivity)

Sensitivity, also known as the TPR, measures the proportion

998

Ahmad et al. Deep learning for CT image classification

of actual positives that are correctly identified by the model.

$$True\ Positive\ rate = \frac{TP}{TP+FN} \times 100\% \qquad [11]$$

FPR

The FPR calculates the proportion of actual negatives that are incorrectly identified as positives by the model.

$$False\ Positive\ rate = \frac{FP}{PF+TN} \times 100\% \qquad [12]$$

True negative rate (specificity)

Specificity, also known as the true negative rate, measures the proportion of actual negatives that are correctly identified by the model.

$$True\ Negative\ rate = \frac{TN}{TN+FP} \times 100\% \qquad [13]$$

False negative rate

The false negative rate calculates the proportion of actual positives that are incorrectly identified as negatives by the model.

$$False\ Negative\ rate = \frac{FN}{FN+FP} \times 100\% \qquad [14]$$

Precision-recall curve

The precision-recall curve is created by plotting precision against recall at various threshold settings.

Precision = TP/(TP + FP) which is values on the y-axis while Recall = TP/(TP + FN) values on the x-axis.

For each threshold value usually from 0 to 1:

$$Precision\text{-}Recall\ Curve = \frac{Precision}{Recall} \qquad [15]$$

Confusion matrix

A confusion matrix is a table that summarizes the model's performance by comparing actual class labels with predicted class labels, showing true positives, true negatives, false positives, and false negatives.

Classification error

Classification error calculates the proportion of misclassified

instances out of the total instances.

$$Accuracy = \frac{FP+FN}{TP+TN+FP+FN} \qquad [16]$$

ROC curve

The ROC curve is a fundamental evaluation tool used in the analysis of CT images to assess the performance of ML models in binary classification tasks, the ROC curve provides valuable insights into the model's ability to distinguish between positive and negative instances at various classification thresholds.

The ROC curve is a graphical representation of the TPR (sensitivity) against the FPR (1 – specificity) at different classification thresholds. Each point on the ROC curve represents the trade-off between sensitivity and specificity at a particular threshold setting. A model with high sensitivity and a low FPR will have a curve that approaches the upper left corner of the plot, indicating better performance. AUC-ROC is a common metric derived from the ROC curve, providing a single value that summarizes the model's performance across all possible classification thresholds. The AUC-ROC score ranges from 0 to 1, where a score closer to 1 indicates a better-performing model.

The ROC curve offers a visual representation of the model's performance in distinguishing between positive and negative instances within the CT images. It helps in evaluating the model's sensitivity and specificity trade-off and provides insights into the model's overall classification performance. The ROC curve is particularly useful for assessing the model's performance in scenarios where different classification thresholds need to be considered, or when the dataset is imbalanced. By analyzing the ROC curve and calculating the AUC-ROC score, researchers and clinicians can gain a comprehensive understanding of the model's ability to classify positive and negative instances accurately within the CT images.

Therefore, the ROC curve is a valuable evaluation tool in CT image analysis, providing a visual representation of the model's performance in binary classification tasks. It is recommended to use the ROC curve in conjunction with other evaluation metrics to gain a complete understanding of the model's performance and effectiveness in diagnosing and analyzing CT images accurately.

The ROC curve is plotted using the TPR = TP/(TP + FN) and the FPR = FP/(FP + TN). The ROC curve is calculated using TPR and FPR at various threshold levels and plotted TPR against FPR.

$$ROC\ Curve = \frac{TPR}{FPR} \qquad [17]$$

Precision at k (P@k)

P@k is an important evaluation metric used in the analysis of CT images to measure the proportion of relevant instances among the top k predictions made by the model. P@k plays a crucial role in evaluating the model's ability to accurately identify and prioritize relevant features or abnormalities within the CT images.

Precision at k is calculated as the number of relevant instances in the top k predictions divided by k. The formula for P@k is:

Precision at k measures the proportion of relevant instances among the top k predictions made by the model.

$$Precision\ at\ K\ (P@K) = \frac{Number\ of\ relevant\ documents\ in\ top\ k}{k} \qquad [18]$$

The P@k provides insights into the model's precision in identifying relevant features or abnormalities within the CT images. A high P@k score indicates that the model is effective in prioritizing and accurately identifying relevant instances within the top k predictions, leading to more efficient and targeted diagnostic outcomes. P@k is particularly useful when evaluating the model's performance in scenarios where prioritizing relevant instances is critical, such as identifying specific abnormalities or diseases within the CT images. By focusing on the precision of the model's top predictions, P@k offers valuable insights into the model's ability to provide accurate and relevant diagnostic information.

Therefore, precision at k is a valuable evaluation metric in CT image analysis, offering a measure of the model's precision in identifying relevant instances within the CT images. It is recommended to use P@k in conjunction with other evaluation metrics, such as recall, F1-score, and specificity, to gain a comprehensive understanding of the model's performance and effectiveness in diagnosing and analyzing CT images accurately.

Recall at k (R@k)

R@k is a significant evaluation metric used in the analysis of CT images to measure the proportion of relevant instances identified among the top k predictions made by the model. In the context of medical imaging, particularly CT imaging, R@k plays a crucial role in evaluating the model's ability to capture and recall relevant features or abnormalities within the CT images.

Recall at k is calculated as the number of relevant documents in the top k predictions divided by the total number of relevant documents in the dataset. The formula for R@k is:

$$Recall\ at\ K\ (R@K) = \frac{Number\ of\ relevant\ documents\ in\ top\ k}{Total\ number\ of\ relevant\ documents} \qquad [19]$$

R@k provides insights into the model's ability to recall and identify relevant instances within the top k predictions. A high R@k score indicates that the model is effective in capturing and recalling relevant features or abnormalities within the CT images, leading to more comprehensive and accurate diagnostic outcomes. R@k is particularly useful when evaluating the model's performance in scenarios where capturing all relevant instances is critical, such as detecting specific abnormalities or diseases within the CT images. By focusing on the recall of relevant instances within the top predictions, R@k offers valuable insights into the model's ability to provide comprehensive and accurate diagnostic information.

Therefore, recall at k is a valuable evaluation metric in CT image analysis, offering a measure of the model's recall in identifying relevant instances within the CT images. It is recommended to use R@k in conjunction with other evaluation metrics, such as precision, F1-score, and specificity, to gain a comprehensive understanding of the model's performance and effectiveness in diagnosing and analyzing CT images accurately.

Dice similarity coefficient (DSC)

DSC is a common evaluation metric used in medical image analysis, including CT imaging, to assess the agreement between the predicted segmentation masks and the ground truth annotations. The DSC is particularly useful for evaluating the accuracy and overlap of segmented regions, such as identifying lesions, tumors, or abnormalities within the CT images.

The DSC is calculated using the formula:

$$DSC = (2*|A \cap B|)/(|A|+|B|) \qquad [20]$$

Where: |A∩B| represents the intersection between the predicted segmentation mask (A) and the ground truth annotation (B).

|A| and |B| denotes the total number of voxels in the predicted mask and the ground truth annotation,

respectively.

