



Research Article

Deep Learning–Based Retinoblastoma Protein Subtyping of Pulmonary Large-Cell Neuroendocrine Carcinoma on Small Hematoxylin and Eosin–Stained Specimens

Teodora E. Trandafir^a, Frank W.J. Heijboer^b, Farhan Akram^a, Jules L. Derks^{b,c}, Yunlei Li^a, Lisa M. Hillen^d, Ernst-Jan M. Speel^{a,d}, Zsolt Megyesfalvi^{e,f,g}, Balazs Dome^{e,f,g,h}, Andrew P. Stubbs^a, Anne-Marie C. Dingemans^b, Jan H. von der Thüsen^{a,*}

^a Department of Pathology and Clinical Bioinformatics, Erasmus University Medical Center, Rotterdam, The Netherlands; ^b Department of Respiratory Medicine, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, The Netherlands; ^c Department of Pulmonary Diseases, GROW School for Oncology and Reproduction, Maastricht University Medical Centre, Maastricht, The Netherlands; ^d Department of Pathology, GROW School for Oncology and Reproduction, Maastricht University Medical Centre, Maastricht, The Netherlands; ^e Department of Thoracic Surgery, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria; ^f Department of Thoracic Surgery, Semmelweis University and National Institute of Oncology, Budapest, Hungary; ^g National Koranyi Institute of Pulmonology, Budapest, Hungary; ^h Department of Translational Medicine, Lund University, Lund, Sweden

ARTICLE INFO

Article history:

Received 29 October 2024

Revised 15 April 2025

Accepted 30 April 2025

Available online 8 May 2025

Keywords:

deep learning
hematoxylin and eosin
large-cell neuroendocrine carcinoma
retinoblastoma 1
subtyping

ABSTRACT

Molecular subtyping of pulmonary large-cell neuroendocrine carcinoma (LCNEC) based on retinoblastoma protein (pRb) expression may influence systemic treatment decisions. Current histomorphologic assessments of hematoxylin and eosin–stained tissue samples cannot reliably differentiate LCNEC molecular subtypes. This study explores the potential of deep learning (DL) to identify histologic patterns that distinguish these subtypes, by developing a custom convolutional neural network to predict the binary expression of pRb in small LCNEC tissue samples. Our model was trained, cross-validated, and tested on tissue microarray cores from 143 resection specimens and biopsies from 21 additional patients, with corresponding immunohistochemical pRb status. The best-performing DL model achieved a patient-wise balanced accuracy value of 0.75 and an area under the receiver operating characteristic curve value of 0.77 when tested on biopsies, significantly outperforming the hematoxylin and eosin–based subtype classification by lung pathologists. Explainable artificial intelligence techniques further highlighted coarse chromatin patterns and distinct nucleoli as distinguishing features for pRb retained status. Meanwhile, pRb lost cases were characterized by limited cytoplasm and morphologic similarities with small cell lung cancer. These findings suggest that DL analysis of small histopathology samples could ultimately replace immunohistochemical pRb testing. Such a development may assist in guiding chemotherapy decisions, particularly in metastatic cases.

© 2025 THE AUTHORS. Published by Elsevier Inc. on behalf of the United States & Canadian Academy of Pathology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

These authors contributed equally: Teodora E. Trandafir and Frank W.J. Heijboer.

These authors contributed equally: Anne-Marie C. Dingemans and Jan H. von der Thüsen.

* Corresponding author.

E-mail address: j.vonderthusen@erasmusmc.nl (J.H. von der Thüsen).

Introduction

Large-cell neuroendocrine carcinoma (LCNEC) is a high-grade tumor in the spectrum of lung neuroendocrine neoplasms



(NENs). In comparison with the indolent typical carcinoid or atypical carcinoid, LCNEC metastasizes frequently, and the diagnosis is often made in an advanced stage.¹ According to the latest World Health Organization (WHO) classification, the diagnosis of LCNEC can be made in the presence of >10 mitoses/2 mm², substantial necrosis, neuroendocrine morphology, and at least 1 positive neuroendocrine immunohistochemical (IHC) marker (CD56, chromogranin A, or synaptophysin).^{2,3} However, the WHO stresses that a diagnosis of LCNEC should be reserved for resection specimens and not be made on a biopsy owing to the potential lack of morphologic detail.⁴ However, other studies have suggested additional tools to facilitate diagnosing LCNEC on a biopsy specimen.^{5,6}

Mutation analysis has identified the following 2 subtypes in LCNEC: non–small cell-like (“NSCLC-like”) and small cell-like (“SCLC-like”).^{7,8} SCLC-like LCNEC is characterized by mutations of tumor protein p53 (*TP53*) and loss of retinoblastoma 1 (*RB1*), which is a hallmark of SCLC.⁹ However, NSCLC-like LCNEC lacks *RB1* and *TP53* coalteration but is characterized by typical NSCLC-associated mutations such as serine/threonine kinase 11 (*STK11*) mutations, Kirsten rat sarcoma viral oncogene homolog (*KRAS*) mutations, and kelch-like ECH-associated protein 1 (*KEAP1*) mutations.¹⁰ Retention or loss of *RB1* is of great importance for patients with LCNEC. When carrying a wild-type *RB1* gene or preserved expression of its gene product retinoblastoma protein (pRb), significantly longer overall survival was seen when treated with platinum–gemcitabine or taxane chemotherapy (NSCLC regimen) vs platinum–etoposide chemotherapy (SCLC regimen)¹¹; however, this was not seen in all studies.¹²

