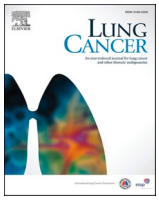
Lung Cancer 186 (2023) 107413

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan



Artificial intelligence-based recurrence prediction outperforms classical histopathological methods in pulmonary adenocarcinoma biopsies

F. Akram [a,1], J.L. Wolf [a,b,1], T.E. Trandafir [a], Anne-Marie C. Dingemans [c], A.P. Stubbs [a,2], J.H. von der Thüsen [a,2,*]

[a] Department of Pathology and Clinical Bioinformatics, Erasmus Medical Center, Rotterdam, The Netherlands
[b] Institute of Tissue Medicine and Pathology, University of Bern, Bern, Switzerland
[c] Department of Pulmonary Diseases, Erasmus MC Cancer Center, University Medical Center, Rotterdam, The Netherlands

ARTICLE INFO

Keywords:
Lung adenocarcinoma
Artificial intelligence
Convolutional neural network
Recurrence prediction

ABSTRACT

Introduction: Between 10 and 50% of early-stage lung adenocarcinoma patients experience local or distant recurrence. Histological parameters such as a solid or micropapillary growth pattern are well-described risk factors for recurrence. However, not every patient presenting with such a pattern will develop recurrence. Designing a model which can more accurately predict recurrence on small biopsy samples can aid the stratification of patients for surgery, (neo-)adjuvant therapy, and follow-up.
Material and Methods: In this study, a statistical model on biopsies fed with histological data from early and advanced-stage lung adenocarcinomas was developed to predict recurrence after surgical resection. Additionally, a convolutional neural network (CNN)-based artificial intelligence (AI) classification model, named AI-based Lung Adenocarcinoma Recurrence Predictor (AILARP), was trained to predict recurrence, with an ImageNet pre-trained EfficientNet that was fine-tuned on lung adenocarcinoma biopsies using transfer learning. Both models were validated using the same biopsy dataset to ensure that an accurate comparison was demonstrated.
Results: The statistical model had an accuracy of 0.49 for all patients when using histology data only. The AI classification model yielded a test accuracy of 0.70 and 0.82 and an area under the curve (AUC) of 0.74 and 0.87 on patch-wise and patient-wise hematoxylin and eosin (H&E) stained whole slide images (WSIs), respectively.
Conclusion: AI classification outperformed the traditional clinical approach for recurrence prediction on biopsies by a fair margin. The AI classifier may stratify patients according to their recurrence risk, based only on small biopsies. This model warrants validation in a larger lung biopsy cohort.

1. Introduction

Lung adenocarcinoma (LUAD) accounts for more than 50 % of all lung cancer subtypes and is nowadays more often diagnosed at an early stage due to advances in screening methods [1,2]. Despite curative-intent therapy by either surgery or stereotactic ablative radiotherapy (SABR), 10–50 % of these patients experience recurrence during follow-up [3–7].

In 2011, the International Association for the Study of Lung Cancer (IASLC), the American Thoracic Society (ATS) and the European Respiratory Society (ERS) developed a novel histological classification of LUADs. This classification consists of five different growth patterns defined as lepidic, acinar, papillary, micropapillary and solid [8]. Based on these patterns, LUAD can further be separated into low-grade (lepidic dominant), intermediate-grade (acinar and papillary predominant) and high-grade (solid and micropapillary predominant) classes [9]. The distinction of these five different histological subtypes is important since they correlate with patient survival. This has been demonstrated and validated in numerous studies [10–13]. In addition, several studies were able to link LUAD-specific recurrence to the presence of high-grade patterns for both stereotactic ablative radiotherapy (SABR) [7] and surgery [5,14]. Even when present in small proportions, the high-grade pattern can influence prognosis [15]. Furthermore, early-stage LUAD with a high-grade pattern can present different time-dependent and

* Corresponding author.
E-mail address: j.vonderthusen@erasmusmc.nl (J.H. von der Thüsen).
[1] shared first authorship.
[2] shared last authorship.

https://doi.org/10.1016/j.lungcan.2023.107413
Received 23 July 2023; Received in revised form 22 October 2023; Accepted 25 October 2023
Available online 4 November 2023

side-specific recurrence patterns [6,16]. Beyond histological and clinical criteria, recurrence prediction has also been evaluated using genetic data, albeit with contradictory results [17–20].

Most guidelines recommend postoperative surveillance of patients with non-small cell lung cancer (NSCLC) regardless of the clinical stage [21–23]. These recommendations incorporate medical history, physical examination and routine radiologic examinations at various time intervals. Histological characteristics such as the presence of high-grade growth patterns and time-dependent and site-specific observations of recurrences in early-stage LUAD could stratify patients for individualized follow-up and treatment, such as (neo-)adjuvant systemic therapy [24,25], to prevent an early recurrence.

Whilst most studies used the Cox proportional hazards (PH) model to predict the clinical outcome, this linear model has some limitations compared to nonlinear analyses which might better reflect real-world clinical data. To overcome this problem, numerous studies have tested deep learning models to predict recurrence, recurrence-free survival (RFS) and overall survival (OS) in NSCLC, using clinical, radiological and pathological variables as input for both patients treated with radiation therapy [26] and by surgery [27,28]. However, no stratification between the different tumor types was done, and only Lee et al. performed subgroup analysis on stage IA and IB NSCLC [27]. Jones et al. developed a genomic-pathological risk model for early stage LUAD patients in which they demonstrated that the combination of molecular and clinicopathologic features improved risk stratification and prediction of recurrence compared to traditional Tumor Node Metastasis (TNM) classifications [29]. However, it remained unclear whether biopsy or resection specimens were incorporated into the model. Furthermore, all the above-named clinical-pathological studies, except the one by Leeman et al., were performed on resected early-stage LUADs.

Most studies using a histology-based deep learning approach focused on WSI-level cancer type classification in resection specimens [30–35]. The publication by Shim et al. represents the only known study in which a DeepRePath model was developed and trained based on a deep convolutional neural network (CNN) using multi-scale pathology images to predict the prognosis of patients with early-stage LUAD [36]. To the best of our knowledge no LUAD recurrence prediction studies have been published that incorporate image-based features and deep learning to predict recurrence on preoperative biopsy specimens. This is of interest since risk stratification might influence not only the follow-up but also the choice of therapy. Thus, high-risk patients may benefit from a combination of a more radical surgical approach with extended lymph node dissection together with a closer follow-up to prevent and detect early recurrences, compared to low-risk patients. Furthermore, recurrence prediction on biopsy specimens is indispensable in cases of SABR or pre-operative neoadjuvant therapy in which a fully assessable resection specimen is not available.