The DSC provides a quantitative measure of the spatial overlap between the predicted segmentation and the ground truth, ranging from 0 (no overlap) to 1 (perfect overlap). A higher DSC score indicates a better agreement between the predicted and ground truth segmentation masks, reflecting the accuracy of the classification model in delineating specific ROIs within the CT images. By using the DSC as an evaluation metric in CT classification tasks, researchers and clinicians can assess the model's performance in accurately identifying and segmenting abnormalities or structures of interest within the CT images. The DSC offers insights into the model's ability to capture and delineate specific regions, providing a quantitative measure of segmentation accuracy and overlap.

Overall, the DSC-based CT classification offers a robust and quantitative assessment of the model's performance in segmenting and classifying ROIs within the CT images. By utilizing the DSC metric, researchers can evaluate the accuracy and spatial agreement of the model's predictions with the ground truth annotations, enhancing the reliability and effectiveness of CT image analysis and classification tasks.

Jaccard index (JI) is defined

The JI, also known as the Jaccard similarity coefficient or Jaccard coefficient, is a statistical measure used to evaluate the similarity between two sets. The JI is commonly employed to assess the agreement or overlap between predicted segmentation masks and ground truth annotations.

The JI is defined by the following equation:

$$JI = \frac{2*|A \cap B|}{|A \cup B|} \qquad [21]$$

In the equation, the numerator represents the intersection of Sets A and B, which corresponds to the number of elements that are common to both sets. The denominator represents the union of Sets A and B, which includes all elements present in either set. The JI ranges from 0 to 1, where a score of 0 indicates no overlap between the sets, and a score of 1 signifies complete overlap or perfect agreement. A higher JI value indicates a greater similarity or agreement between the sets being compared.

The JI provides a quantitative measure of the spatial overlap or similarity between the predicted segmentation mask (Set A) and the ground truth annotation (Set B). By calculating the JI, researchers can assess the accuracy and

agreement of the model's segmentation results with the reference annotations, offering insights into the model's performance in delineating specific ROIs within the CT images. Overall, the JI serves as a valuable metric in CT image analysis for quantifying the overlap and agreement between predicted and ground truth segmentation masks. By utilizing the JI, researchers can evaluate the accuracy and spatial correspondence of the model's predictions with the annotated regions, facilitating a more comprehensive assessment of the model's performance in CT image classification tasks.

These evaluation metrics are essential for assessing the performance of DL models in CT image classification tasks and can help researchers and clinicians evaluate the effectiveness of their models accurately.

Challenge and future directions

CT image classification using DL and FMs faces multifaceted challenges. Data scarcity and imbalance, particularly for rare conditions, hinder model development. Annotation quality and inter-observer variability affect training data reliability. Image quality issues, such as noise and artifacts, impact classification accuracy. Variability in anatomy and pathology presentation complicates model generalization. Computational challenges include resource intensity and model interpretability. Regulatory and ethical concerns involve data privacy and compliance. Generalization across diverse populations and robustness to variations in imaging protocols remain significant hurdles. Clinical adoption faces integration and validation challenges (11). Explainability and interpretability of model decisions are crucial for clinical trust. Additionally, the need for substantial computational resources, potential biases in training data, and the complexity of FMs further complicate their implementation. Addressing these challenges requires interdisciplinary collaboration, improved data collection and annotation processes, advanced model architectures, and strategies to enhance interpretability and generalizability while maintaining ethical standards and clinical relevance.

In addition to the challenges mentioned, CT image classification using DL and FMs faces significant hurdles related to image complexity and dataset limitations. CT images present intricate structural details and subtle variations that pose unique challenges for DL models. The three-dimensional nature of CT scans, with multiple slices representing different anatomical planes, increases the complexity of feature extraction and interpretation.

Moreover, the presence of diverse anatomical structures, varying tissue densities, and potential pathologies within a single image necessitates sophisticated modeling approaches to capture relevant features accurately. The high dynamic range of CT images, typically represented in Hounsfield units, requires careful preprocessing and normalization techniques to ensure optimal model performance. Additionally, the presence of imaging artifacts, such as beam hardening, motion artifacts, and metal artifacts, can significantly impact image quality and pose challenges for accurate classification.

The scarcity of large-scale, high-quality CT image datasets remains a significant bottleneck in the development of robust DL and FM models. Several factors contribute to this limitation: data collection challenges, acquiring diverse and representative CT datasets is hindered by patient privacy concerns, regulatory restrictions, and the high cost associated with CT imaging. This limitation is particularly acute for rare diseases or specific patient subpopulations (10).

Annotation burden: the process of annotating CT images is time-consuming, labor-intensive, and requires expert knowledge. This results in a limited availability of accurately labeled datasets, especially for complex classification tasks (11). Class imbalance, medical imaging datasets often suffer from severe class imbalance, with normal cases significantly outnumbering pathological cases. This imbalance can lead to biased model performance and reduced sensitivity to rare conditions (12). Lack of standardization, variations in imaging protocols, scanner types, and reconstruction algorithms across different healthcare institutions introduce heterogeneity in CT datasets, complicating the development of generalizable models (14). Limited multi-modal data, the integration of CT images with other clinically relevant data, such as electronic health records or genomic information, is often limited, hindering the development of comprehensive multi-modal classification models (1).

These challenges related to image complexity and dataset limitations compound the difficulties in developing accurate and reliable CT image classification models. Addressing these issues requires innovative approaches in data augmentation, synthetic data generation, and transfer learning techniques (178). Furthermore, collaborative efforts to establish large-scale, multi-institutional CT image repositories with standardized annotation protocols are essential to advance the field. Future research directions should focus on developing DL architectures specifically tailored to handle the complexities of CT images, incorporating domain knowledge into model design, and exploring novel semi-supervised and self-supervised learning approaches to leverage unlabeled data effectively. Additionally, the development of robust evaluation metrics and benchmarking datasets that account for the unique challenges in CT image classification is crucial for assessing model performance and facilitating meaningful comparisons across different approaches. By addressing these challenges comprehensively, the field can move towards more accurate, reliable, and clinically applicable CT image classification models, ultimately enhancing diagnostic capabilities and improving patient care (12).

A crucial challenge in developing robust DL models for CT image classification is the significant diversity in CT scan protocols. This diversity presents a substantial obstacle that needs to be addressed. Protocol Variability, CT scan protocols vary widely across institutions, regions, and even individual radiologists. These variations can include differences in slice thickness, reconstruction algorithms, contrast use, patient positioning, and scanner settings. Such diversity in protocols can lead to inconsistencies in image features and quality, potentially confounding DL models (97). Need for Protocol Re-evaluation: The diverse range of CT protocols needs to be critically re-evaluated as potential obstacles to building effective DL models. Variations in protocols can introduce unintended biases and reduce the generalizability of models across different clinical settings (178). Standardization efforts, there is a pressing need to evaluate and standardize CT protocols specifically to build high-quality datasets for DL. Standardization would involve establishing consensus guidelines on key parameters such as slice thickness, reconstruction kernels, and dose levels that optimize image quality for both human interpretation and AI analysis. The lack of protocol standardization directly affects the quality and consistency of datasets used for training and validating DL models. Inconsistent protocols can lead to variability in image features, potentially reducing the effectiveness of learned representations (179). Challenges in Multi-center Studies, The diversity in protocols poses significant challenges for multi-center studies and the development of broadly applicable AI models. It complicates the process of aggregating data from multiple sources and can introduce site-specific biases (191).

