

Improving diagnosis and prognosis of lung cancer using vision transformers: a scoping review

Hazrat Ali, Farida Mohsen, Zubair Shah

Item type

Journal Contribution

Terms of use

This work is licensed under a [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license

This version is available at

https://manara.qnl.qa/articles/journal_contribution/Improving_diagnosis_and_prognosis_of_lung_cancer_using_vision_transformers_a_scoping_review

Access the item on Manara for more information about usage details and recommended citation.

Posted on Manara – Qatar Research Repository on

2023-09-15

RESEARCH ARTICLE

Open Access



Improving diagnosis and prognosis of lung cancer using vision transformers: a scoping review

Hazrat Ali^{1*}, Farida Mohsen¹ and Zubair Shah^{1*}

Abstract

Background Vision transformer-based methods are advancing the field of medical artificial intelligence and cancer imaging, including lung cancer applications. Recently, many researchers have developed vision transformer-based AI methods for lung cancer diagnosis and prognosis.

Objective This scoping review aims to identify the recent developments on vision transformer-based AI methods for lung cancer imaging applications. It provides key insights into how vision transformers complemented the performance of AI and deep learning methods for lung cancer. Furthermore, the review also identifies the datasets that contributed to advancing the field.

Methods In this review, we searched Pubmed, Scopus, IEEEExplore, and Google Scholar online databases. The search terms included intervention terms (vision transformers) and the task (i.e., lung cancer, adenocarcinoma, etc.). Two reviewers independently screened the title and abstract to select relevant studies and performed the data extraction. A third reviewer was consulted to validate the inclusion and exclusion. Finally, the narrative approach was used to synthesize the data.

Results Of the 314 retrieved studies, this review included 34 studies published from 2020 to 2022. The most commonly addressed task in these studies was the classification of lung cancer types, such as lung squamous cell carcinoma versus lung adenocarcinoma, and identifying benign versus malignant pulmonary nodules. Other applications included survival prediction of lung cancer patients and segmentation of lungs. The studies lacked clear strategies for clinical transformation. SWIN transformer was a popular choice of the researchers; however, many other architectures were also reported where vision transformer was combined with convolutional neural networks or UNet model. Researchers have used the publicly available lung cancer datasets of the lung imaging database consortium and the cancer genome atlas. One study used a cluster of 48 GPUs, while other studies used one, two, or four GPUs.

Conclusion It can be concluded that vision transformer-based models are increasingly in popularity for developing AI methods for lung cancer applications. However, their computational complexity and clinical relevance are important factors to be considered for future research work. This review provides valuable insights for researchers in

*Correspondence:

Hazrat Ali
haali2@hbku.edu.qa
Zubair Shah
zshah@hbku.edu.qa

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the field of AI and healthcare to advance the state-of-the-art in lung cancer diagnosis and prognosis. We provide an interactive dashboard on lung-cancer.onrender.com/.

Keywords Adenocarcinoma, Artificial Intelligence, Convolutional neural networks, Deep learning, Diagnosis, Lung Cancer, Medical imaging, Segmentation, Survival prediction, Vision Transformers

Introduction

Lung cancer is a highly prevalent and fatal form of cancer globally [1, 2]. Over the last few decades, medical imaging techniques have played an increasingly vital role in diagnosing, prognosis, survival prediction, and early detection of lung cancer, eventually aiding in effective cure and prevention. Such techniques make use of lung computed tomography (CT), X-rays, positron emission tomography (PET), and magnetic resonance imaging (MRI). Traditionally, medical images in clinical work have been interpreted and analyzed by trained radiologists who use their expertise and experience to make accurate diagnoses. However, the manual interpretation of medical images can be time-consuming, prone to human error, and affected by intra-observer as well as inter-observer variability.

Artificial intelligence (AI) methods, particularly deep learning models, have played a vital role in automating image processing in the past few years and have been gaining increasing attention in medical imaging [3, 4]. AI methods dominated by convolutional neural networks (CNNs) [5, 6] have revolutionized the realm of medical imaging with their capability of learning complex representations enabling the automated diagnosis of diseases and the detection of abnormalities. They have demonstrated remarkable improvements in various medical imaging applications and modalities, including MRI [7, 8], CT [9], endoscopy [10], and radiography [11, 12], to name a few. However, the advent of transformers apprised researchers of CNNs' major drawback, i.e., the inability to capture long-range dependencies such as the extraction of contextual information and the non-local correlation of objects.

Recently, Dosovitskiy et al. [13] sought to apply the success of transformers in natural language processing to image processing. They developed a vision transformer to capture long-term dependencies within an image by treating image classification as a sequence prediction task for a series of image patches. On several benchmark datasets, the vision transformer and its derived instances demonstrated state-of-the-art (SOTA) performance and gained popularity in several computer vision tasks, including classification [13], segmentation [14], and detection [15]. The use of vision transformers has also been cross-pollinated into the medical image field, where they are used for image segmentation [16], synthesis [17], and disease diagnosis, resulting in SOTA performances. For lung cancer imaging applications, the use of vision

transformers has gained attention for different applications, including cancer classification, tumor segmentation, nodule detection, and survival prediction. Much new vision transformer-based AI methods for lung cancer imaging applications have recently been published by researchers.

Our scoping review aims to present a comprehensive overview of the recent studies that developed vision transformer-based AI methods for lung cancer imaging. While there are a few related reviews in the literature [18–21]; they differ in their focus and coverage. For example, the review in [18] covers the applications of vision transformers in medical imaging; however, it is not specific to lung cancer imaging applications and covers many different medical imaging applications. Similarly, the reviews in [19, 20] cover other AI methods for cancer imaging but do not include vision transformers, while the review in [21] covers AI methods for lung cancer applications but covers only pathology imaging and does not cover all the imaging modalities. Besides, it does not cover the recent developments of vision transformers for lung cancer imaging, as the review was published much earlier. To the best of our knowledge, no review study focuses specifically on the utilization of vision transformers for medical imaging in lung cancer. Therefore, our review is the first comprehensive review that focuses specifically on the use of vision transformers for medical imaging in lung cancer, providing a thorough overview of the current state of the field. Table 1 shows the literature review comparison.

The primary aim of our scoping review is to synthesize scientific literature by answering the following research questions, as listed in Fig. 1C.

We are confident that this review will provide a comprehensive text on the recent developments in vision transformer-based lung cancer imaging applications.