This study introduces AI-based Lung Adenocarcinoma Recurrence Predictor (AILARP), an AI-powered classifier based on CNNs, designed to predict recurrence in Lung Adenocarcinoma (LUAD) within a 5-year timeframe using H&E WSIs from biopsies. AILARP leverages data extracted from image patches sourced from tumor regions in LUAD biopsy samples, focusing on color, spatial, and texture features. The primary goal of AILARP is to predict recurrence using pre-operative biopsy samples, providing crucial information for treatment decisions and follow-up planning. We also conducted a comparative analysis, contrasting AILARP with a traditional statistical approach. This traditional method relies on histological parameters such as growth patterns, nuclear grade, fibrosis, and inflammation, employing regression techniques to predict recurrence outcomes. Fig. 1 visually demonstrates the usage of biopsy samples in both approaches, facilitating a direct comparison between the two models.

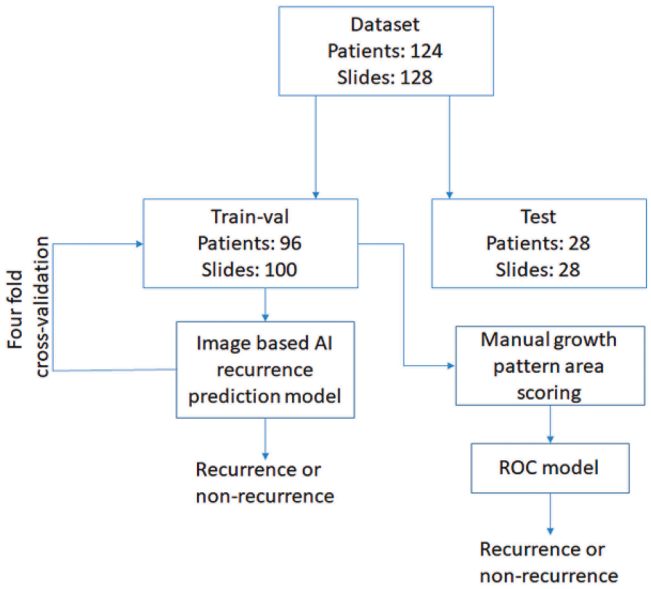


Fig. 1. Data distribution and experimental design for both models.

2. Materials and methods

2.1. Growth pattern analysis and recurrence prediction

A cohort of 128 histology specimens of 124 patients who underwent diagnostic small lung biopsies was selected following the criteria from our previous publication (Wolf. et al [37]) and independently reviewed by two pathologists (JvT and JW), who were blinded to the diagnosis as well as the clinical data (institutional ethical approval obtained under MEC-2020–0732). The biopsies were reviewed according to the IASLC/ATS/ERS classification using 5 % increments for recording the different histological subtypes. Tumors were divided into lepidic, acinar, papillary, micropapillary and solid predominant, where the predominant pattern was defined as the pattern with the largest percentage. This scoring method was defined as the *continuous growth pattern*. For the d*ominant growth pattern*, the dominant pattern was set 1 when present, the other patterns were set 0 and for the *cumulative growth pattern* the acinar and papillary pattern as well as the solid and micropapillary patterns were fused to intermediate and high-grade pattern respectively and then set 1 when present as dominant and 0 when absent [8]. In addition to determining the predominant growth pattern, a semi quantitative analysis of the nuclear grade, desmoplastic stroma reaction and inflammation within the tumor was made as reported earlier [37]. For all patients, clinical data concerning the medical follow-up from the patients' medical records were also collected. This included date of histopathological diagnosis, follow-up time, survival data (date of death) and duration before a recurrence event. Since we were interested in a clinically relevant recurrence outcome, we divided the group into two classes, based on the absence or occurrence of a recurrence event recorded after therapy.

The statistical model is based on a Cox proportional-hazards model which is a regression model to investigate the association between the recurrence time of patients and a set of predictor variables. To assess the predictive performance of the histological data, receiver operating curve (ROC) analyses were performed based on the predictions provided by the Cox model. The whole dataset was randomized and divided into a training (n = 98) and validation set (n = 28). To find the best-performing model, various combinations of the different growth patterns together with additional histological parameters such as nuclear grading, fibrosis and inflammation were used. An overview of the parameter combinations used for training is provided in supplementary table S1. The "ROCR" package (R-version 4.0.5) was used for analysis.

This analysis was compared to the recurrence prediction of the deep learning classification model.

2.2. AI-based recurrence prediction

In this section, we propose a deep learning framework for recurrence prediction, named AILARP, that uses color-, spatial and texture-based features from the image patches taken from LUAD biopsy specimens. To build the deep learning framework, we first split the patient data belonging to the two classes (i.e., a recurrence event recorded after therapy, or no recurrence recorded) into training, validation, and test samples to prevent data leakage. The test samples remained untouched during model training and were solely used for evaluation. In the data split, the training set contained 75 patients (21 recurred and 54 non-recurred), the validation set included 25 patients (7 recurred and 18 non-recurred), and the test set comprised 28 patients (7 recurred and 21 non-recurred); see supplementary table S2. The selected patients within the test set were identical to the validation set in the statistical model.