Addressing these protocol-related challenges requires collaborative efforts between radiologists, medical physicists, and AI researchers. Initiatives to develop and implement standardized CT protocols for AI

applications, while maintaining flexibility for clinical needs are crucial. Such efforts would not only improve the quality and consistency of datasets but also enhance the generalizability and clinical applicability of DL models in CT image classification. Furthermore, developing methods to harmonize or normalize images from diverse protocols, either through pre-processing techniques or by incorporating protocol information directly into the model architecture, could help mitigate the impact of protocol variability on model performance (192). By focusing on these protocol-related challenges and working towards standardization, the field can move towards creating more reliable, consistent, and high-quality datasets. This, in turn, will support the development of more robust and generalizable DL models for CT image classification, ultimately improving their clinical utility and impact on patient care. Recent advancements in image processing technologies, such as GE True Fidelity, offer promising opportunities to improve the quality of CT images. These solutions can enhance image clarity and reduce noise, potentially leading to more accurate nodule detection and classification. Incorporating images processed with these advanced technologies into training datasets may help overcome some of the current challenges in building effective DL models. The discussion should consider whether leveraging such technologies could be a viable strategy for improving dataset quality and model performance, particularly in situations where high-quality images are essential for accurate diagnosis (179).

Limitations of the study

This paper aims to review and discuss various DL-based systems for diagnosing COVID-19, and nodules using CT images. While the paper covers many important aspects found in the literature, it also identifies several limitations that should be addressed in future research. Firstly, the focus is on describing diagnosis systems based on DL and FM techniques without providing detailed explanations of the underlying mathematical concepts, assuming a certain level of domain-specific knowledge. Secondly, specific details of the neural networks, such as layer specifications, learning rates, and optimization techniques, are not thoroughly discussed, with readers directed to related references for more information. Thirdly, while the review discusses the diagnosis of nodules and COVID-19 from a computer vision perspective, it does not present qualitative results of the diagnosis in CT images. Fourthly, while

many of the reviewed systems show high accuracy rates, the real-world reliability of these systems is not adequately evaluated. Lastly, the paper does not include computer code or practical examples to demonstrate the significant results of the reviewed, nodule and COVID-19 diagnosis systems.

Conclusions

In conclusion, this review highlights the significant advancements in CT image classification for detecting COVID-19 and lung nodules. It traces the evolution from traditional methods to early deep DL models, culminating in sophisticated architectures tailored for CT analysis. By discussing diverse approaches, including CNNs, RNNs, GANs, and FMs like BERT, GPT, CLIP, and ViT. We emphasize their crucial role in enhancing classification accuracy. The review identifies key challenges such as data scarcity, model generalization, and clinical integration, stressing the necessity for improved network architectures and larger pre-training datasets. The methodologies discussed can serve as valuable tools for medical teams in densely populated areas, where rapid diagnosis is essential. However, obstacles remain, including insufficient labeled data, reproducibility across multi-center datasets, and the difficulty in differentiating COVID-19 from other pneumonia cases due to similar CT imaging characteristics. To address these challenges, there is an urgent need for intelligent and precise CAD systems that leverage adaptive DL models, incorporating active and incremental learning approaches. Future research directions should focus on innovations in model design, multi-modal learning, and real-world applications, ultimately driving advancements in personalized medicine and precision diagnostics. This comprehensive review underscores the transformative potential of DL in revolutionizing CT image classification, paving the way for improved disease detection and enhanced clinical decision-making.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the

Narrative Review reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-24-1400/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-24-1400/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Liu CJ, Zhang L, Sun Y, Geng L, Wang R, Shi KM, Wan JX. Application of CT and MRI images based on an artificial intelligence algorithm for predicting lymph node metastasis in breast cancer patients: a meta-analysis. BMC Cancer 2023;23:1134.
2. Sun Z, Zhang N, Li Y, Xu X. A systematic review of chest imaging findings in COVID-19. Quant Imaging Med Surg 2020;10:1058-79.
3. Miles K, Charnsangavej C, Cuenod C. Computed tomography perfusion: a historical perspective Anne E Miles and Kenneth A Miles. Multi-Detector Computed Tomography in Oncology: CRC Press; 2007. p. 17-30.
4. Hounsfield GN. Computerized transverse axial scanning (tomography). 1. Description of system. Br J Radiol 1973;46:1016-22.
5. Ahmad IS, Li N, Wang T, Liu X, Dai J, Chan Y, Liu H, Zhu J, Kong W, Lu Z, Xie Y, Liang X. COVID-19 Detection via Ultra-Low-Dose X-ray Images Enabled by Deep Learning. Bioengineering (Basel) 2023;10:1314.
6. Lan T, Zeng F, Yi Z, Xu X, Zhu M. ICNoduleNet: Enhancing Pulmonary Nodule Detection Performance on

7. Sharp Kernel CT Imaging. IEEE J Biomed Health Inform 2024;28:4751-60.
7. Javed R, Abbas T, Khan AH, Daud A, Bukhari A, Alharbey R. Deep learning for lungs cancer detection: a review. Artificial Intelligence Review 2024;57:197.
8. Thanoon MA, Zulkifley MA, Mohd Zainuri MAA, Abdani SR. A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images. Diagnostics (Basel) 2023.
9. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, Ashrafian H, Darzi A. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit Med 2021;4:65.
10. Dong D, Tang Z, Wang S, Hui H, Gong L, Lu Y, Xue Z, Liao H, Chen F, Yang F, Jin R, Wang K, Liu Z, Wei J, Mu W, Zhang H, Jiang J, Tian J, Li H. The Role of Imaging in the Detection and Management of COVID-19: A Review. IEEE Rev Biomed Eng 2021;14:16-29.
11. Gupta K, Bajaj V. Deep learning models-based CT-scan image classification for automated screening of COVID-19. Biomed Signal Process Control 2023;80:104268.
12. Szczykutowicz TP, Toia GV, Dhanantwari A, Nett B. A review of deep learning CT reconstruction: concepts, limitations, and promise in clinical practice. Current Radiology Reports 2022;10:101-15.
13. Luo X, Li Z, Xu C, Zhang B, Zhang L, Zhu J, Huang P, Wang X, Yang M, Chang S. Semi-Supervised Thyroid Nodule Detection in Ultrasound Videos. IEEE Trans Med Imaging 2024;43:1792-803.
14. Siddique N, Paheding S, Elkin CP, Devabhaktuni V. U-net and its variants for medical image segmentation: A review of theory and applications. IEEE Access 2021;9:82031-57.
15. Xu C, Guo H, Xu M, Duan M, Wang M, Liu P, Luo X, Jin Z, Liu H, Wang Y. Automatic coronary artery calcium scoring on routine chest computed tomography (CT): comparison of a deep learning algorithm and a dedicated calcium scoring CT. Quant Imaging Med Surg 2022;12:2684-95.
16. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 2012;25.
17. Zhang W, Salmi A, Jiang F, Yang C. Enhancing Pulmonary Nodule Detection Rate using 3D Convolutional Neural Networks with Optical Flow Frame Insertion Technique. IEEE Access. 2024;12:112881-95.
18. Shi C, Shao Y, Shan F, Shen J, Huang X, Chen C, Lu

1004

Ahmad et al. Deep learning for CT image classification

Y, Zhan Y, Shi N, Wu J, Wang K, Gao Y, Shi Y, Song F. Development and validation of a deep learning model for multicategory pneumonia classification on chest computed tomography: a multicenter and multireader study. Quant Imaging Med Surg 2023;13:8641-56.