Results

Search results

The search retrieved 314 studies. However, 92 studies were duplicates that we removed. We removed 183 studies according to the exclusion/inclusion criteria in the title and abstract screening phase. In the remaining 39 studies, we removed eight more studies after the full-text reading phase, as they did not fulfill the inclusion criteria. We added three additional studies through forward/backward referencing. Finally, we included 34 unique studies [24–57] in the review (also see Appendix 1 for all

Table 1 Literature comparison with previous review studies

| Reference | Year | Scope and coverage | Differences with our review |
|--|---------------|--|---|
| Transformers in Medical Image Analysis [18] | August 2022 | It focuses on the use of transformers for various medical imaging applications. It is not specific to lung cancer imaging. It does not cover many recent studies on vision transformers in lung cancer imaging. | Our review focuses specifically on the use of vision transformers for lung cancer imaging. Our review covers many recent studies published in later 2022. |
| Artificial intelligence in lung cancer: current applications and perspectives [19] | November 2022 | It focuses on the current state of AI in lung cancer. It covers traditional machine learning and deep learning methods for lung nodule detection and segmentation. It does not cover vision transformer-based approaches. | Our review focuses specifically on the use of vision transformers for lung cancer imaging. |
| Artificial intelligence techniques for cancer detection in medical image processing: A review [20] | May 2021 | It focuses on a broad range of AI methods. It covers many different types of cancer imaging applications. It is not specific to vision transformers. It is not specific to lung cancer. It does not cover many recent studies. | Our review focuses specifically on the use of vision transformers for medical imaging in lung cancer. Our review is specific to vision transformers Our review is specific to lung cancer. Our review covers recent studies. |
| Artificial Intelligence in Lung Cancer Pathology Image Analysis [21] | November 2019 | It focuses on AI methods for pathology image analysis of lung cancer. It does not cover vision transformers. It does not cover recent studies. | Our review covers the use of transformers for medical imaging in lung cancer. Our review covers different imaging modalities, including pathology and CT scans. Our review is specific to vision transformers. Our review covers recent studies. |
| Recent advances of Transformers in medical image analysis: A comprehensive review [22] | March 2023 | It focuses on transformers for various medical imaging applications. It is not specific to lung cancer imaging and covers many studies on COVID-19. It does not cover many recent studies on vision transformers in lung cancer imaging. | Our review focuses specifically on the use of vision transformers for lung cancer imaging. Our review covers many recent studies published in later 2022. |
| Machine Learning for Lung Cancer Diagnosis, Treatment, and Prognosis [23] | October 2022 | It covers different machine learning and deep learning techniques. It does not cover any study on vision transformer applications. It is not specific to imaging modality and covers studies on different data modalities for lung cancer. | Our review focuses on vision transformer applications in lung cancer imaging. Our review focuses on imaging data for lung cancer. |

the included studies). Figure 1 shows the flowchart for the different phases of the study selection and the number of studies retained in each phase. Readers may access an interactive dashboard on lung-cancer.onrender.com/. (Loading may take up to 60 s)

Demographics of the included studies

In the included studies, half of the studies ($n=17$) were journal articles, while 16 studies were published in conference proceedings. Only one study was published as a thesis. Most of the studies ($n=28$) were published in 2022, four studies were published in 2021, and only two studies were published in 2020. Of the articles published in 2022, five studies were published in June, five were published in September, and four were published in August. Of the four studies published in 2021, no study was published in the first eight months. The included studies were published by researchers from seven different countries (first-author country affiliation). Researchers from China published almost two-third ($n=21$) of the studies, while researchers from the USA published approximately one-fourth ($n=8$) of the studies. Researchers from India, Saudi Arabia, Pakistan, Canada, and South Korea published one study each. Figure 2 shows a

summary of the year-wise, and Fig. 3 shows the country-wise demographics of the included studies. Table 2 summarizes the demographics of the included studies.

Main tasks addressed in the studies

In the 34 studies included in this review, one-third of studies ($n=11$) [24–35] performed classification of different types of lung cancers. Nearly half of the studies ($n=15$) [35, 43, 45–57] used vision transformer-based models to predict the growth of tumors or the course of cancer. Of these, eight studies [35, 43, 48, 53–57] developed vision transformer-based models for survival prediction of lung cancer patients. Six studies [36–43] addressed the task of segmentation of tumor or lung nodules. One study [44] performed lung nodule detection. Few studies performed more than one task. For example, one study [35] performed the classification of lung cancer types and reported performance for survival prediction too. Similarly, one study [42] performed segmentation of lungs and detection of nodules, and one study [43] reported segmentation of lungs and survival prediction of patients. Figure 4 shows a mapping of the different tasks addressed in the included studies.

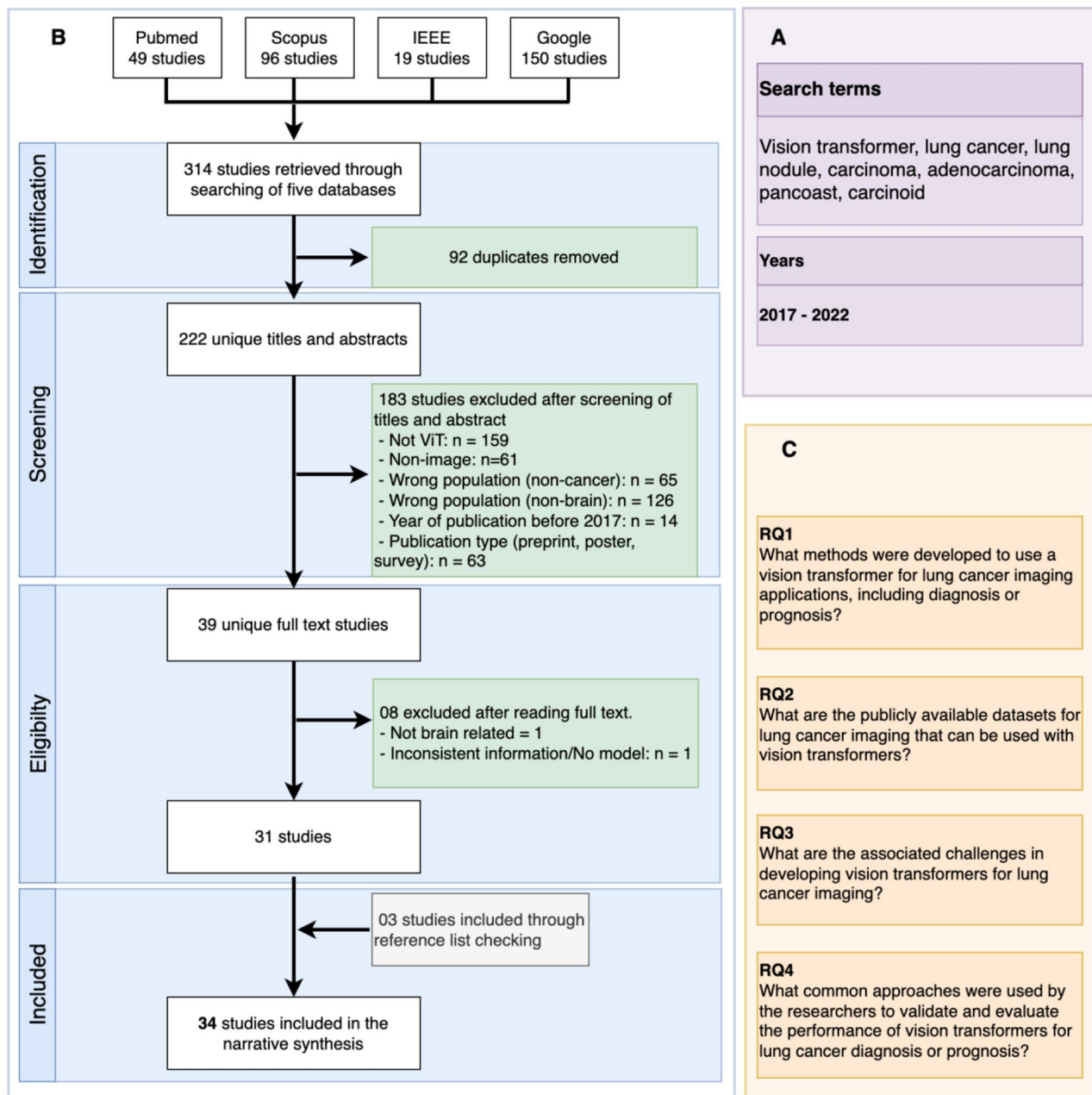


Fig. 1 (a) Search terms used. (b) The PRISMA-ScR flowchart for the selection of the included studies. (c) Research questions

Key implementation details

In the included studies, vision transformers were combined with CNNs, UNet, or graph networks. In the included studies, seven studies [25, 28, 34, 37, 51, 57] combined vision transformers with CNNs, three studies [36, 38, 52] used vision transformers in combination with UNet model, one study [41] combined both CNN and UNet with vision transformer, one study combined vision transformer with ResNet model. One study [24] combined the mask R-CNN model with a vision transformer to perform segmentation followed by classification. Two

studies [48, 55] explored the use of graph networks in combination with vision transformers. Six studies [24, 28, 32, 46, 49, 50] used SWIN transformer as their backbone transformer architecture.