In this work, we used a deep learning convolutional neural network (CNN) approach to design a classification model for the recurrence prediction of LUAD using the H&E WSI biopsy images. First, small patches of size 512 × 512 pixels at a magnification of 20× (0.46 um/pixel) were extracted from the WSI images to train a CNN-based classification model for the recurrence prediction of LUAD. An extracted image patch was discarded when no cancer region was found in addition to surrounding tissue. In CNN-based classification models, feature extraction consists of several convolutional and pooling layer pairs. Each convolutional layer is a collection of digital filters to perform the convolution operation on the input data. In addition, the pooling layer is used as a dimensionality reduction layer and decides the threshold. However, traditional CNN-based classification models have a higher number of parameters, making models computationally expensive and difficult to train while having less data. To solve the computational complexity problem, we used an EfficientNet model [38] that uniformly

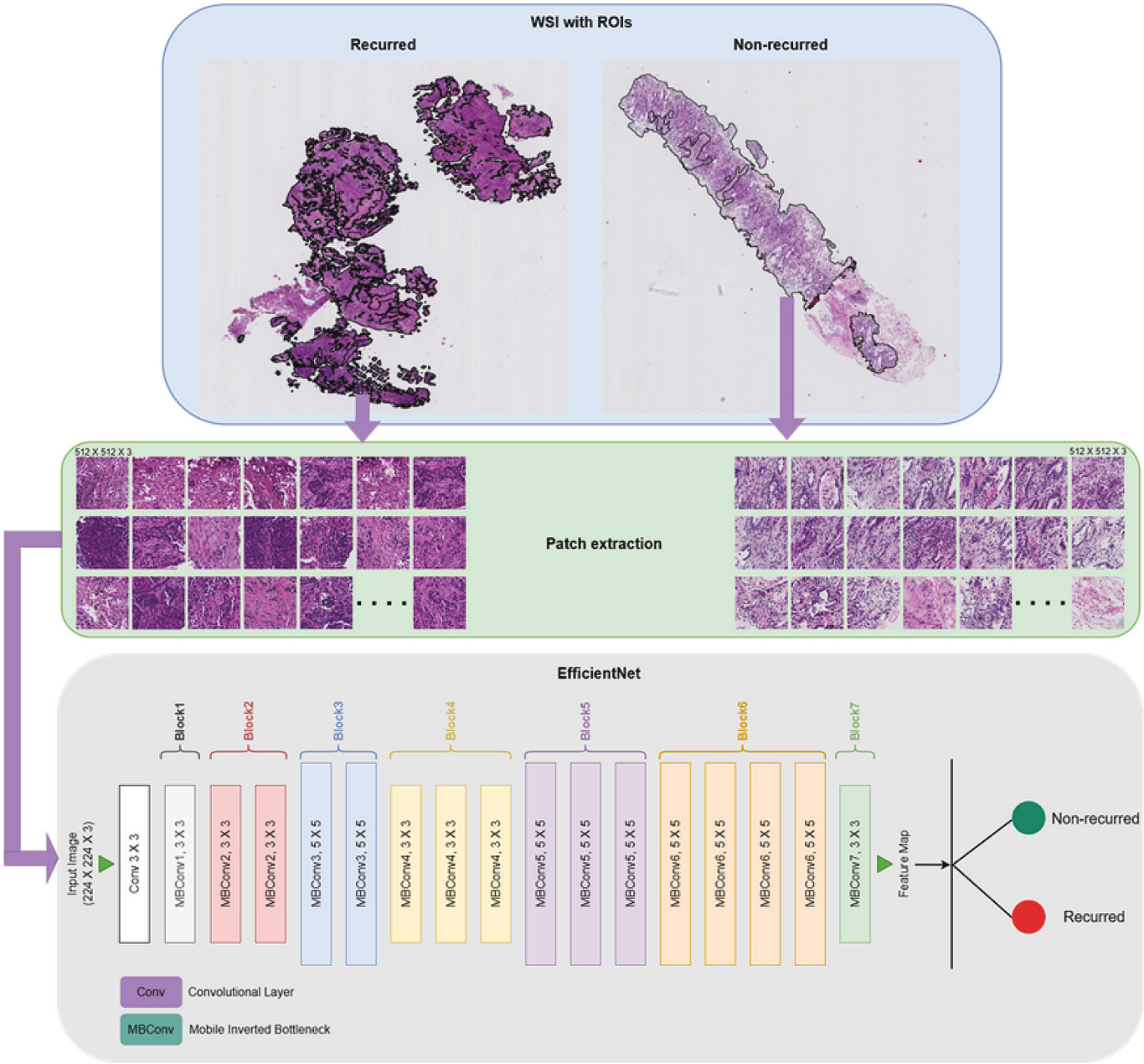


Fig. 2. Proposed AI framework for LUAD recurrence prediction.

scales all depth/width/resolution dimensions using a *compound coefficient*. We used the classical EfficientNetV2-B0 network that is based on the inverted bottleneck residual blocks of MobileNetV2 [35] and squeeze-and-excitation blocks. In the model design, to enrich the feature space and prevent overtraining, we incorporate transfer learning into the pipeline [38]. Transfer learning is used to improve a learner from one domain by transferring information from a related domain [30].

In the experimental design a transfer learning approach is utilized that involves training an EfficientNetv2-b0 model [39] on image tiles resized to dimensions of 224 × 224 pixels. To enable this training, the original input images, initially sized at 512x512 pixels, are adjusted to the model's specified input size of 224x224 pixels. This resizing operation is imperative due to the model's specific requirements for input size. Fig. 2 illustrates the framework design of our proposed LUADRecAI method.

MATLAB was used for image data pre-processing, Python environment and Pytorch library were used in the model design. The experiments were performed in a four-fold cross-validation fashion. The data was split patient-wise to ensure there is no data leakage during the training process and test data was not mixed in the training and validation set.

It is important to emphasize that for both methods, namely the statistical model and the AILARP model, the same test set is utilized to evaluate the performance of both models. This consistency ensures that both models are assessed using identical data, enabling a fair and direct comparison of their respective performances. Fig. 1 illustrates the process of splitting patient-level WSIs into training and test sets, which were then utilized for both experiments.

3. Results

3.1. Clinical dataset

A cohort of 128H&E WSI biopsy specimens of 124 patients was reviewed, of which 35 patients experienced recurrence and 89 were recurrence free. The median follow-up time was 1.98 years (or 724 days), ranging between 0.05 and 9.981 years. An early loss to follow-up was caused by transfer or death of the patient, the latter due to reasons other than lung cancer. To gain a more comprehensive understanding of the dataset, we have provided a summary of the clinical characteristics in supplementary table S3. Notably, the patients have been categorized into two groups: those who experienced recurrences within a 5-year period and those who did not.

3.2. Growth pattern analysis and recurrence prediction

ROC analysis of histological parameters, containing growth pattern, nuclear grading, fibrosis, and inflammation revealed a maximal area under the curve (AUC) of 0.49 (see Fig. 3). Different combinations of growth patterns were tested together with the above listed histological parameters. Two model combinations showed an identical AUC: selecting the dominant growth pattern and selecting the dominant and the worst growth pattern with a cut-off of 20 % respectively. An overview of the different models tested is provided in supplementary table S4.