19. Khatami F, Saatchi M, Zadeh SST, Aghamir ZS, Shabestari AN, Reis LO, Aghamir SMK. A meta-analysis of accuracy and sensitivity of chest CT and RT-PCR in COVID-19 diagnosis. Sci Rep 2020;10:22402.

20. Cortés E, Sánchez S. Deep Learning Transfer with AlexNet for chest X-ray COVID-19 recognition. IEEE Latin America Transactions 2021;19:944-51.

21. Zouch W, Sagga D, Echtioui A, Khemakhem R, Ghorbel M, Mhiri C, Hamida AB. Detection of COVID-19 from CT and Chest X-ray Images Using Deep Learning Models. Ann Biomed Eng 2022;50:825-35.

22. Hage Chehade A, Abdallah N, Marion J-M, Hatt M, Oueidat M, Chauvet P. A Systematic Review: Classification of Lung Diseases from Chest X-Ray Images Using Deep Learning Algorithms. SN Computer Science 2024;5:405.

23. Li J, Chen J, Tang Y, Wang C, Landman BA, Zhou SK. Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives. Med Image Anal 2023;85:102762.

24. Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020.

25. Oquab M, Darcet T, Moutakanni T, Vo H, Szafraniec M, Khalidov V, et al. Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:230407193. 2023.

26. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al., editors. Learning transferable visual models from natural language supervision. International Conference on Machine Learning; 2021: PMLR.

27. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al., editors. Segment anything. Proceedings of the IEEE/CVF International Conference on Computer Vision; 2023.

28. He S, Bao R, Li J, Grant PE, Ou Y. Accuracy of segment-anything model (sam) in medical image segmentation tasks. arXiv preprint arXiv:230409324. 2023.

29. Liu Y, Zhang J, She Z, Kheradmand A, Armand M. Samm (segment any medical model): A 3d slicer integration to sam. arXiv preprint arXiv:230405622. 2023.

30. Zhang S, Metaxas D. On the challenges and perspectives of foundation models for medical image analysis. Med Image Anal 2024;91:102996.

31. Butoi VI, Ortiz JJG, Ma T, Sabuncu MR, Guttag J, Dalca AV, editors. Universeg: Universal medical image segmentation. Proceedings of the IEEE/CVF International Conference on Computer Vision; 2023.

32. Li X, Jia M, Islam MT, Yu L, Xing L. Self-Supervised Feature Learning via Exploiting Multi-Modal Data for Retinal Disease Diagnosis. IEEE Trans Med Imaging 2020;39:4023-33.

33. Wu C, Zhang X, Zhang Y, Wang Y, Xie W. Towards generalist foundation model for radiology. arXiv preprint arXiv:230802463. 2023.

34. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, Rajpurkar P. Foundation models for generalist medical artificial intelligence. Nature 2023;616:259-65.

35. Ali N, Yadav J, editors. Computer-Aided Detection and Diagnosis of Lung Nodules Using CT Scan Images: An Analytical Review. Proceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021; 2021: Springer.

36. Christina Sweetline B, Vijayakumaran C, editors. A Comprehensive Survey on Deep Learning-Based Pulmonary Nodule Identification on CT Images. International Conference on Advances in Data-driven Computing and Intelligent Systems; 2022: Springer.

37. Cao W, Wu R, Cao G, He Z. A comprehensive review of computer-aided diagnosis of pulmonary nodules based on computed tomography scans. IEEE Access 2020;8:154007-23.

38. Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, Velazquez ER, Lambin P, Gu Y, Balagurunathan Y, Eikman E, Gatenby RA, Eschrich S, Gillies RJ. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. PLoS One 2015;10:e0118261.

39. Machtay M, Duan F, Siegel BA, Snyder BS, Gorelick JJ, Reddin JS, et al. Prediction of survival by [18F] fluorodeoxyglucose positron emission tomography in patients with locally advanced non–small-cell lung cancer undergoing definitive chemoradiation therapy: results of the ACRIN 6668/RTOG 0235 trial. J Clin Oncol 2013;31:3823.

40. Wang L. Deep Learning Techniques to Diagnose Lung Cancer. Cancers (Basel) 2022.

41. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM,

Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.

42. Morozov SP, Andreychenko AE, Pavlov N, Vladzymyrskyy A, Ledikhova N, Gombolevskiy V, et al. Mosmeddata: Chest ct scans with covid-19 related findings dataset. arXiv preprint arXiv:200506465. 2020.

43. Ma X, Ng M, Xu S, Xu Z, Qiu H, Liu Y, Lyu J, You J, Zhao P, Wang S, Tang Y, Cui H, Yu C, Wang F, Shao F, Sun P, Tang Z. Development and validation of prognosis model of mortality risk in patients with COVID-19. Epidemiol Infect 2020;148:e168.

44. Hofmanninger J, Prayer F, Pan J, Röhrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. Eur Radiol Exp 2020;4:50.

45. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020;323:1239-42.

46. Liu H, Liu F, Li J, Zhang T, Wang D, Lan W. Clinical and CT imaging features of the COVID-19 pneumonia: Focus on pregnant women and children. J Infect 2020;80:e7-e13.

47. Martínez Chamorro E, Díez Tascón A, Ibáñez Sanz L, Ossaba Vélez S, Borruel Nacenta S. Radiologic diagnosis of patients with COVID-19. Radiologia (Engl Ed) 2021;63:56-73.

48. Chen J, Wu L, Zhang J, Zhang L, Gong D, Zhao Y, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. Sci Rep 2020;10:19196.

49. Fang M, He B, Li L, Dong D, Yang X, Li C, et al. CT radiomics can help screen the coronavirus disease 2019 (COVID-19): a preliminary study. Science China Information Sciences 2020;63:1-8.

50. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X, Xu B. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). Eur Radiol 2021;31:6096-104.

51. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, et al. A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia. Engineering (Beijing) 2020;6:1122-9.

52. Jin C, Chen W, Cao Y, Xu Z, Tan Z, Zhang X, Deng L, Zheng C, Zhou J, Shi H, Feng J. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. Nat Commun 2020;11:5088.

53. Song Y, Zheng S, Li L, Zhang X, Zhang X, Huang Z, Chen J, Wang R, Zhao H, Chong Y, Shen J, Zha Y, Yang Y. Deep Learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) With CT Images. IEEE/ACM Trans Comput Biol Bioinform 2021;18:2775-80.

54. Kaur T, Gandhi TK. Classifier Fusion for Detection of COVID-19 from CT Scans. Circuits Syst Signal Process 2022;41:3397-414.