Almost half of the studies (n=18) [24, 26–28, 30, 31, 33, 35–37, 41, 42, 45–49, 53] reported that their implementation was in Pytorch framework, while one study [25] reported the use of TensorFlow and Keras frameworks. The remaining studies did not specify the framework used.

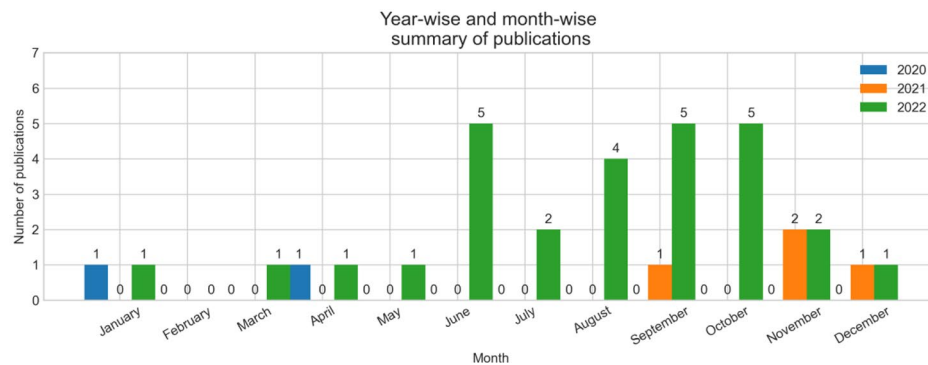


Fig. 2 Summary of year-wise and month-wise number of publications

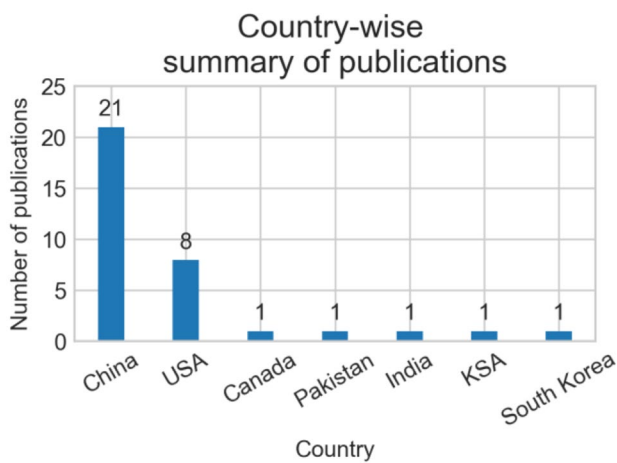


Fig. 3 Country-wise number of publications

Three studies [24, 26, 40] reported the use of a single Nvidia RTX 2080Ti GPU that usually comes with 11 GB memory, while one study [53] reported the use of four Nvidia RTX 2080Ti GPUs with 12 GB memory. Two studies [25, 36] reported the use of Nvidia P100 GPU, where the authors in [25] accessed the GPU via the Kaggle computational platform. Four studies [28, 39, 42, 57] reported the use of Nvidia V100 GPUs. Of these, one study [39] used four GPUs, one study [42] used two GPUs, and one study [57] used a single V100 GPU. Three studies [27, 33, 37] used a single Nvidia RTX3090 GPU, while one study [41] used a combination of two Nvidia RTX3090 GPUs. Three studies [45, 54, 55] used a single Nvidia GTX 1080 or 1080Ti GPU with 11 GB memory. One study [47] used Nvidia Titan-XP GPU. The largest number of GPUs usage was reported by [44], who used 48 Nvidia V100 GPUs. The remaining studies did not provide information on GPU usage.

Types of data used in the studies

In the included studies, 22 studies reported the publicly available use of data, six studies reported experiments

Table 2 Demographics of the included studies

| | | | |
|---------------------|--------------|-------------------|-------------------|
| Year | Year | Month | Number of studies |
| | 2022 (n =28) | January | 1 |
| | | March | 1 |
| | | April | 1 |
| | | May | 1 |
| | | June | 5 |
| | | July | 2 |
| | | August | 4 |
| | | September | 5 |
| | | October | 5 |
| | | November | 2 |
| | | December | 1 |
| | | 2021 (n =4) | September |
| | October | | 2 |
| | December | | 1 |
| | 2020 (n =2) | January | 1 |
| April | | 1 | |
| Countries | Country | Number of studies | |
| | China | 21 | |
| | USA | 8 | |
| | Canada | 1 | |
| | India | 1 | |
| | Saudi Arabia | 1 | |
| | South Korea | 1 | |
| | Pakistan | 1 | |
| Type of publication | Venue | Number of studies | |
| | Journal | 17 | |
| | Conference | 16 | |
| | Thesis | 1 | |

on privately collected data, and six studies used both public and private datasets. In the included studies, 23 developed models for 2D image data while 11 developed models for volumetric data. Nearly two-third ($n=21$) of the included studies used computed tomography (CT) scans of lung, while one-third ($n=11$) studies used histopathology or whole slide images of lungs. One study used PET, while another used CT and MRI scans. Table 3 summarizes the use of types of data in the included studies.