3.3. AI-based recurrence prediction

The experiments were performed in a four-fold cross-validation fashion, which yields outcomes as shown in Table 1 and supplementary table S5. The upper half of supplementary table S5 shows that while training the model using the training set and tuning its parameters on the validation set, fold1 yields the best patch-wise classification validation accuracy of 0.77, whereas fold3 provides the best validation precision, recall, and F1-score of 0.70, 0.72, and 0.71, respectively. The upper half of Table 1 shows that after training and tuning the AI models

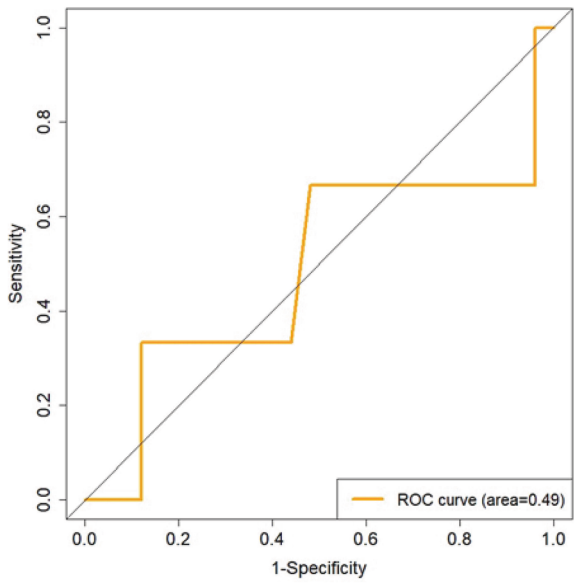


Fig. 3. An ROC curve of the statistical model on test set using dominant growth pattern and the dominant together with the worst growth pattern with a cut-off of 20% in predicting recurrence outcome.

Table 1
Test results using patch-wise (top half) and patient-wise (bottom half) four-fold cross-validation.

	Fold split	Accuracy	Precision	Recall	F1 score
Patch-wise	1	**0.70**	**0.66**	**0.59**	**0.62**
	2	0.67	0.60	0.58	0.59
	3	0.68	0.60	0.57	0.58
	4	0.68	0.61	0.55	0.58
Patient-wise	1	**0.82**	**0.90**	**0.64**	**0.75**
	2	0.82	0.90	0.64	0.75
	3	0.79	0.89	0.57	0.69
	4	0.79	0.89	0.57	0.69

on the training and validation sets in a four-fold fashion, on the first fold, the proposed deep learning framework provided the best patch-wise classification test accuracy, precision, recall, and F1-score of 0.70, 0.66, 0.59, and 0.62, respectively, compared to the rest of the folds.

After obtaining the patch-wise results, we translated the results back to WSI (patient-wise). For example, let x be an image input, $pred_c(x)$ the correctly predicted outcome on the image patches, $labels(x)$ the actual ground truth for that outcome, M the correctly predicted patches and N the number of patches per WSI. Then the patient's correct prediction distribution $patient_{dist}$ can be computed using (1). Using the correct prediction distribution for a WSI, the patient-wise prediction $pred$ can be computed using (2), where 0.5 is the threshold used to generate the binary outcome for the patient-wise prediction.

$$patient_{dist} = \frac{\sum_{i=1}^{M} pred_c(x)}{\sum_{j=1}^{N} labels(x)} \tag{1}$$

$$pred = \begin{cases} 1, patient_{dist} > 0.5 \\ 0, patient_{dist} \leq 0.5, \end{cases} \tag{2}$$

The bottom half of supplementary table S5 shows patient-wise validation results on the used four-fold cross-validation setup. It shows that the third fold provides the best accuracy, precision, recall, and F1-score of 0.80, 0.89, 0.64, and 0.74 compared to the other folds. Although the proposed deep learning framework yielded the best patch-wise accuracy on the first fold, it provided the best patient-wise accuracy on the third fold because images used in the first fold had more similar features

found in the test set than in the third fold. The bottom half of Table 1 shows patient-wise test results on the used four-fold cross-validation setup. It shows that the first fold provides the best accuracy, precision, recall, and F1-score of 0.82, 0.90, 0.64, and 0.75, respectively, compared to the rest of the folds.

Fig. 4A and 4B show ROCs of the first fold for validation and test sets computed using patch-wise image outcomes from the H&E WSI. For both ROCs that yield AUCs of 0.65 and 0.74 on the validation set and test, respectively, the proposed recurrence prediction model performed better than the traditional approach discussed earlier.

Fig. 4C and 4D show ROCs of the first fold for the test sets computed on patient-wise (WSI) outcomes. We noticed a fair improvement by translating the prediction outcomes from patches to the patient level (WSI). For both ROCs that yield AUCs of 0.83 and 0.87 on the validation set and test, the classification prediction of the model improved modestly by translating the prediction outcomes from patches to the patient (WSI) AUCs 0.18 and 0.13, respectively. Consequently, the ROCs computed using patient-wise outcomes show that the proposed recurrence prediction model performed better than the earlier traditional approach which showed an AUC of 0.49.

There are unresolved questions about the explainability of various AI models, which makes it harder to entirely rely on them, especially in medical diagnosis, prognosis, and decision-making, where one incorrect prediction can have serious consequences. In this paper, we therefore used the gradient-weighted class activation mapping (gradCAM) algorithm to trace back the activated hot spots from the last hidden layer as a form of heat map from which pathologists could learn about the crucial patterns that lead to the correct prediction of the recurrence outcome. For this, we have taken tumor regions only to train the model. Fig. 5 shows a qualitative example of both the recurrence and non-recurrence patients, which are correctly predicted by the proposed recurrence prediction AI framework. In the gradCAM-based generated heatmap, yellow (light) color represents the features of the non-recurrence (positive) class and the blue (dark) color represents the recurrence (negative) class.