55. Mishra AK, Das SK, Roy P, Bandyopadhyay S. Identifying COVID19 from Chest CT Images: A Deep Convolutional Neural Networks Based Approach. J Healthc Eng 2020;2020:8843664.

56. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, Cao K, Liu D, Wang G, Xu Q, Fang X, Zhang S, Xia J, Xia J. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology 2020;296:E65-E71.

57. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, Wang M, Qiu X, Li H, Yu H, Gong W, Bai Y, Li L, Zhu Y, Wang L, Tian J. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. Eur Respir J 2020;56:2000775.

58. Gaur P, Malaviya V, Gupta A, Bhatia G, Pachori RB, Sharma D. COVID-19 disease identification from chest CT images using empirical wavelet transformation and transfer learning. Biomed Signal Process Control 2022;71:103076.

59. Soares E, Angelov P, Biaso S, Froes MH, Abe DK. SARSCoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. MedRxiv 2020:2020.04. 24.20078584.

60. Goel T, Murugan R, Mirjalili S, Chakrabartty DK. Automatic Screening of COVID-19 Using an Optimized Generative Adversarial Network. Cognit Comput 2021. [Epub ahead of print]. doi: 10.1007/s12559-020-09785-7.

61. Lu SY, Zhang Z, Zhang YD, Wang SH. CGENet: A Deep Graph Model for COVID-19 Detection Based on Chest CT. Biology (Basel) 2021.

62. Basu A, Sheikh KH, Cuevas E, Sarkar R. COVID-19 detection from CT scans using a two-stage framework. Expert Syst Appl 2022;193:116377.

63. Wu X, Hui H, Niu M, Li L, Wang L, He B, Yang X, Li L, Li H, Tian J, Zha Y. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study. Eur J Radiol 2020;128:109041.

64. Yousefzadeh M, Esfahanian P, Movahed SMS, Gorgin S,

1006

Ahmad et al. Deep learning for CT image classification

Rahmati D, Abedini A, Nadji SA, Haseli S, Bakhshayesh Karam M, Kiani A, Hoseinyazdi M, Roshandel J, Lashgari R. ai-corona: Radiologist-assistant deep learning framework for COVID-19 diagnosis in chest CT scans. PLoS One 2021;16:e0250952.

65. Wang B, Jin S, Yan Q, Xu H, Luo C, Wei L, et al. AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system. Appl Soft Comput 2021;98:106897.

66. Javaheri T, Homayounfar M, Amoozgar Z, Reiazi R, Homayounieh F, Abbas E, et al. Covidctnet: An open-source deep learning approach to identify covid-19 using ct image. arXiv preprint arXiv:200503059. 2020.

67. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comput Biol Med 2020;121:103795.

68. Cifci MA. Deep learning model for diagnosis of corona virus disease from CT images. Int J Sci Eng Res 2020;11:273-8.

69. Elghamrawy S, Hassanien AE. Diagnosis and prediction model for COVID-19 patient's response to treatment based on convolutional neural networks and whale optimization algorithm using CT images. MedRxiv. 2020:2020.04. 16.20063990.

70. He X, Yang X, Zhang S, Zhao J, Zhang Y, Xing E, et al. Sample-efficient deep learning for COVID-19 diagnosis based on CT scans. medrxiv. 2020:2020.04. 13.20063941.

71. Liu B, Liu P, Dai L, Yang Y, Xie P, Tan Y, et al. Assisting scalable diagnosis automatically via CT images in the combat against COVID-19. Sci Rep 2021;11:4145.

72. Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. Deep learning-based detection for COVID-19 from chest CT using weak label. MedRxiv. 2020:2020.03. 12.20027185.

73. Hasan AM, Al-Jawad MM, Jalab HA, Shaiba H, Ibrahim RW, Al-Shamasneh AR. Classification of Covid-19 Coronavirus, Pneumonia and Healthy Lungs in CT Scans Using Q-Deformed Entropy and Deep Learning Features. Entropy (Basel) 2020.

74. Amyar A, Modzelewski R, Li H, Ruan S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. Comput Biol Med 2020;126:104037.

75. Alaiad AI, Mugdadi EA, Hmeidi II, Obeidat N, Abualigah L. Predicting the severity of COVID-19 fromlung CT images using novel deep learning. Journal of Medical and

Biological Engineering 2023;43:135-46.

76. Ullah N, Khan JA, El-Sappagh S, El-Rashidy N, Khan MS. A holistic approach to identify and classify COVID-19 from chest radiographs, ECG, and CT-scan images using shufflenet convolutional neural network. Diagnostics 2023;13:162.

77. Silva P, Luz E, Silva G, Moreira G, Silva R, Lucio D, et al. COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis. Informatics in Medicine Unlocked 2020;20:100427.

78. Shi F, Xia L, Shan F, Song B, Wu D, Wei Y, Yuan H, Jiang H, He Y, Gao Y, Sui H, Shen D. Large-scale screening to distinguish between COVID-19 and community-acquired pneumonia using infection size-aware classification. Phys Med Biol 2021;66:065031.

79. Shi W, Peng X, Liu T, Cheng Z, Lu H, Yang S, Zhang J, Wang M, Gao Y, Shi Y, Zhang Z, Shan F. A deep learning-based quantitative computed tomography model for predicting the severity of COVID-19: a retrospective study of 196 patients. Ann Transl Med 2021;9:216.

80. Afif M, Ayachi R, Said Y, Atri M. Deep learning-based technique for lesions segmentation in CT scan images for COVID-19 prediction. Multimedia Tools and Applications 2023;82:26885-99.

81. Khan SH, Alahmadi TJ, Alsahfi T, Alsadhan AA, Mazroa AA, Alkahtani HK, et al. COVID-19 infection analysis framework using novel boosted CNNs and radiological images. Sci Rep 2023;13:21837.

82. Sharma S. Drawing insights from COVID-19-infected patients using CT scan images and machine learning techniques: a study on 200 patients. Environmental Science and Pollution Research 2020;27:37155-63.

83. Kathamuthu ND, Subramaniam S, Le QH, Muthusamy S, Panchal H, Sundararajan SCM, et al. A deep transfer learning-based convolution neural network model for COVID-19 detection using computed tomography scan images for medical applications. Advances in Engineering Software 2023;175:103317.

84. Motwani A, Shukla PK, Pawar M, Kumar M, Ghosh U, Alnumay W, et al. Enhanced framework for COVID-19 prediction with computed tomography scan images using dense convolutional neural network and novel loss function. Computers and Electrical Engineering 2023;105:108479.

85. Gozes O, Frid-Adar M, Greenspan H, Browning PD, Zhang H, Ji W, et al. Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring

using deep learning ct image analysis. arXiv preprint arXiv:200305037. 2020.

86. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Lung infection quantification of COVID-19 in CT images with deep learning. arXiv preprint arXiv:200304655. 2020.

87. Sahoo P, Saha S, Mondal S, Gowda S, editors. Vision transformer based covid-19 detection using chest ct-scan images. 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); 2022: IEEE.

88. Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V. A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. Chaos Solitons Fractals 2020;140:110190.

89. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020.

90. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018.