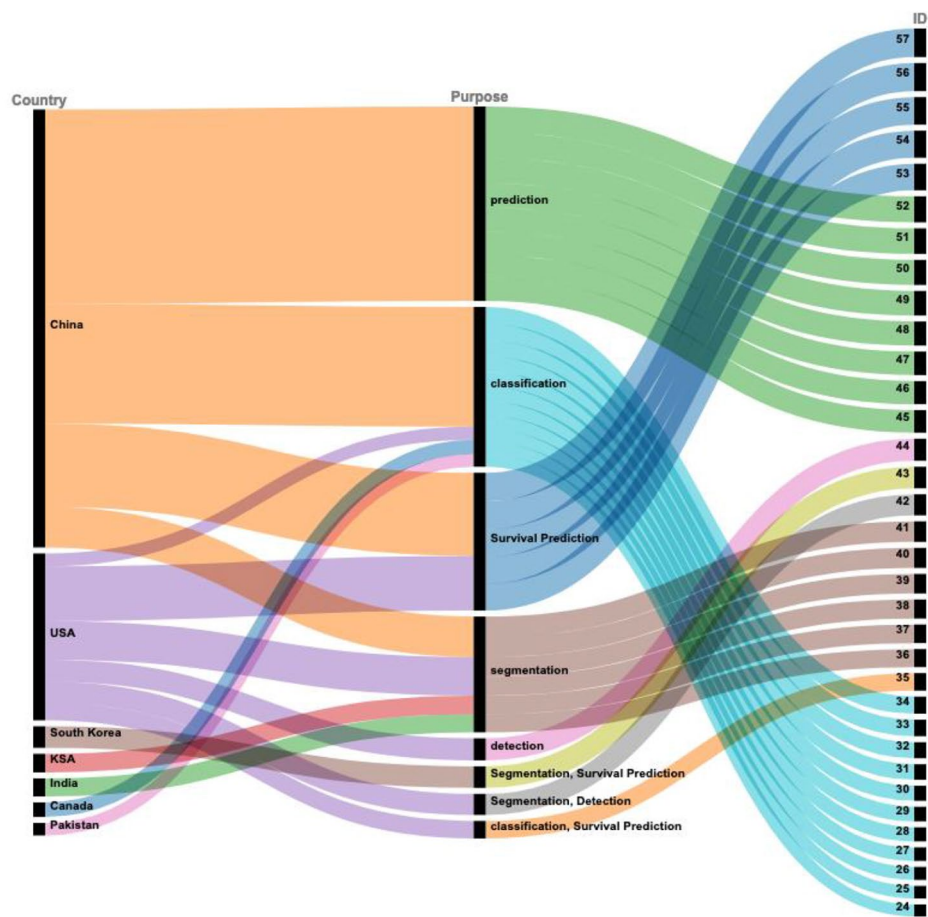


Fig. 4 Different tasks addressed in the included studies. The main tasks included classification of lung cancer types, segmentation of lungs, survival prediction for cancer patients, and prediction of course of the disease

| Table 3 Model parameter estimates with the entire student sample | | Used by |
|--|--------------------------------------|---|
| Availability of dataset | Public data (n = 22) | [25–28], [30–32], [34–40], [42, 44], [51–56] |
| | Private data (n = 6) | [24, 29, 33, 41, 46, 49] |
| | Public and private data (n = 6) | [43, 45, 47, 48, 50, 57] |
| Dimensionality of data | 2D models (2D data) (n = 23) | [24], [26–28], [30–32], [34–37], [40, 41], [47–50], [52–57] |
| | 3D models (volumetric data) (n = 11) | [25, 29, 33, 38, 39], [42–46], [51] |
| Modality of image data | CT (n = 21) | [29, 30, 32, 33, 33, 34], [36–42], [44–52] |
| | Histopathology (n = 11) | (n = 11) [24–28], [35], [53–57] |
| | PET (n = 1) | [43] |
| | CT and MRI (n = 1) | [39] |

Figure 5 shows the number of studies that used different modalities of data. Figure 6 shows the Venn diagram for the number of studies using public versus private data.

Datasets used in the studies
In the included studies, six studies [30–32, 40, 44, 52] the Lung Imaging Database Consortium (LIDC-IDRI) dataset, five studies [27, 28, 54–56] used The Cancer Genome Atlas (TCGA) datasets, four studies [34, 36–38] used the LUNA16 dataset. Table 4 summarizes the datasets used in the included studies along with the URL for the publicly available datasets.

Evaluation metrics
The most commonly used evaluation metrics in the included studies were accuracy and area under the ROC curve (AUC), each reported in 16 studies. Other popular metrics were specificity reported in 11 studies, sensitivity reported in nine studies, dice similarity coefficient reported in seven studies, and concordance index reported in six studies. Both precision and recall

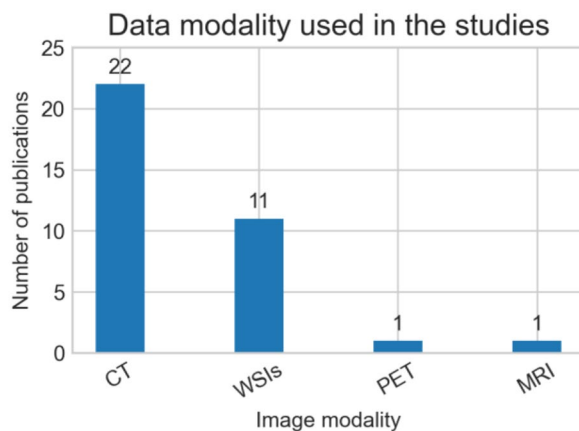


Fig. 5 Different data modalities used in the included studies

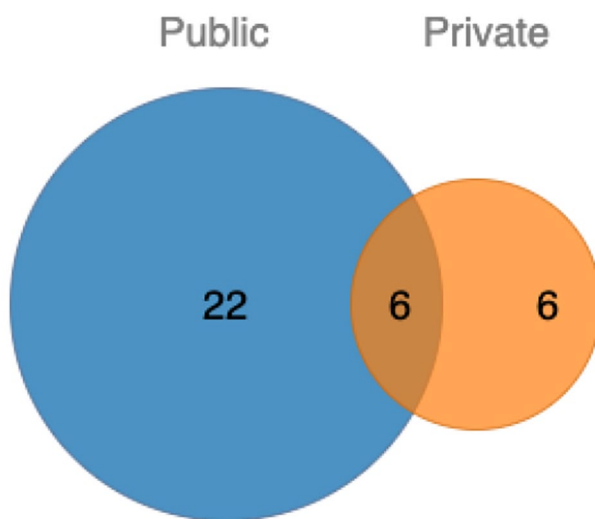


Fig. 6 Venn diagrams showing the contribution of public versus private datasets used in the included studies

measures were reported in five studies each. Other metrics were the F1 score, mean absolute error, mean absolute error, root mean square error, and Kappa score, each reported by one study. Figure 7 summarizes the number of studies using different evaluation metrics.

Almost one-third of studies ($n=11$) reported splitting the data into training, validation, and test sets, while five reported splitting the data into training and test sets only. Similarly, eight studies used a 5-fold cross-validation scheme, while six used a 10-fold cross-validation scheme to evaluate the performance of their methods.

In the included studies, only nine studies [26–28, 35, 37] provided a GitHub link for the implementation code.

Data preprocessing

In the included studies, only 19 out of 34 provided some information on the data preprocessing. Of these, ten

studies [26, 28, 30–32, 44–46, 56, 57] reported that they have done patch extraction. The patches are extracted with fixed sizes such as $96 \times 96 \times 96$ volumetric CT patches in [44] or 32×32 patches from 2D images in [30]. These patches are usually extracted with fixed sizes to help in reducing computational overload and load the samples in the memory. Six studies [25, 29, 32, 34, 37, 48] reported that they resized or reshaped the input image data into fixed dimensions such as 128×128 pixels in [25] or resized in the spatial domain such as $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ in [34] or $64 \text{ mm} \times 64 \text{ mm} \times 36 \text{ mm}$ in [48]. Three studies [31, 38, 42] reported that they applied normalization to the data, for example, by transforming the values in the $[0, 1]$ range. Five studies [30, 33, 34, 36, 38] reported different image augmentation techniques such as random rotation, random flipping, random affine, random shearing, zooming, horizontal and vertical flipping, and shifting, applied to the data before using the data for model training. In 15 studies [24, 27, 35, 39–41, 43, 47, 49–55], no details are provided on the data preprocessing.