Fig. 6 shows a zoomed-in cropped region of both original and heatmap images of recurrence and non-recurrence biopsy samples shown in Fig. 5. In the heatmap shown in Fig. 6B and 6D, yellow regions refer to features that play a key role in the non-recurrence outcome prediction. Fig. 6A shows a region of interest (reconstructed from multiple patches) from a patient who was assigned a recurrence outcome. The image patch shows a tumor area with high cellularity. The tumor cells had enlarged and hyperchromatic nuclei, an increased nucleus-to-cytoplasm ratio, and were surrounded by only scant collagenous stroma. The dominant architectural pattern in the biopsy was assessed as solid growth. Fig. 6B shows the corresponding heatmap used by the proposed model. The heatmap shows very few spots selected in the yellow color, which means that almost no features were activated that correspond to the non-recurrence outcome. Hence, the result belongs to the recurrence outcome. Similarly, Fig. 6C depicts a case assigned to a non-recurrence

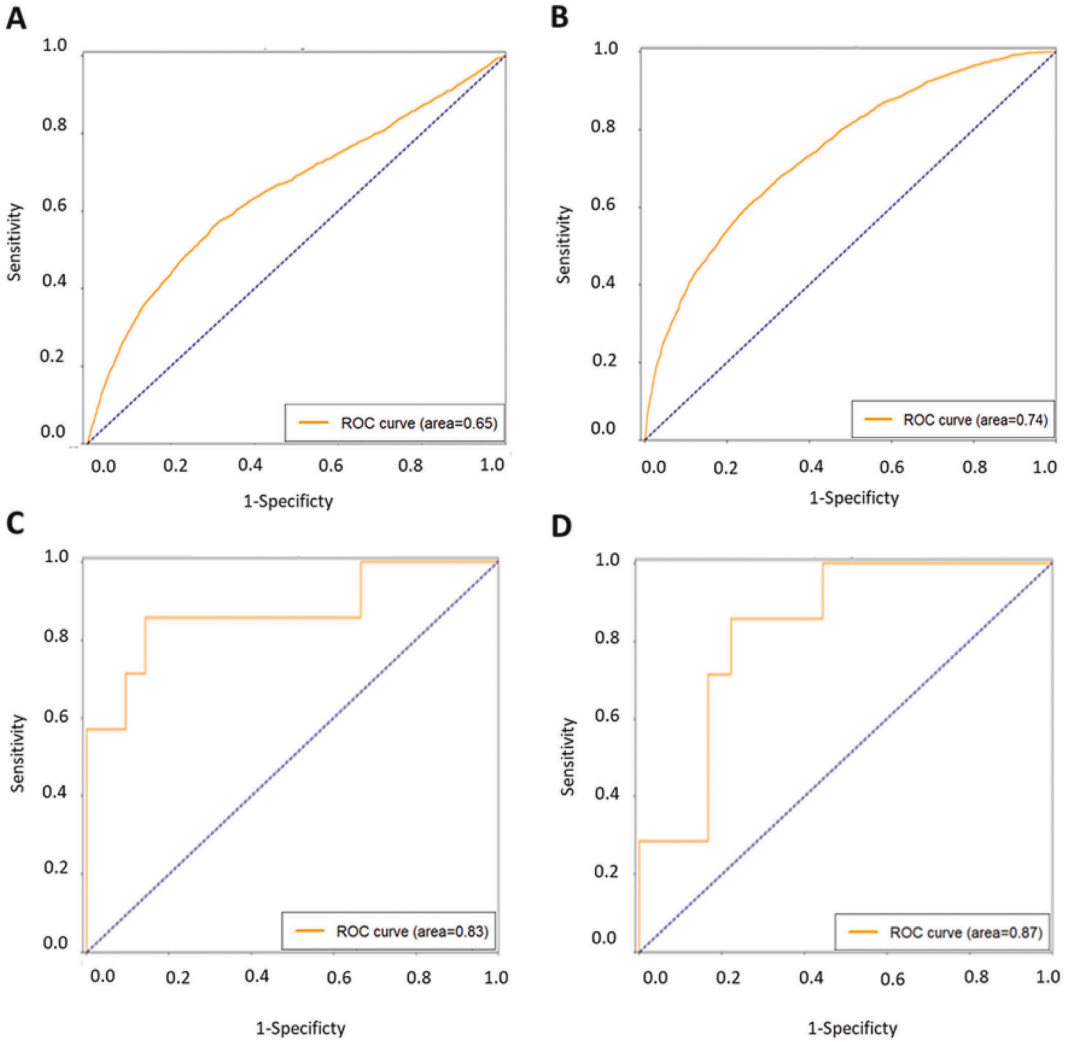


Fig. 4. Validation and test ROCs for patient-wise best outcome ((A) validation and (B) test) and image patch-wise test outcome ((C) validation and (D) test).