Based on histomorphology alone, that is, using hematoxylin and eosin (H&E)–stained tissue samples, it is not yet possible to faithfully differentiate these molecular subclasses of LCNEC, except that SCLC-like LCNEC tends to exhibit higher proliferation rates than NSCLC-like LCNEC.⁷ A potential method to identify cytomorphologic features, which may distinguish between SCLC- and NSCLC-like LCNEC, is the use of deep convolutional neural networks (CNNs). With the emergence of digital pathology, deep learning (DL) has reached state-of-the-art results in classifying and prognosticating certain types of cancer using histopathological images.¹³ Several studies have proposed successful models to classify and prognosticate different lung cancers.^{14–18} Nevertheless, the applicability of artificial intelligence (AI) strategies, particularly DL algorithms, in LCNEC digital pathology, is currently limited, and often restricted to limited cohorts, thus necessitating further developments. For example, Yang et al¹⁹ proposed a DL model to subclassify lung carcinoma into lung adenocarcinoma, squamous cell carcinoma, LCNEC, SCLC, and nonneoplastic lung tissue in biopsy specimens, with an average area under the receiver operating characteristic curve (ROC-AUC) of 0.90. Furthermore, Ilić et al²⁰ developed methods to distinguish histologic subtypes of pulmonary NENs, such as SCLC, LCNEC, and atypical carcinoid from digitized resection specimens, with an average AUC of 0.93. Concerning cytologic smears, Gonzalez et al²¹ designed a DL algorithm for classifying high-grade NENs, including SCLC and LCNEC, in fine needle aspirations. To the best of our knowledge, no DL model has been reported for the prediction of molecular subtypes in LCNEC.

In this study, we employed DL strategies in a pioneering analysis of the phenotypical/histopathological differences between SCLC-like and NSCLC-like LCNEC in small tissue samples stained with H&E (Fig. 1). The 2 LCNEC molecular subtypes are defined IHC based on the binary expression of pRb in adjacent tissue sections. Using limited tissue information, such as biopsies,

our model may ultimately guide the assignment of a suitable chemotherapeutic strategy in the metastatic setting.

Materials and Methods

Patients and Tissue Cohorts

Patient samples were gathered via 2 independent nationwide LCNEC cohorts. The first cohort consisted of 128 patients diagnosed with stage IA to IIIB (sixth/seventh TNM) LCNEC on a resection specimen during the period between January 2003 and December 2012 in the Netherlands. All resection specimens were independently reviewed by 2 experienced pulmonary pathologists (J.H.v.d.T. and L.M.H.) to establish the diagnosis of LCNEC,² in addition to the initially established diagnosis and neuroendocrine marker (CD56, chromogranin A, and synaptophysin) evaluation. In case of any doubt, the samples were excluded from the study (Supplementary Fig. S1).

The second nationwide cohort comprised 21 biopsy and 15 resection specimens from different patients with LCNEC between January 2019 and April 2024, collected prospectively and retrospectively. For these cases, the diagnosis of LCNEC had been established by a panel of 5 thoracic pathologists (including L.M.H. and J.H.v.d.T.), by strictly following the 2021 WHO guidelines for resection and biopsy classification. Biopsies were included only if sufficient tissue was available for evaluation of neuroendocrine morphologies, and if crush artefacts and extensive necrosis did not limit the differentiation from SCLC. If available, the corresponding resection specimen was used to confirm the biopsy diagnosis. Along with the H&E specimen, neuroendocrine markers (CD56, chromogranin A, and synaptophysin) have been evaluated to clarify the LCNEC diagnosis using IHC. Consensus was reached when 4 out of 5 pathologists (including L.M.H. and J.H.v.d.T.) agreed upon the diagnosis of LCNEC. When consensus was not reached, a case was discussed in a consensus meeting. The diagnosis of LCNEC was then established when the majority of participants agreed. In addition, we included an external cohort of 115 SCLC confirmed patients provided by the Medical University of Vienna, Austria, and the National Koranyi Institute of Pulmonology, Hungary.²² In addition to the original classification, the SCLC diagnosis was independently confirmed by a study pathologist (J.H.v.d.T.).²³

The study protocol was approved by the medical ethical committee of the Erasmus Medical Center, Rotterdam, the Netherlands (MEC-2022-0130).