Deep learning approaches such as unsupervised and self-supervised learning eliminate the need for manual feature engineering and feature selection. So, the included studies do not mention feature selection methods as they extract features using deep learning.

Discussion

Principle results

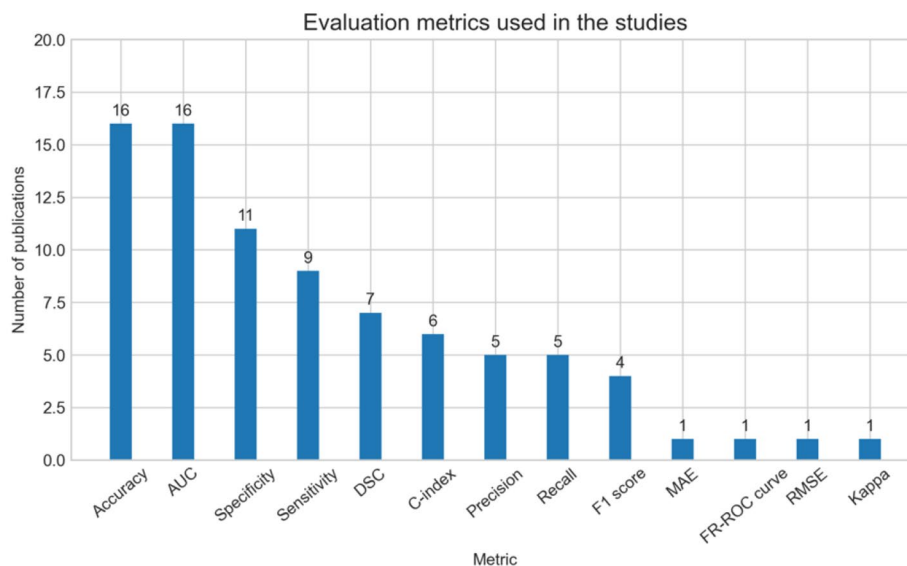
This study provides an overview of recent literature on the utilization of vision transformer-based artificial intelligence models for enhancing the diagnosis, prognosis, and classification of lung cancer. In the review, we did not find any studies before 2020. This is not surprising as vision transformers were proposed in 2017, and their use in medical imaging has recently gained popularity. Most of these studies were published in 2022, reflecting the growing interest in developing vision transformer-based approaches for lung cancer applications. However, the diversity of authors was limited, as researchers from China or the United States authored 85% of the studies.

The popularity of vision transformers for classification tasks has driven a majority of the studies reviewed in this work to employ them for classifying different types of lung cancers. The classification tasks included separating lung squamous cell carcinoma from lung adenocarcinoma, identifying benign versus malignant pulmonary nodules, and determining the invasiveness of lung adenocarcinomas. The studies also used vision transformers to predict lung cancer's severity or growth, thus aiding in survival predictions for patients. Some of the studies were limited to segmenting lung nodules.

Vision transformers effectively capture the long-range context in the input data, while CNNs tend to excel in

Table 4 Datasets used in the included studies

| Dataset name | URL | Used by |
|--|---|-----------------------|
| LC25000 | https://github.com/tampapath/lung_colon_image_set | [24] |
| National Lung Screening Trial (NLST) | https://www.cancer.gov/types/lung/research/nlst | [26, 47, 50] |
| Non-small cell lung cancer (NSCLC) | https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/studied-cancers/lung-adenocarcinoma | [27, 48] |
| Lung Squamous Cell Carcinoma (TCGA-LUSC) | https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=16056484 | [44, 54, 55] |
| Lung Adenocarcinoma (TCGA-LUAD) | https://www.cancerimagingarchive.net/collections/tcga-luad/ | [28, 53, 56] |
| Lung Imaging Database Consortium (LIDC-IDRI) | https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254 | [30–32], [40, 44, 52] |
| LUNA16 | https://luna16.grand-challenge.org/Data/ | [34, 36, 37, 44] |
| Tianchi Lung Nodule Detection dataset | https://tianchi.aliyun.com/competition/entrance/231601/introduction | [34] |
| Cbioportal | https://www.cbioportal.org/ | [35] |
| Medical Segmentation Decathlon | http://medicaldecathlon.com/ | [42] |
| LUNG1 | https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics | [43] |
| LUNGx | https://wiki.cancerimagingarchive.net/display/Public/SPIE-AAPM+Lung+CT+Challenge | [44] |
| NSCLC Radiogenomics | https://wiki.cancerimagingarchive.net/display/Public/NSCLC+Radiogenomics | [27, 39, 48] |
| NLSTt (derived from NLST) | https://github.com/liaw05/STMixer | [51] |
| Shanghai pulmonary hospital | Private | [33, 48] |
| Huadong Hospital dataset | Private | [45] |
| West China Hospital of Sichuan University | Private | [46] |
| Shanxi Provincial People's Hospital | Private | [47, 50] |
| CHCAMS | Private | [57] |

**Fig. 7** Evaluation metrics used in the included studies. DSC: dice similarity coefficient. MAE: mean absolute error. RMSE: root mean square error

capturing short-range dependencies. This is why many studies reviewed in this work combined vision transformers and CNNs, either through cascade or parallel connections or by incorporating vision transformer attention mechanisms into CNNs. Researchers integrated vision transformers with UNet or Mask R-CNN models for segmentation tasks. Due to its inherent benefits, the SWIN

transformer was frequently used as the backbone architecture for lung cancer imaging applications. The most popular framework for implementing vision transformer-based models in the studies was Pytorch, with all but one study (that used Tensorflow and Keras) reporting the use of Pytorch as the implementation framework.

Vision transformer-based models, in general, are computationally demanding. The computational demands of vision transformer models were addressed in the studies, with some utilizing multiple GPUs and one using a cluster of 48 GPUs, while others demonstrated that implementation on a single GPU was feasible.

Since the most commonly addressed task was the classification of lung cancers; hence, the studies reported accuracy and area under the ROC curve. The concordance index was a common evaluation metric for studies that reported survival prediction. In machine learning models, it is common to split the data in training, validation, and test sets (or training and test sets); however, some studies did not specify the evaluation mechanism and data split.

Practical and research implications

In developing AI models based on vision transformers, the availability of public datasets plays a crucial role. More than two-thirds of the studies utilized publicly available datasets for lung cancer imaging analysis. To encourage further growth in this field, it is imperative to have a rich resource of large-scale public datasets for lung cancer. In our review, the most commonly used imaging modality for lung cancer analysis was CT scans, followed by histopathology images. The use of PET and MRI was found to be less common. The Lung Imaging Database Consortium (LIDC) and The Cancer Genome Atlas (TCGA) offer extensive datasets for lung cancer (and other cancers) for researchers to utilize.

Despite the promising outcomes of vision transformer-based AI methods in analyzing lung cancer, they have limitations, such as the reliance on a significant amount of computational resources, including clusters of numerous GPUs, which may not be accessible to many research laboratories. Moreover, their practical implementation in a clinical setting remains unverified. Hence, there is a pressing need to advance toward developing computationally efficient training methods for vision transformers. The analyzed studies failed to furnish a comprehensive understanding of the interpretability of vision transformer-based models. This information is critical in applications such as predicting the survival of lung cancer patients, as it provides a deeper insight into the progression of the disease and the related risk factors. Additionally, the majority of studies (73%) did not provide access to their implementation code, hindering the ability of other researchers to reproduce the results or build upon the vision transformer-based models for lung cancer analysis. The absence of such links further reduces the reproducibility of the reported studies.