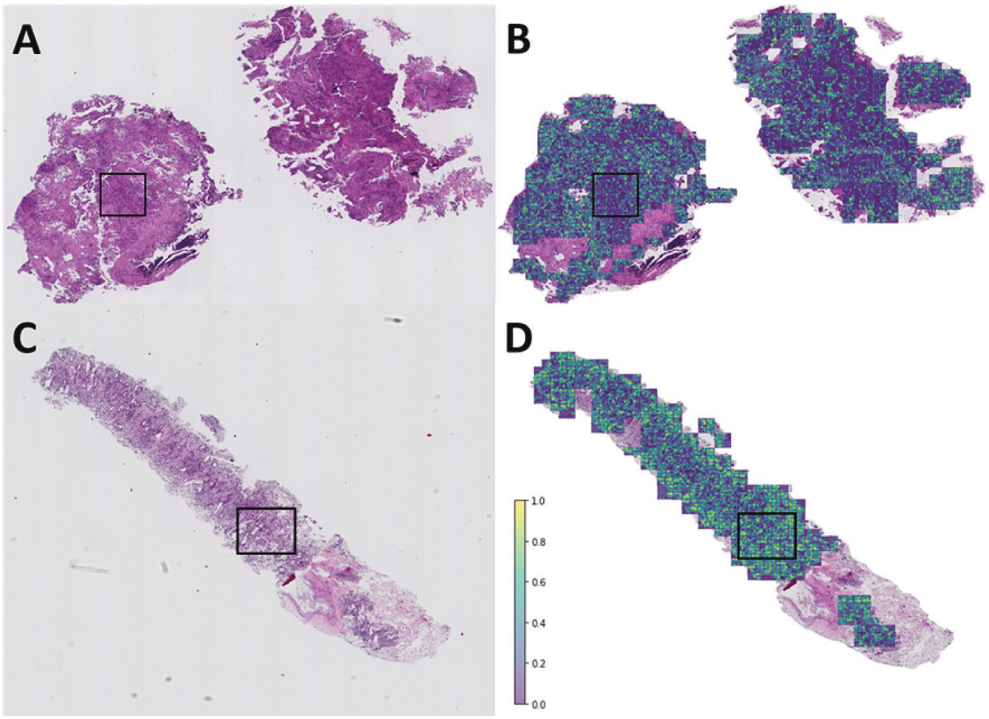


Fig. 5. Hotspot detection in recurred and non-recurred biopsy samples using gradCAM on patient level WSI. (A) Original WSI (recurred patient), (B) WSI with heatmap based hotspot, (C) original WSI (non-recurred patient), (D) WSI with heatmap based hotspot. Hotspots indicate areas of feature extraction where yellow (light) color represents features related to non-recurrence class and blue (dark) color represents features representing recurrence class. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

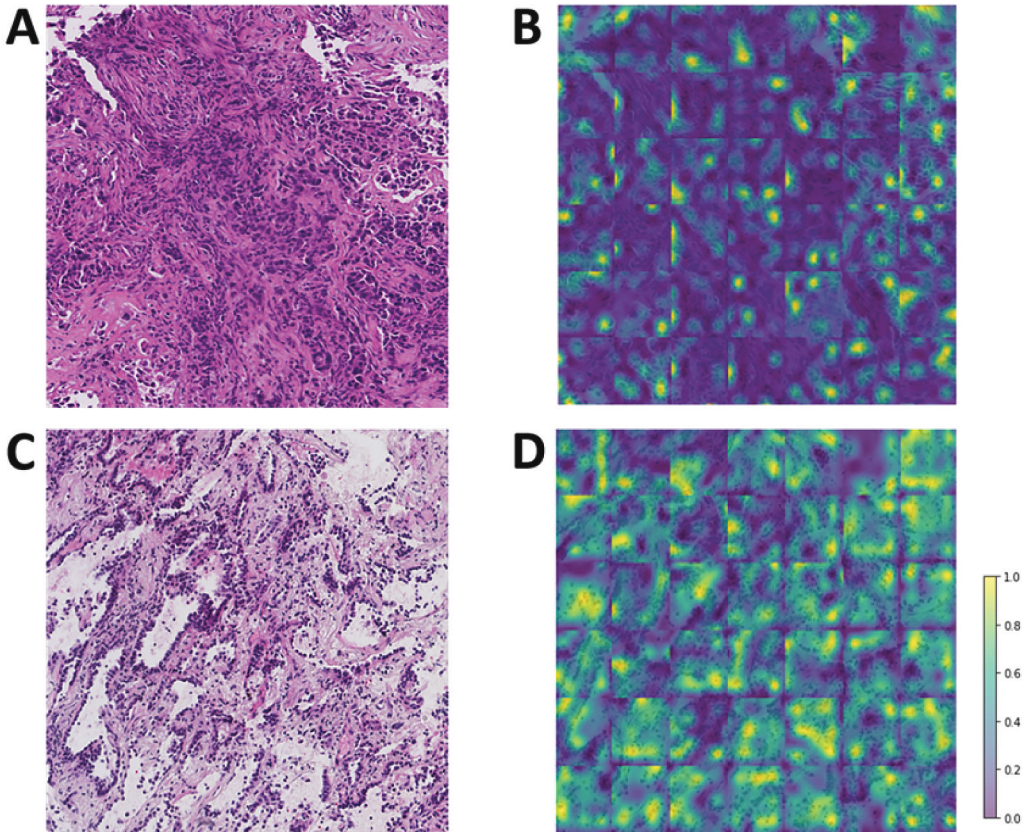


Fig. 6. Zoomed-in crops of the hotspot in recurred and non-recurred biopsy samples shown in Fig. 5. (A) Crop from original WSI (recurred patient), (B) Crop from WSI with heatmap based hotspot, (C) Crop from original WSI (non-recurred patient), (D) from WSI with heatmap based hotspot.

outcome that presented with lower cell density, and smaller tumor cells. The dominant pattern was represented by lepidic growth. Fig. 6D shows the heatmap of the proposed model generated using a selected region of interest. As expected, it shows a considerable number of cellular structures in yellow color, which means it has more features corresponding to the non-recurrence outcome.

4. Conclusions and discussion

In this study, we compare a classical statistical model for predicting recurrence in early and locally advanced stage LUAD using histopathological evaluation data and a deep learning (CNN) approach to develop a classification model for recurrence prediction using H&E WSI biopsy images. The deep learning classification framework proposed in this publication had superior recurrence prediction as compared to the traditional approach utilizing histological information. This classification model yielded a test accuracy of 0.70 and 0.82 and an AUC of 0.74 and 0.87 on patch-wise and patient-wise H&E WSI respectively, compared to the traditional model which showed an AUC of 0.49. These findings for our AI classifier are comparable with those of Shim et al., who developed DeepRePath, a model based on a deep CNN using multiscale pathology images from resection specimens. Their model achieved an AUC of 0.77 and 0.76 in their two LUAD cohorts [36]. On a cellular level, among other items (such as necrosis), they also described nuclear features such as discohesive tumor cells and hyperchromatic nuclei as hallmarks for recurrence. This latter feature was also observed in our heatmaps from the recurrence positive cases. A major difference between both approaches is the amount of tissue and patients analyzed. Compared to Shim et al., we used a smaller cohort and trained and validated our model on biopsy specimens.

It is important to note that by applying our deep learning strategy on entire tumor specimens in biopsies with only the recurrence status as label, we did not impose the classical IASLC/ATS/ERS LUAD scoring rules as features for the classification. Instead using gradCAM for explainable AI, we aimed at identifying the features that may be associated to the recurrence status by the proposed CNN. We compared our deep learning model with the traditional statistical model using ROC analysis which used different histological features to build the model to demonstrate differences in the model's performance. An explanation for the inferior performance of the traditional statistical model might be the exclusion of additional histological features which in resected specimens are known to be accompanied by a higher chance of recurrence. For example, the presence of (lymph-) angio-invasion which is a significant prognostic factor in resected early-stage adenocarcinoma was not included [40–42]. Additionally, the presence or absence of necrosis also seems predictive [43,44], a feature which was also found to be relevant in the DeepRePath model by Shim et al [36]. Furthermore, pleural invasion is correlated with survival, and its impact on survival has even led to the inclusion in the TNM classification [45,46]. However, in contrast to necrosis or lymphatic vessel invasion, pleural invasion is almost impossible to diagnose when dealing with small biopsy samples.