91. Tekade R, Rajeswari K, editors. Lung cancer detection and classification using deep learning. 2018 fourth international conference on computing communication control and automation (ICCUBEA); 2018: IEEE.

92. Wang S, Dong L, Wang X, Wang X. Classification of Pathological Types of Lung Cancer from CT Images by Deep Residual Neural Networks with Transfer Learning Strategy. Open Med (Wars) 2020;15:190-7.

93. Chen J, Ma Q, Wang W, editors. A lung cancer detection system based on convolutional neural networks and natural language processing. 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT); 2021: IEEE.

94. Humayun M, Sujatha R, Almuayqil SN, Jhanjhi NZ. A Transfer Learning Approach with a Convolutional Neural Network for the Classification of Lung Carcinoma. Healthcare (Basel) 2022.

95. Al-Yasriy HF, Al-Husieny MS, Mohsen FY, Khalil EA, Hassan ZS, editors. Diagnosis of lung cancer based on CT scans using CNN. IOP Conference Series: Materials Science and Engineering; 2020: IOP Publishing.

96. Al-Huseiny MS, Sajit AS. Transfer learning with GoogLeNet for detection of lung cancer. Indonesian Journal of Electrical Engineering and Computer Science 2021;22:1078-86.

97. Raza R, Zulfiqar F, Khan MO, Arif M, Alvi A, Iftikhar MA, et al. Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. Engineering Applications of Artificial Intelligence 2023;126:106902.

98. Mohamed TIA, Oyelade ON, Ezugwu AE. Automatic detection and classification of lung cancer CT scans based on deep learning and ebola optimization search algorithm. PLoS One 2023;18:e0285796.

99. Gulsoy T, Kablan EB, editors. Diagnosis of lung cancer based on CT scans using Vision Transformers. 2023 14th International Conference on Electrical and Electronics Engineering (ELECO); 2023: IEEE.

100. Sahin ME, Ulutas H, Yuce E, Erkoc MF. Detection and classification of COVID-19 by using faster R-CNN and mask R-CNN on CT images. Neural Comput Appl 2023;35:13597-611.

101. Gunraj H, Wang L, Wong A. COVIDNet-CT: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases From Chest CT Images. Front Med (Lausanne) 2020;7:608525.

102. Shi W, Tong L, Zhu Y, Wang MD. COVID-19 Automatic Diagnosis With Radiographic Imaging: Explainable Attention Transfer Deep Neural Networks. IEEE J Biomed Health Inform 2021;25:2376-87.

103. Yu X, Lu S, Guo L, Wang SH, Zhang YD. ResGNet-C: A graph convolutional neural network for detection of COVID-19. Neurocomputing (Amst) 2021;452:592-605.

104. Mondal AK, Bhattacharjee A, Singla P, Prathosh AP. xViTCOS: Explainable Vision Transformer Based COVID-19 Screening Using Radiography. IEEE J Transl Eng Health Med 2022;10:1100110.

105. Harmon SA, Sanford TH, Xu S, Turkbey EB, Roth H, Xu Z, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. Nat Commun 2020;11:4080.

106. Ouyang X, Huo J, Xia L, Shan F, Liu J, Mo Z, Yan F, Ding Z, Yang Q, Song B, Shi F, Yuan H, Wei Y, Cao X, Gao Y, Wu D, Wang Q, Shen D. Dual-Sampling Attention Network for Diagnosis of COVID-19 From Community Acquired Pneumonia. IEEE Trans Med Imaging 2020;39:2595-605.

107. Sun L, Mo Z, Yan F, Xia L, Shan F, Ding Z, Song B, Gao W, Shao W, Shi F, Yuan H, Jiang H, Wu D, Wei Y, Gao Y, Sui H, Zhang D, Shen D. Adaptive Feature Selection Guided Deep Forest for COVID-19 Classification With Chest CT. IEEE J Biomed Health Inform 2020;24:2798-805.

108. Wang J, Bao Y, Wen Y, Lu H, Luo H, Xiang Y, Li X, Liu C, Qian D. Prior-Attention Residual Learning for More Discriminative COVID-19 Screening in CT Images.

1008

Ahmad et al. Deep learning for CT image classification

IEEE Trans Med Imaging 2020;39:2572-83.

109. Butt C, Gill J, Chun D, Babu BA. Deep learning system to screen coronavirus disease 2019 pneumonia. Appl Intell 2023;53:4874.

110. Huang G, Wei X, Tang H, Bai F, Lin X, Xue D. A systematic review and meta-analysis of diagnostic performance and physicians' perceptions of artificial intelligence (AI)-assisted CT diagnostic technology for the classification of pulmonary nodules. J Thorac Dis 2021;13:4797.

111. Yadav G, Maheshwari S, Agarwal A, editors. Contrast limited adaptive histogram equalization based enhancement for real time video system. 2014 international conference on advances in computing, communications and informatics (ICACCI); 2014: IEEE.

112. Veldhuizen TL, Jernigan ME, editors. Grid filters for local nonlinear image restoration. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat No 98CH36181); 1998: IEEE.

113. Tian J, Tian J, Wang Y, Dai X, Zhang X. Medical image processing and analysis. Molecular Imaging: Fundamentals and Applications 2013:415-69.

114. Prabha DS, Kumar JS, editors. Performance analysis of image smoothing methods for low level of distortion. 2016 IEEE International Conference on Advances in Computer Applications (ICACA); 2016: IEEE.

115. Kociołek M, Strzelecki M, Obuchowicz R. Does image normalization and intensity resolution impact texture classification? Comput Med Imaging Graph 2020;81:101716.

116. Lehmann TM, Gönner C, Spitzer K. Survey: interpolation methods in medical image processing. IEEE Trans Med Imaging 1999;18:1049-75.

117. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, et al. Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing 1987;39:355-68.

118. Gungor MA. A comparative study on wavelet denoising for high noisy CT images of COVID-19 disease. Optik (Stuttg) 2021;235:166652.

119. Zhu Y, Wang C, Dong C, Zhang K, Gao H, Yuan C. High-Frequency Normalizing Flow for Image Rescaling. IEEE Trans Image Process 2023;32:6223-33.

120. Luo Z, Shi D, Gan W-S, Huang Q. Delayless generative fixed-filter active noise control based on deep learning and bayesian filter. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2023.

121. Jung W, Jeon E, Kang E, Suk HI. EAG-RS: A Novel Explainability-Guided ROI-Selection Framework for ASD Diagnosis via Inter-Regional Relation Learning. IEEE Trans Med Imaging 2024;43:1400-11.

122. Zhang C, He W, Liu L, Dai J, Salim Ahmad I, Xie Y, Liang X. Volumetric feature points integration with bio-structure-informed guidance for deformable multi-modal CT image registration. Phys Med Biol 2023.

123. He W, Zhang C, Dai J, Liu L, Wang T, Liu X, Jiang Y, Li N, Xiong J, Wang L, Xie Y, Liang X. A statistical deformation model-based data augmentation method for volumetric medical image segmentation. Med Image Anal 2024;91:102984.

124. Li N, Zhou S, Zhao G, Zhang Z, Xie Y, Liang X. Iterative stripe artifact correction framework for TOF-MRA. Comput Biol Med 2021;134:104456.