In our review, studies from China and USA dominated the literature where the healthcare tools are advanced, and thus, the new techniques can, in general,

be integrated with less effort. However, there is a lack of studies from developing countries. However, there is a scarcity of studies from developing countries. It is anticipated that increased contributions from these countries would aid in addressing the challenges of lung cancer in underdeveloped economies, where the disease is more prevalent due to socio-economic reasons.

The included studies greatly varied in how they reported the different datasets' usage or the number of images for training, validation, or test sets. For example, many studies reported the values of accuracy, sensitivity, specificity, or AUC, the number of samples in the test set varied between them, or the cross-validation strategy differed (or was even absent) in some of the studies. Accordingly, this review does not provide a quantitative summary of the results reported in the included studies for two reasons. First, the review aimed to identify the recent AI methods that used vision transformers for lung cancer imaging applications. Second, the review included many studies that vary in how they report quantitative metrics for the outcomes or how they organize their data; hence, establishing a direct summarization of the results is not practical. We believe that future systematic reviews should also cover the clinical relevance of the methods for lung cancer applications. This review did not find any implementation of vision transformers-based methods for mobile devices. Mobile devices will carry a significant role in transforming cancer care, and porting of highly accurate and effective strategies for cancer diagnosis and classification to mobile devices will open new dimensions in future digital healthcare by facilitating ease of use and accessibility. The included studies were inconsistent in reporting the training time required for the model. For example, the reviewers could not find this information in most of the studies or did not compare how the models would behave on different hardware and whether the training/inference time would see a major reduction. It is expected that providing web-based demos for the proposed models, in general, will increase the interest of doctors, physicians, and students in exploring the potential of vision transformers for lung cancer applications. However, this review did not find web-based platforms that used vision transformers for lung cancer applications.

Strengths and limitations

Strengths

With the recent popularity of vision transformer-based AI methods in medical imaging, there has been a growing interest in reviews on the topic [18, 58, 59, 22, 23]. However, we did not find any previous review on vision transformers for lung cancer imaging. This is the first review covering the classification, diagnosis, and

prognosis applications of vision transformers for lung cancer imaging.

In this review, we have summarized the key vision transformer-based methods for lung cancer applications that will help the readers and the researchers to identify the potential opportunities and related challenges in developing advanced methods for lung cancer analysis. In the review, we followed the guidelines of the PRISMA-ScR [60]. We included the most relevant studies from popular scientific databases that cover technical and healthcare literature. We overcame bias in study selection by adapting an independent selection mechanism of studies by two reviewers that a third reviewer validated. We identified the key areas and gaps in the vision transformer-based methods to which researchers may contribute. To the best of our knowledge, this is the first comprehensive review that explores the role of vision transformers in improving lung cancer classification and prognosis. Furthermore, it covers the most recent studies reported by the researchers. Hence, this review will be beneficial for researchers and practitioners interested in the transformation of digital healthcare, in general, and lung cancer, in particular.

Limitations

Since this review covered imaging-based applications only, clinical factors and living habits of lung cancer patients were not covered in the included studies, which would otherwise provide key information in the course of the disease. We did not evaluate the code as this was beyond the scope of this review. Since the included studies varied in terms of the datasets, or the number of samples/patients used, it was impossible to establish a direct comparison of their performances on the classification or prognosis of lung cancer. This review does not provide a discussion on the training delays due to two reasons. Firstly, such information was not provided in the included studies. Secondly, different research groups may vary greatly in their access to computational resources and GPUs. We understand that the interest in using and developing newer architectures of vision transformer-based AI methods for lung cancer imaging is growing rapidly. Hence, we cannot rule out the possibility that several other studies may come out while this work is being drafted, despite our best efforts to include the most recent studies until December 2022. This review covers studies published in English, so, relevant studies in other languages (if any) are not included.

Methods

In the review, we followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [60] guidelines to perform the study search and synthesis of the data.

Search strategy

An extensive search of scientific databases, including PubMed, Scopus, IEEE Xplore, Google Scholar, and MEDLINE (via PubMed), was conducted to identify the relevant studies. The study search was performed on December 21, 2022. Reference list checking was also performed for additional relevant studies. Only the first 150 relevant studies from Google Scholar were considered for the review, as search results beyond this number rapidly lost relevance and were not pertinent to the scoping review topic. The search terms were defined through consultation with domain experts and on the basis of the previous literature. The search terms included terms based on the target anatomy (e.g., lung cancer) and the intervention (e.g., transformers). The detailed search strings used in the study can be found in Appendix 2.

Search eligibility criteria

In this scoping review, we focused on exploring the recent advancements and applications of vision transformers in lung cancer medical imaging. We analyzed studies published until December 2022 in English that involved the utilization of vision transformers for various purposes related to lung cancer imaging, such as classification of lung cancer types, prediction of the cancer growth, detection of nodule, survival prediction of lung cancer patients, and segmentation of lungs. Studies that used any medical imaging modality such as MRI, CT, X-ray, and histopathology images were considered. Only original research published in peer-reviewed journals, conference proceedings, or book chapters was considered.

Studies that did not use vision transformers specifically but utilized other deep learning methods, such as CNNs and Generative Adversarial Networks (GANs), were excluded from the review. Additionally, studies that used transformers for non-imaging data, such as text data and electronic health records (EHRs), were excluded. Moreover, studies that used transformers for cancers other than lung cancer were also excluded. Studies identified as non-English text, review articles, preprints, editorials, proposals, conference abstracts, commentaries, and editor letters were also excluded. No restrictions were in place on the country of publications, the models' complexity, the reported methods' performance, and the modality of imaging data.

Study selection and data extraction

We used the Rayyan web-based review management tool [61] to conduct the initial screening and study selection process. One reviewer (H.A.) performed the literature search. After eliminating duplicates, two reviewers (F.M.) and (H.A.) independently screened the titles and abstracts of the studies to identify eligible studies. The studies that successfully passed the initial title and

abstract screening were selected for the full-text screening phase. Any disagreements during the process were resolved through discussion and through validation by a third reviewer (Z.S.). An evidence form was created and tested on three studies to establish a systematic and precise data extraction process (also see Appendix 3). Data extracted from the studies included the titles, first author's name, publication date and venue, the country of the first author's institution, the study application, the imaging type, the transformer type, the data source (public or private), data size, the validation methods, and the evaluation metrics. Additionally, information regarding the required hardware resources was also extracted. Moreover, the studies' challenges and suggested solutions were extracted, along with the challenges encountered and proposed solutions in the studies. Two reviewers (F.M. and H.Z.) conducted the data extraction, and any discrepancies were resolved through discussions and mutual consensus.

Data synthesis

We followed a narrative approach to synthesize the data extracted from the included studies. We categorized the data in terms of the specific tasks addressed in them, such as classification of lung cancer type, prediction of the course of cancer, survival prediction of the cancer patients, and segmentation of lungs. Based on the models developed in the included studies, we categorized them into those using 3D models and those using 2D models. We also cataloged the studies based on the use of public versus privately developed datasets, the method of validation of the results, and the reproducibility of the results.