Another reason for the inferior model performance of the classical approach is the data selection. In the present model, we only focused on histology data to better compare CNN with the traditional approach and did not take clinical or molecular data into consideration. From our previous work [37] and from the work of Moreira et al. [9] we had the experience that feeding the model with additional clinical data can improve its performance. Thus Moreira et al. achieved an AUC of 0.77 for recurrence prediction when combining clinical data with the dominant and second dominant growth pattern on resected LUAD specimens. This indicates that combining imaging data with clinical information can increase the model's performance.

An advantage of using a recurrence prediction model on small biopsies is its broad availability. Most patients undergo a transbronchial or CT-guided biopsy during diagnostic workup when presenting with a pulmonary mass. Recurrence prediction can be used to guide the extent of the surgical procedure and it is of clinical interest predominantly when considering the duration of follow-up and the need for adjuvant treatment. Whilst in WSI from resected patients, the removed tumor might offer more information, however, not all patients undergo surgical therapy, which necessitates a reliable model predicting recurrence on small biopsy samples. This is especially true for patients who are selected for SABR or who undergo neoadjuvant treatment.

The disadvantage of recurrence prediction on small biopsies is the fact that only limited amounts of tumor tissue are analyzed and that biopsy specimens are prone to sampling errors. In a biopsy with an acinar pattern there may be an adjacent high-grade pattern such as solid or micropapillary which might be missed, but highly influences prognosis. Even when CNN seems to focus more on nuclear than architectural features the problem of sampling error is not solved, since high-grade growth patterns often show a higher nuclear grade [37].

Whilst our deep learning model was trained and validated only on a small cohort ($n = 124$) we still have a high performance as compared to the existing statistical model. We are planning to increase the number of biopsies, ideally as an external cohort scanned under different settings to reduce bias, to train, validate and test our deep learning model for recurrence prediction.

In summary, the classical model used in the current study can moderately discriminate between recurrence and non-recurrence. However, due to the high recurrence rates of early-stage LUAD, there is a need for a more reliable assessment of recurrence even on a low amount of tissue. Therefore, we developed a deep learning classifier that works purely on image patches of tumorous tissue without a need for additional information on the clinical history or mutational state, which simplifies data collection. This makes the classifier universally deployable on lung biopsy WSI with an acceptable AUC of 0.87 on patient-wise WSI. This type of classifier may help to determine follow-up regimens and to select the correct type and extent of curative surgical or radiotherapeutic treatment in early-stage adenocarcinoma. Moreover, it may in future guide the use of neo-adjuvant or adjuvant therapy, thus leading to potential de-escalation of (*peri*-operative) systemic treatment in cases which are not likely to recur following curative tumor ablation.

Author contributions

F. Akram designed and carried out the AI experiments, produced figures and tables, and wrote the manuscript.

J.L. Wolf composed the patient cohort, performed annotations and carried out statistical analysis, and produced graphs.

T.E. Trandafir annotated the data, assisted in AI experiments, and contributed to writing the manuscript.

Anne-Marie C. Dingemans supervised the work and reviewed the manuscript.

A.P. Stubbs supervised the work and reviewed the manuscript.

J.H. von der Thüsen conceived the study, selected and contributed to composing the patient cohort, and co-authored the manuscript.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors Janina Wolf, Teodora Trandafir, Farhan Akram, and Andrew Stubbs declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work of this paper. Anne-Marie Dingemans reports consulting fees from Sanofi, Amagen, Bayer, Roche, Astra Zeneca and Boehringer Ingelheim and payment honoraria from Eli Lilly Pfizer and Astra Zeneca. Jan von der Thüsen reports consulting fees from Astra Zeneca, Eli Lilly, Pfizer, Roche, Roche Diagnostics and MSD.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lungcan.2023.107413.