125. John J, Mini M. Multilevel thresholding based segmentation and feature extraction for pulmonary nodule detection. Procedia Technology 2016;24:957-63.

126. Ayshath Thabsheera A, Thasleema T, Rajesh R. Lung cancer detection using CT scan images: A review on various image processing techniques. Data Analytics and Learning: Proceedings of DAL 2018;2018:413-9.

127. Javaid M, Javid M, Rehman MZ, Shah SI. A novel approach to CAD system for the detection of lung nodules in CT images. Comput Methods Programs Biomed 2016;135:125-39.

128. Elavarasu M, Govindaraju K. Effectiveness of filtering methods in enhancing pulmonary carcinoma image quality: a comparative analysis. International Journal of Electrical & Computer Engineering (2088-8708) 2024;14:358-65.

129. Vignesh V, Kothavari K, editors. Classification and detection of lung nodules using virtual dual energy in CXR images. 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE); 2014: IEEE.

130. Fotin SV, Reeves AP, Biancardi AM, Yankelevitz DF, Henschke CI, editors. A multiscale Laplacian of Gaussian filtering approach to automated pulmonary nodule detection from whole-lung low-dose CT scans. Medical Imaging 2009: Computer-Aided Diagnosis; 2009: SPIE.

131. Ahmed I, Chehri A, Jeon G, Piccialli F. Automated Pulmonary Nodule Classification and Detection Using Deep Learning Architectures. IEEE/ACM Trans Comput Biol Bioinform 2023;20:2445-56.

132. Ozdemir O, Russell RL, Berlin AA. A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans. IEEE Trans

Med Imaging 2020;39:1419-29.

133. Li F, Huang H, Wu Y, Cai C, Huang Y, Ding X, editors. Lung nodule detection with a 3d convnet via iou self-normalization and maxout unit. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019: IEEE.

134. Masood A, Yang P, Sheng B, Li H, Li P, Qin J, Lanfranchi V, Kim J, Feng DD. Cloud-Based Automated Clinical Decision Support System for Detection and Diagnosis of Lung Cancer in Chest CT. IEEE J Transl Eng Health Med 2020;8:4300113.

135. El-Regaily SA, Salem MAM, Aziz MHA, Roushdy MI. Multi-view Convolutional Neural Network for lung nodule false positive reduction. Expert Systems with Applications 2020;162:113017.

136. Wang J, Wang J, Wen Y, Lu H, Niu T, Pan J, et al. Pulmonary nodule detection in volumetric chest CT scans using CNNs-based nodule-size-adaptive detection and classification. IEEE Access 2019;7:46033-44.

137. Zheng S, Cornelissen LJ, Cui X, Jing X, Veldhuis RNJ, Oudkerk M, van Ooijen PMA. Deep convolutional neural networks for multiplanar lung nodule detection: Improvement in small nodule identification. Med Phys 2021;48:733-44.

138. Ali I, Muzammil M, Haq IU, Amir M, Abdullah S. Efficient lung nodule classification using transferable texture convolutional neural network. IEEE Access 2020;8:175859-70.

139. Zuo W, Zhou F, He Y. An Embedded Multi-branch 3D Convolution Neural Network for False Positive Reduction in Lung Nodule Detection. J Digit Imaging 2020;33:846-57.

140. Zuo W, Zhou F, Li Z, Wang L. Multi-resolution CNN and knowledge transfer for candidate classification in lung nodule detection. IEEE Access 2019;7:32510-21.

141. Zheng S, Guo J, Cui X, Veldhuis RNJ, Oudkerk M, van Ooijen PMA. Automatic Pulmonary Nodule Detection in CT Scans Using Convolutional Neural Networks Based on Maximum Intensity Projection. IEEE Trans Med Imaging 2020;39:797-805.

142. Liu J, Cao L, Akin O, Tian Y. Accurate and robust pulmonary nodule detection by 3D feature pyramid network with self-supervised feature learning. arXiv preprint arXiv:190711704. 2019.

143. Wang D, Zhang Y, Zhang K, Wang L, editors. Focalmix: Semi-supervised learning for 3d medical image detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020.

144. Zhai P, Tao Y, Chen H, Cai T, Li J. Multi-task learning for lung nodule classification on chest CT. IEEE Access 2020;8:180317-27.

145. Liu S, Setio AAA, Ghesu FC, Gibson E, Grbic S, Georgescu B, Comaniciu D. No Surprises: Training Robust Lung Nodule Detection for Low-Dose CT Scans by Augmenting With Adversarial Attacks. IEEE Trans Med Imaging 2021;40:335-45.

146. Shi Z, Hao H, Zhao M, Feng Y, He L, Wang Y, et al. A deep CNN based transfer learning method for false positive reduction. Multimedia Tools and Applications 2019;78:1017-33.

147. Zhou Z, Sodha V, Siddiquee MMR, Feng R, Tajbakhsh N, Gotway MB, Liang J. Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis. Med Image Comput Comput Assist Interv 2019;11767:384-93.

148. Riquelme D, Akhloufi MA. Deep learning for lung cancer nodules detection and classification in CT scans. AI 2020;1:28-67.

149. Ren Y, Tsai M-Y, Chen L, Wang J, Li S, Liu Y, et al. A manifold learning regularization approach to enhance 3D CT image-based lung nodule classification. International Journal of Computer Assisted Radiology and Surgery 2020;15:287-95.

150. Al-Shabi M, Lee HK, Tan M. Gated-dilated networks for lung nodule classification in CT scans. IEEE Access 2019;7:178827-38.

151. Liu L, Dou Q, Chen H, Qin J, Heng PA. Multi-Task Deep Model With Margin Ranking Loss for Lung Nodule Analysis. IEEE Trans Med Imaging 2020;39:718-28.

152. Xie Y, Xia Y, Zhang J, Song Y, Feng D, Fulham M, Cai W. Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT. IEEE Trans Med Imaging 2019;38:991-1004.

153. Apostolopoulos ID. Experimenting with Convolutional Neural Network Architectures for the automatic characterization of Solitary Pulmonary Nodules' malignancy rating. arXiv preprint arXiv:200306801. 2020.

154. Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H. Automated Lung Nodule Detection and Classification Using Deep Learning Combined with Multiple Strategies. Sensors (Basel) 2019.

155. Harsono IW, Liawatimena S, Cenggoro TW. Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning. Journal of King Saud University-Computer and Information Sciences 2022;34:567-77.

156. Liao F, Liang M, Li Z, Hu X, Song S. Evaluate the

Malignancy of Pulmonary Nodules Using the 3-D Deep Leaky Noisy-OR Network. IEEE Trans Neural Netw Learn Syst 2019;30:3484-95.

157. Yang J, Deng H, Huang X, Ni B, Xu Y, editors. Relational learning between multiple pulmonary nodules via deep set attention transformers. 2020 IEEE 17th international symposium on biomedical imaging (ISBI); 2020: IEEE.

158. Balagurunathan Y, Schabath MB, Wang H, Liu Y, Gillies RJ. Quantitative Imaging features Improve Discrimination of Malignancy in Pulmonary nodules. Sci Rep 2019;9:8528.