Conclusion

In this work, we undertook a scoping review of 34 studies investigating the development and implementation of AI methods in lung cancer imaging, specifically using vision transformer models. Our review work indicates that vision transformer-based methods have been developed for the classification of lung cancer types and survival prediction of lung patients. Most reported methods have achieved performance propelling forward the field of AI for lung cancer imaging. The included studies evaluated the performance in terms of accuracy, the area under the ROC curve, and the concordance index. Additionally, we cataloged publicly available datasets for lung cancer imaging. Despite these advancements, we also identified areas for improvement, such as reducing model complexity, bridging the gap between clinical practice and vision transformer-based AI methods, and increasing geographical diversity in published studies.

Moreover, there is an urgent need to develop explainable vision transformer models for lung cancer imaging, as this will enhance the trust and acceptance of these methods among all stakeholders. We anticipate that our findings will provide a valuable reference text for researchers and students in the interdisciplinary fields of medical AI and cancer imaging. Vision transformers struggle to generalize well when data is limited. To improve vision transformers' generalization for lung cancer imaging, we advocate for the acquisition of larger and more diverse datasets of lung images with different modalities. While combining vision transformers with CNNs is common, a simple cascade arrangement might not effectively capture local and global features crucial for lung nodule detection and classification. Moreover, the parallel use of vision transformers and CNNs necessitates carefully filtering redundant information to overcome computational overhead. Future research must focus on developing pipelines that optimize the complementary performance of these architectures. Likewise, multimodal AI techniques have been proven effective in healthcare data; thus, vision transformers-based pipelines should be explored for processing lung cancer imaging data of multiple modalities, such as CT and PET. Lung nodules and large tumors are usually available in only few of the samples. So, data imbalance remains a challenge as most publicly available datasets come with a smaller number of large tumors. Thus, multi-institutional and multi-center collaborative efforts are needed to ensure large and diverse data of lung cancer that can help in better model generalization. Furthermore, research efforts should prioritize addressing the explainability and interpretability of vision transformers' performance in identifying tumor-related imaging components or discerning significant features influencing the model's prognosis behavior. Considering the resource-intensive nature of transformer architectures, we urge the development of resource-efficient implementation methods for vision transformers-based approaches. By doing so, we can advance toward clinical translation and real-time integration of vision transformers-based methods in lung cancer care.

Abbreviations

| | |
|------|----------------------------------|
| AI | Artificial Intelligence |
| AUC | Area under ROC curve |
| CNN | Convolutional Neural Networks |
| CT | Computed Tomography |
| DSC | Dice similarity coefficient |
| GPU | Graphics Processing Unit |
| LIDC | Lung Imaging Database Consortium |
| MAE | Mean absolute error |
| PET | Positron Emission Tomography |
| RMSE | Root mean square error |
| SOTA | State-of-the-art |
| TCGA | The Cancer Genome Atlas |
| WSI | Whole slide imaging |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12880-023-01098-z>.

Appendix 1: List of included studies

Appendix 2: Search strings

Appendix 3: Data extraction form

Acknowledgements

Not applicable.

Authors' contributions

H. A. contributed to the conception, design, literature search, data selection, data synthesis, data extraction, and drafting. F. M. contributed to the data synthesis, data extraction, and drafting. Z. S. contributed to the drafting and critical revision of the manuscript. All authors gave their final approval and accepted accountability for all aspects of the work.

Funding

Open Access funding provided by Qatar National Library.

Data Availability

All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

Received: 29 March 2023 / Accepted: 5 September 2023

Published online: 15 September 2023

References

- Cruz CSD, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. *Clin Chest Med*. 2011;32(4):605–44.
- de Groot PM, Wu CC, Carter BW, Munden RF. The epidemiology of lung cancer. *Transl Lung Cancer Res*. 2018;7(3):220.
- Lewis SJ, Gandomkar Z, Brennan PC. Artificial Intelligence in medical imaging practice: looking to the future. *J Med Radiat Sci*. 2019;66(4):292–5.
- Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31–8.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6):84–90.
- Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016.
- Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Für Med Phys*. 2019;29(2):102–27.
- Ali H, et al. The role of generative adversarial networks in brain MRI: a scoping review. *Insights Imaging*. 2022;13(1):1–15.
- Würfl T, Ghesu FC, Christlein V, Maier A. "Deep learning computed tomography," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part III 19*, Springer, 2016, pp. 432–440.
- Min JK, Kwak MS, Cha JM. Overview of deep learning in gastrointestinal endoscopy. *Gut Liver*. 2019;13(4):388.
- Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*. 2017;284(2):574–82.
- Iqbal T, Ali H. Generative adversarial network for medical images (MI-GAN). *J Med Syst*. 2018;42:1–11.
- Dosovitskiy A et al. "An image is worth 16x16 words: Transformers for image recognition at scale," *ArXiv Prepr. ArXiv201011929*, 2020.
- Zheng S et al. "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 6881–6890.
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. "End-to-end object detection with transformers," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, Springer, 2020, pp. 213–229.
- Gao X et al. "COVID-VIT: Classification of Covid-19 from 3D CT chest images based on vision transformer model," in *2022 3rd International Conference on Next Generation Computing Applications (NextComp)*, IEEE, 2022, pp. 1–4.
- Watanabe S, Ueno T, Kimura Y, Mishina M, Sugimoto N. Generative image transformer (GIT): unsupervised continuous image generative and transformable model for [123 I] FP-CIT SPECT images. *Ann Nucl Med*. 2021;35:1203–13.
- He K, et al. Transformers in medical image analysis: a review. *Intell Med*. 2022;3(1):59–78. <https://doi.org/10.1016/j.imed.2022.07.002>.
- Chassagnon G et al. Artificial intelligence in lung cancer: current applications and perspectives. *Jpn J Radiol*, pp. 1–10, 2022.
- Kaur C, Garg U. "Artificial intelligence techniques for cancer detection in medical image processing: A review," *Mater. Today Proc*, 2021.
- Wang S, et al. Artificial intelligence in lung cancer pathology image analysis. *Cancers*. 2019;11(11):1673.
- Xia K, Wang J. Recent advances of Transformers in medical image analysis: a comprehensive review. *MedComm—Future Med*. 2023;2(1):e38. <https://doi.org/10.1002/mef2.38>.
- Li Y, Wu X, Yang P, Jiang G, Luo Y. Machine learning for Lung Cancer diagnosis, treatment, and prognosis. *Genomics Proteom Bioinf*. 2022;20(5):850–66.
- Chen Y, Feng J, Liu J, Pang B, Cao D, Li C. Detection and classification of Lung Cancer cells using swin transformer. *J Cancer Ther*. 2022;13(7):464–75.
- Aitazaz T, Tubaishat A, Al-Obeidat F, Shah B, Zia T, Tariq A. Transfer learning for histopathology images: an empirical study. *Neural Comput Appl*. 2022. <https://doi.org/10.1007/s00521-022-07516-7>.
- Zheng Y, et al. A graph-transformer for whole slide image classification. *IEEE Trans Med Imaging*. 2022;41(11):3003–15. <https://doi.org/10.1109/TMI.2022.3176598>.
- Shao Z, Bian H, Chen Y, Wang Y, Zhang J, Ji X. "Transmil: Transformer based correlated multiple instance learning for whole slide image classification," in *Advances in neural information processing systems*, 2021, pp. 2136–2147.
- Wang X, et al. Transformer-based unsupervised contrastive learning for histopathological image classification. *Med Image Anal*. 2022;81. <https://doi.org/10.1016/j.media.2022.102559>.
- Heidarian S. "Capsule Network-based COVID-19 Diagnosis and Transformer-based Lung Cancer Invasiveness Prediction via Computerized Tomography (CT) Images," Doctoral thesis, Concordia University, 2022.
- Liu D, Liu F, Tie Y, Qi L, Wang F. Res-trans networks for lung nodule classification. *Int J Comput Assist Radiol Surg*. 2022;17(6):1059–68. <https://doi.org/10.1007/s11548-022-02576-5>.
- Wang R, Zhang Y, Yang J. *TransPND: A Transformer Based Pulmonary Nodule Diagnosis Method on CT Image*, vol. 13535. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13535. 2022. https://doi.org/10.1007/978-3-031-18910-4_29.
- Wu P, Chen J, Wu Y. "Swin Transformer based benign and malignant pulmonary nodule classification," in *Proceedings of SPIE - The International Society for Optical Engineering*, 2022. <https://doi.org/10.1117/12.2656809>.
- Xiong Y, Du B, Xu Y, Deng J, She Y, Chen C. "Pulmonary Nodule Classification with Multi-View Convolutional Vision Transformer," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–7. <https://doi.org/10.1109/IJCNN55064.2022.9892716>.
- Yang J, Deng H, Huang X, Ni B, Xu Y. "Relational Learning Between Multiple Pulmonary Nodules via Deep Set Attention Transformers," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 1875–1878. <https://doi.org/10.1109/ISBI45749.2020.9098722>.
- Chen RJ et al. "Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning," in *2022 IEEE/CVF Conference on Computer Vision and*