References

[1] Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening, N. Engl. J. Med. 365 (5) (2011) 395–409.
[2] H.J. de Koning, et al., Reduced lung-cancer mortality with volume CT screening in a randomized trial, N. Engl. J. Med. 382 (6) (2020) 503–513.
[3] J.J. Hung, et al., Prognostic factors of postrecurrence survival in completely resected stage I non-small cell lung cancer with distant metastasis, Thorax 65 (3) (2010) 241–245.
[4] R. Maeda, et al., Risk factors for tumor recurrence in patients with early-stage (stage I and II) non-small cell lung cancer: patient selection criteria for adjuvant chemotherapy according to the seventh edition TNM classification. Chest, 2011. **140**(6): p. 1494-1502.
[5] M. Ito, et al., Pathological high malignant grade is higher risk of recurrence in pN0M0 invasive lung adenocarcinoma, even with small invasive size, Thorac. Cancer 12 (23) (2021) 3141–3149.
[6] T. Koike, et al., Characteristics and timing of recurrence during postoperative surveillance after curative resection for lung adenocarcinoma, Surg. Today 47 (12) (2017) 1469–1475.
[7] J.E. Leeman, et al., Histologic subtype in core lung biopsies of early-stage lung adenocarcinoma is a prognostic factor for treatment response and failure patterns after stereotactic body radiation therapy, Int. J. Radiat. Oncol. Biol. Phys. 97 (1) (2017) 138–145.
[8] W.D. Travis, et al., International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma, J. Thorac. Oncol. 6 (2) (2011) 244–285.
[9] A.L. Moreira, et al., A grading system for invasive pulmonary adenocarcinoma: a proposal from the international association for the study of lung cancer pathology committee, J. Thorac. Oncol. 15 (10) (2020) 1599–1610.
[10] A. Yoshizawa, et al., Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases, Mod. Pathol. 24 (5) (2011) 653–664.
[11] A. Warth, et al., The novel histologic international association for the study of lung cancer/american thoracic society/european respiratory society classification system of lung adenocarcinoma is a stage-independent predictor of survival, J. Clin. Oncol. 30 (13) (2012) 1438–1446.
[12] P.A. Russell, et al., Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new international association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary lung adenocarcinoma classification, J. Thorac. Oncol. 6 (9) (2011) 1496–1504.
[13] J.J. Hung, et al., Predictive value of the international association for the study of lung cancer/American Thoracic Society/European Respiratory Society classification of lung adenocarcinoma in tumor recurrence and patient survival, J. Clin. Oncol. 32 (22) (2014) 2357–2364.
[14] E. Yi, et al., Pathological prognostic factors of recurrence in early stage lung adenocarcinoma, ANZ J. Surg. 88 (4) (2018) 327–331.
[15] H.W. Jeon, et al., Significant difference in recurrence according to the proportion of high grade patterns in stage IA lung adenocarcinoma, Thorac. Cancer 12 (13) (2021) 1952–1958.
[16] X. Liu, et al., Different pathologic types of early stage lung adenocarcinoma have different post-operative recurrence patterns, Thorac. Cancer 12 (15) (2021) 2205–2213.
[17] H. Liao, et al., Mutational status of main driver genes influences the prognosis of stage I-III lung adenocarcinoma patients underwent radical surgery, Transl. Cancer Res. 10 (7) (2021) 3286–3298.
[18] C. Deng, et al., Genetic-pathological prediction for timing and site-specific recurrence pattern in resected lung adenocarcinoma, Eur. J. Cardiothorac. Surg. 60 (5) (2021) 1223–1231.
[19] I.A. Kim, et al., Targeted next-generation sequencing analysis for recurrence in early-stage lung adenocarcinoma, Ann. Surg. Oncol. 28 (7) (2021) 3983–3993.
[20] P.J. Kneuertz, et al., Prognostic value and therapeutic implications of expanded molecular testing for resected early stage lung adenocarcinoma, Lung Cancer 143 (2020) 60–66.
[21] P.E. Postmus, et al., Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol, 2017. **28**(suppl_4): p. iv1-iv21.
[22] H.G. Colt, et al., Follow-up and surveillance of the patient with lung cancer after curative-intent therapy: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines, Chest 143 (5 Suppl) (2013) e437S–e454S.
[23] NVALT, Niet kleincellig longcarcinoom - Follow up NSCLC-patient na curatieve behandeling. 2020, Federatie medisch specialisten. p. 306-316.
[24] P.M. Forde, et al., Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer, N. Engl. J. Med. 386 (21) (2022) 1973–1985.
[25] Y.L. Wu, et al., Osimertinib in resected EGFR-mutated non-small-cell lung cancer, N. Engl. J. Med. 383 (18) (2020) 1711–1723.
[26] S. Hindocha, et al., A comparison of machine learning methods for predicting recurrence and death after curative-intent radiotherapy for non-small cell lung cancer: development and validation of multivariable clinical prediction models, EBioMedicine 77 (2022), 103911.
[27] B. Lee, et al., DeepBTS: prediction of recurrence-free survival of non-small cell lung cancer using a time-binned deep neural network, Sci. Rep. 10 (1) (2020) 1952.
[28] Y. She, et al., Development and validation of a deep learning model for non-small cell lung cancer survival, JAMA Netw. Open 3 (6) (2020) e205842.
[29] G.D. Jones, et al., A genomic-pathologic annotated risk model to predict recurrence in early-stage lung adenocarcinoma, JAMA Surg. 156 (2) (2021).
[30] K.Z. Simonyan, A very deep convolutional networks for large-scale image recognition, in ICLR. 2015. p. 14.
[31] C. Szegedy, et al., Inception-v4, inception-ResNet and the impact of residual connections on learning. proceedings of the AAAI conference on artificial intelligence, 2017. **31**(1).
[32] M. Kriegsmann, et al., Deep learning for the classification of small-cell and non-small-cell lung cancer, Cancers (basel) 12 (6) (2020).
[33] F. Kanavati, et al., A deep learning model for the classification of indeterminate lung carcinoma in biopsy whole slide images, Sci. Rep. 11 (1) (2021) 8110.
[34] H. Yang, et al., Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: a retrospective study, BMC Med. 19 (1) (2021) 80.
[35] M. Sandler, M. Zhu, A. Zhmoginov, L.-C. Chen, MobileNetV2: inverted residuals and linear bottlenecks, in IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018. p. pp. 4510-4520.
[36] W.S. Shim, et al., DeepRePath: identifying the prognostic features of early-stage lung adenocarcinoma using multi-scale pathology images and deep convolutional neural networks, Cancers (basel) 13 (13) (2021).
[37] J.L. Wolf, et al., The value of prognostic and predictive parameters in early-stage lung adenocarcinomas: a comparison between biopsies and resections, Lung Cancer 176 (2022) 112–120.
[38] M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, ArXiv (2019) abs/1905.11946.
[39] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, Journal of Big Data 3 (1) (2016) 9.
[40] K. Funai, et al., Lymphatic vessel invasion is a significant prognostic indicator in stage IA lung adenocarcinoma, Ann. Surg. Oncol. 18 (10) (2011) 2968–2972.
[41] M.J. Schuchert, et al., Impact of angiolymphatic and pleural invasion on surgical outcomes for stage I non-small cell lung cancer. Ann Thorac Surg, 2011. **91**(4): p. 1059-65; discussion 1065.
[42] T. Tsuchiya, et al., Stage IA non-small cell lung cancer: vessel invasion is a poor prognostic factor and a new target of adjuvant chemotherapy, Lung Cancer 56 (3) (2007) 341–348.
[43] M. Inoue, et al., Clinicopathologic factors influencing postoperative prognosis in patients with small-sized adenocarcinoma of the lung, J. Thorac. Cardiovasc. Surg. 135 (4) (2008) 830–836.
[44] C.F. Wu, et al., Recurrence risk factors analysis for stage I non-small cell lung cancer, Medicine (Baltimore) 94 (32) (2015) e1337.
[45] W.D. Travis, et al., Visceral pleural invasion: pathologic criteria and use of elastic stains: proposal for the 7th edition of the TNM classification for lung cancer. J Thorac Oncol, 2008. **3**(12): p. 1384-90.
[46] R. Rami-Porta, et al., Lung cancer - major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin, 2017. **67**(2): p. 138-155.