159. Hussein S, Kandel P, Bolan CW, Wallace MB, Bagci U. Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches. IEEE Trans Med Imaging 2019;38:1777-87.

160. Al-Shabi M, Lan BL, Chan WY, Ng KH, Tan M. Lung nodule classification using deep Local-Global networks. Int J Comput Assist Radiol Surg 2019;14:1815-9.

161. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, Tse D, Etemadi M, Ye W, Corrado G, Naidich DP, Shetty S. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019;25:954-61.

162. Gao R, Huo Y, Bao S, Tang Y, Antic SL, Epstein ES, Balar AB, Deppen S, Paulson AB, Sandler KL, Massion PP, Landman BA. Distanced LSTM: Time-Distanced Gates in Long Short-Term Memory Models for Lung Cancer Detection. Mach Learn Med Imaging 2019;11861:310-8.

163. Chen S, Ma K, Zheng Y. Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:190400625. 2019.

164. Zhang J, Zou W, Hu N, Zhang B, Wang J. S-Net: an S-shaped network for nodule detection in 3D CT images. Phys Med Biol 2024.

165. Yang A, Jin X, Li L, editors. CT images recognition of pulmonary tuberculosis based on improved faster RCNN and U-Net. 2019 10th international Conference on information Technology in Medicine and education (ITME); 2019: IEEE.

166. Wu Y, Wang H, Wu F, editors. Automatic classification of pulmonary tuberculosis and sarcoidosis based on random forest. 2017 10th International congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI); 2017: IEEE.

167. Li L, Huang H, Jin X, editors. AE-CNN classification of pulmonary tuberculosis based on CT images. 2018 9th international conference on information technology in medicine and education (ITME); 2018: IEEE.

168. Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, et al. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recognition. 2017;61:663-73.

169. Guo K, Liu X, Soomro NQ, Liu Y. A novel 2D ground-glass opacity detection method through local-to-global multilevel thresholding for segmentation and minimum bayes risk learning for classification. Journal of Medical Imaging and Health Informatics. 2016;6(5):1193-201.

170. Huang S, Liu X, Han G, Zhao X, Zhao Y, Zhou C, editors. 3D GGO candidate extraction in lung CT images using multilevel thresholding on supervoxels. Medical Imaging 2018: Computer-Aided Diagnosis; 2018: SPIE.

171. Cui W, Wang Y, editors. A Lung Calcification Detection Method through Improved Two-dimensional OTSU and the combined features. 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA); 2019: IEEE.

172. Song Y, Cai W, Huang H, Zhou Y, Feng DD, Yue Wang, Fulham MJ, Chen M. Large Margin Local Estimate With Applications to Medical Image Classification. IEEE Trans Med Imaging 2015;34:1362-77.

173. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans Med Imaging 2016;35:1285-98.

174. Onishi Y, Teramoto A, Tsujimoto M, Tsukamoto T, Saito K, Toyama H, et al. Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. International Journal of Computer Assisted Radiology and Surgery 2020;15:173-8.

175. Gomes Ataide EJ, Ponugoti N, Illanes A, Schenke S, Kreissl M, Friebe M. Thyroid nodule classification for physician decision support using machine learning-evaluated geometric and morphological features. Sensors. 2020;20:6110.

176. Ma L, Liu X, Fei B. Learning with distribution of optimized features for recognizing common CT imaging signs of lung diseases. Phys Med Biol 2017;62:612-32.

177. Kanjanasurat I, Tenghongsakul K, Purahong B, Lasakul A. CNN-RNN Network Integration for the Diagnosis of COVID-19 Using Chest X-ray and CT Images. Sensors (Basel) 2023.

178. Celik G. Detection of Covid-19 and other pneumonia cases from CT and X-ray chest images using deep learning based on feature reuse residual block and depthwise

dilated convolutions neural network. Appl Soft Comput 2023;133:109906.

179. de Jesus Silva LF, Cortes OAC, Diniz JOB. A novel ensemble CNN model for COVID-19 classification in computerized tomography scans. Results in Control and Optimization 2023;11:100215.

180. Tajbakhsh N, Shin JY, Gotway MB, Liang J. Computer-aided detection and visualization of pulmonary embolism using a novel, compact, and discriminative image representation. Med Image Anal 2019;58:101541.

181. Helm EJ, Silva CT, Roberts HC, Manson D, Seed MT, Amaral JG, Babyn PS. Computer-aided detection for the identification of pulmonary nodules in pediatric oncology patients: initial experience. Pediatr Radiol 2009;39:685-93.

182. Armato SG 3rd, McNitt-Gray MF, Reeves AP, Meyer CR, McLennan G, Aberle DR, et al. The Lung Image Database Consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. Acad Radiol 2007;14:1409-21.

183. Martini K, Blüthgen C, Eberhard M, Schönenberger ALN, De Martini I, Huber FA, Barth BK, Euler A, Frauenfelder T. Impact of Vessel Suppressed-CT on Diagnostic Accuracy in Detection of Pulmonary Metastasis and Reading Time. Acad Radiol 2021;28:988-94.

184. Agam G, Armato SG 3rd, Wu C. Vessel tree reconstruction in thoracic CT scans with application to nodule detection. IEEE Trans Med Imaging 2005;24:486-99.

185. Meyer CR, Johnson TD, McLennan G, Aberle DR, Kazerooni EA, Macmahon H, et al. Evaluation of lung MDCT nodule annotation across radiologists and

methods. Acad Radiol 2006;13:1254-65.

186. Ross JC, Miller JV, Turner WD, Kelliher TP. An analysis of early studies released by the Lung Imaging Database Consortium (LIDC). Acad Radiol 2007;14:1382-8.

187. Nair A, Bartlett EC, Walsh SLF, Wells AU, Navani N, Hardavella G, Bhalla S, Calandriello L, Devaraj A, Goo JM, Klein JS, MacMahon H, Schaefer-Prokop CM, Seo JB, Sverzellati N, Desai SR. Variable radiological lung nodule evaluation leads to divergent management recommendations. Eur Respir J 2018;52:1801359.

188. Yi L, Zhang L, Xu X, Guo J. Multi-Label Softmax Networks for Pulmonary Nodule Classification Using Unbalanced and Dependent Categories. IEEE Trans Med Imaging 2023;42:317-28.

189. Dodia S, Annappa B, Mahesh PA. Recent advancements in deep learning based lung cancer detection: A systematic review. Engineering Applications of Artificial Intelligence 2022;116:105490.

190. Ghorpade H, Jagtap J, Patil S, Kotecha K, Abraham A, Horvat N, et al. Automatic Segmentation of Pancreas and Pancreatic Tumor: A Review of a Decade of Research. IEEE Access. 2023;11:108727-45.

191. Saraswat D, Bhattacharya P, Verma A, Prasad VK, Tanwar S, Sharma G, et al. Explainable AI for healthcare 5.0: opportunities and challenges. IEEE Access 2022;10:84486-517.

192. Xiao H, Li Y, Jiang B, Xia Q, Wei Y, Li H. The progress on lung computed tomography imaging signs: A review. Applied Sciences 2022;12:9367.

Quant Imaging Med Surg 2025;15(1):962-1011 | https://dx.doi.org/10.21037/qims-24-1400