- Pattern Recognition (CVPR), New Orleans, LA, USA: IEEE, 2022, pp. 16144–16155. <https://doi.org/10.1109/CVPR52688.2022.01567>.
36. Dhamija T, Gupta A, Gupta S, Anjum R, Katarya, Singh G. Semantic segmentation in medical images through transfused convolution and transformer networks, *Appl Intell*, 2022.
 37. Alahmadi MD. Medical image segmentation with learning semantic and global contextual representation. *Diagnostics*. 2022;12(7). <https://doi.org/10.3390/diagnostics12071548>.
 38. Guo D, Terzopoulos D. "A Transformer-Based Network for Anisotropic 3D Medical Image Segmentation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 8857–8861. <https://doi.org/10.1109/ICPR48806.2021.9411990>.
 39. Jiang J, Tyagi N, Tringale K, Crane C, Veeraraghavan H. *Self-supervised 3D Anatomy Segmentation Using Self-distilled Masked Image Transformer (SMIT)*, vol. 13434. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13434. 2022. https://doi.org/10.1007/978-3-031-16440-8_53.
 40. Wang S, Jiang A, Li X, Qiu Y, Li M, Li F. DPBET: a dual-path lung nodules segmentation model based on boundary enhancement and hybrid transformer. *Comput Biol Med*. 2022;151:106330. <https://doi.org/10.1016/j.combiomed.2022.106330>.
 41. Xie H, Chen Z, Deng J, Zhang J, Duan H, Li Q. "Automatic segmentation of the gross target volume in radiotherapy for lung cancer using transesSEUnet 2.5 D Network," *J. Transl. Med*, vol. 20, no. 1, pp. 1–12, Nov. 2022.
 42. Yang D, Myronenko A, Wang X, Xu Z, Roth HR, Xu D. "T-AutoML: Automated machine learning for lesion segmentation using transformers in 3d medical imaging," presented at the Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 3962–3974.
 43. Dao D-P et al. "Survival Analysis based on Lung Tumor Segmentation using Global Context-aware Transformer in Multimodality," in *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 5162–5169. <https://doi.org/10.1109/ICPR56361.2022.9956406>.
 44. Niu C, Wang G. Unsupervised contrastive learning based transformer for lung nodule detection. *Phys Med Biol*. 2022;67(20). <https://doi.org/10.1088/1361-6560/ac92ba>.
 45. Zhao W, et al. GMILT: a Novel Transformer Network that can noninvasively predict EGFR Mutation Status. *IEEE Trans Neural Netw Learn Syst*. 2022;1–15. <https://doi.org/10.1109/TNNLS.2022.3190671>.
 46. Shao J, et al. Radiogenomic System for non-invasive identification of multiple actionable mutations and PD-L1 expression in Non-Small Cell Lung Cancer based on CT images. *Cancers*. 2022;14(19). <https://doi.org/10.3390/cancers14194823>.
 47. Wang H, et al. Static–dynamic coordinated transformer for Tumor Longitudinal Growth Prediction. *Comput Biol Med*. 2022;148. <https://doi.org/10.1016/j.combiomed.2022.105922>.
 48. Lian J, et al. Early stage NSCLC patients' prognostic prediction with multi-information using transformer and graph neural network model. *eLife*. 2022;11. <https://doi.org/10.7554/eLife.80547>.
 49. Ma X, Xia L, Chen J, Wan W, Zhou W. Development and validation of a deep learning signature for predicting lymph node metastasis in lung adenocarcinoma: comparison with radiomics signature and clinical-semantic model. *Eur Radiol*. 2022. <https://doi.org/10.1007/s00330-022-09153-z>.
 50. Song P et al. "MSTS-Net: malignancy evolution prediction of pulmonary nodules from longitudinal CT images via multi-task spatial-temporal self-attention network," *Int. J. Comput. Assist. Radiol. Surg*, pp. 1–9, Nov. 2022.
 51. Fang J et al. "Siamese Encoder-based Spatial-Temporal Mixer for Growth Trend Prediction of Lung Nodules on CT Scans," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Singapore: Springer, Sep. 2022, pp. 484–494.
 52. Wang H, Zhu H, Ding L. "Accurate Classification of Lung Nodules on CT Image Based on TransUnet," *Front. Public Health*, p. 4664, Dec. 2022.
 53. Chen RJ et al. "Multimodal Co-Attention Transformer for Survival Prediction in Gigapixel Whole Slide Images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3995–4005.
 54. Huang Z, Chai H, Wang R, Wang H, Yang Y, Wu H. "Integration of patch features through self-supervised learning and transformer for survival analysis on whole slide images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Strasbourg, France: Springer, 2021, pp. 561–570.
 55. Wang R, Huang Z, Wang H, Wu H, Biomedicine. (BIBM), IEEE, 2021, pp. 757–760.
 56. Li C, Zhu X, Yao J, Huang J. "Hierarchical Transformer for Survival Prediction Using Multimodality Whole Slide Images and Genomics," in *26th International Conference on Pattern Recognition (ICPR)*, Montreal, QC, Canada: IEEE, 2022, pp. 4256–4262.
 57. Shen Y et al. "Explainable Survival Analysis with Convolution-Involved Vision Transformer," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, pp. 2207–2215.
 58. Shamshad F et al. Transformers in medical imaging: a survey, *ArXiv Prepr ArXiv220109873*, 2022.
 59. Akinyelu AA, Zaccagna F, Grist JT, Castelli M, Rundo L. Brain tumor diagnosis using machine learning, convolutional neural networks, Capsule neural networks and Vision Transformers, Applied to MRI: a Survey. *J Imaging*. 2022;8(8):205.
 60. Tricco AC, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467–73.
 61. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5:1–10.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.