Whole-Slide Imaging

Representative tumor blocks from the first cohort were chosen for constructing tissue microarrays (TMAs) by randomly punching the tumor area of the FFPE blocks to capture tumor heterogeneity (1–3 cores per case, 1.0 mm in diameter). Consecutive 4-μm-thick paraffin sections were cut and stained with H&E (HE600, catalog number 06917259001; Ventana) and pRb IHC (clone 13A10, dilution 1/50, catalog number NCL-LRB 358; DAKO) using standard protocols. We obtained 5 TMA slides encompassing 128 LCNEC cases for subsequent image preprocessing, representing 96 pRb lost and 32 pRb retained cases. Additional characterization of p16, p40, cyclin D1, and Napsin A was used to understand the functionality of pRb.^{24,25} The TMA arrays stained with H&E were scanned using the NanoZoomer 2.0-HT scanner from Hamamatsu (Hamamatsu Photonics KK) at ×40 magnification, with a

resolution of 0.23 $\mu\text{m}/\text{pixel}$. Independent cores comprising sufficient tissue were extracted using the TMA dearrayer functionality of QuPath v.0.5.0²⁶ and exported as separate JPG files at a resolution of 0.46 $\mu\text{m}/\text{pixel}$ ($\times 20$ magnification). Following quality control, 8 cores were automatically excluded owing to enhanced blur or analysis-obstructing artefacts, in a process that did not affect the total number of patients.

Samples from the second LCNEC cohort were not included in TMAs but underwent similar staining and imaging procedures as the first cohort. Following IHC evaluation, we obtained 7 pRb lost biopsies, 2 pRb lost resections, 14 pRb retained biopsies, and 13 pRb retained resections.

In this study, we further used an external cohort of 115 SCLC tissue specimens sampled in 8 TMA slides with maximally 2 cores per case stained and imaged similarly to the rest of the data.

Tumor Tissue Annotation and Image Preprocessing

All tissue WSIs that had passed quality control were manually annotated in QuPath v.0.5.0²⁶ to thoroughly mark the tumor area from surrounding tissue and artefacts. Solely for the resection specimens, 3 random regions with dimensions equating to a TMA core were selected for the analysis. From all tissue samples, nonoverlapping 128×128 -pixel tiles were extracted. Otsu thresholding was applied to effectively separate the white background from the tissue regions within the tiles, ensuring that only tiles with tissue coverage surpassing 80% of the total area were retained.²⁷ As previously highlighted by Illé et al.,²⁰ in resection specimens, tissue areas located at the edges of the specimen (with possible crush and sampling artefacts) could bias a model's performance toward smaller cellular morphologies. Therefore, we automatically excluded marginal tiles from our analysis. Furthermore, samples were discarded if they covered less than 85% tumor area, as marked by the manual annotations. This is under the assumption that an ideal model considers all tumor cells in its decision, when distinguishing between the pRb-based subtypes of LCNEC. Last, patches with a high level of blur were automatically filtered out. We generated a total of 38,777 tiles from the LCNEC TMA set and 8878 tiles from the LCNEC biopsy set through this image-preprocessing process. Similarly, we obtained 7988 tiles from the SCLC TMA set.

Training Process

Following preprocessing, the data set consisted of resection tiles from 143 patients (45 pRb retained and 98 pRb lost), and biopsy tiles from 21 other patients (14 pRb retained and 7 pRb lost). All H&E-stained tissue tiles that passed quality control were grouped patient wise to prevent data leakage among the training, validation, and test sets. The proposed CNN was trained and validated using stratified k-fold cross-validation ($\text{kappa} = 4$) on a randomly selected set of 27,494 resection tiles from 100 patients (44 pRb retained and 56 pRb lost). Subsequently, testing was done on the remaining data of TMAs and biopsies (Fig. 2A). To focus on the behavior of the model on the biopsy data, we separately computed the performance on the biopsy subset of the test set.

All experiments were conducted using Nvidia 3090 RTX (Nvidia) GPU hardware.

Model Design

A custom-made CNN architecture, encompassing 2 convolutional blocks, was constructed to extract hierarchical information

from the input tiles.²⁸ Each convolutional block consisted of 2 convolutional layers for feature extraction with L1 regularization, batch normalization to stabilize the learning process, and a max-pooling layer for spatial reduction (Fig. 2B). Then, the retrieved features were flattened to reduce spatial dimensions and combined by a fully connected layer with L1 regularization. The last dense layer, using sigmoid activation, enabled the final binary classification. The learning process was optimized using the Adam optimizer with learning rate decay, and the learning error was iteratively computed using the weight-balanced binary cross-entropy loss function, with class weights inversely proportional to the number of tiles per class. This CNN architecture had sufficient parameters for the processing of our limited data set of small tiles of 128×128 pixels and effectively captured key distinguishing features from the tiles without an early offset of overfitting to noise. Furthermore, the combination of 5×5 - and 3×3 -pixel kernels allowed our network to capture both local and slightly broader contextual information from the images. Larger kernels (such as 5×5) captured more context per filter such as variations in cell shapes and arrangements, and smaller kernels (like 3×3) helped refine the features learned by the previous layers.

Performance Evaluation

As we expected homogeneous pRb nuclear immunoreactivity throughout the tumor in retained cases, patient-level predictions were obtained using majority voting from the predictions of all corresponding tiles (Fig. 1H, J). At both tile and patient levels, the classification performance was evaluated using confusion matrices, ROC-AUC, accuracy, balanced accuracy, precision, recall, and F1/Dice coefficient.

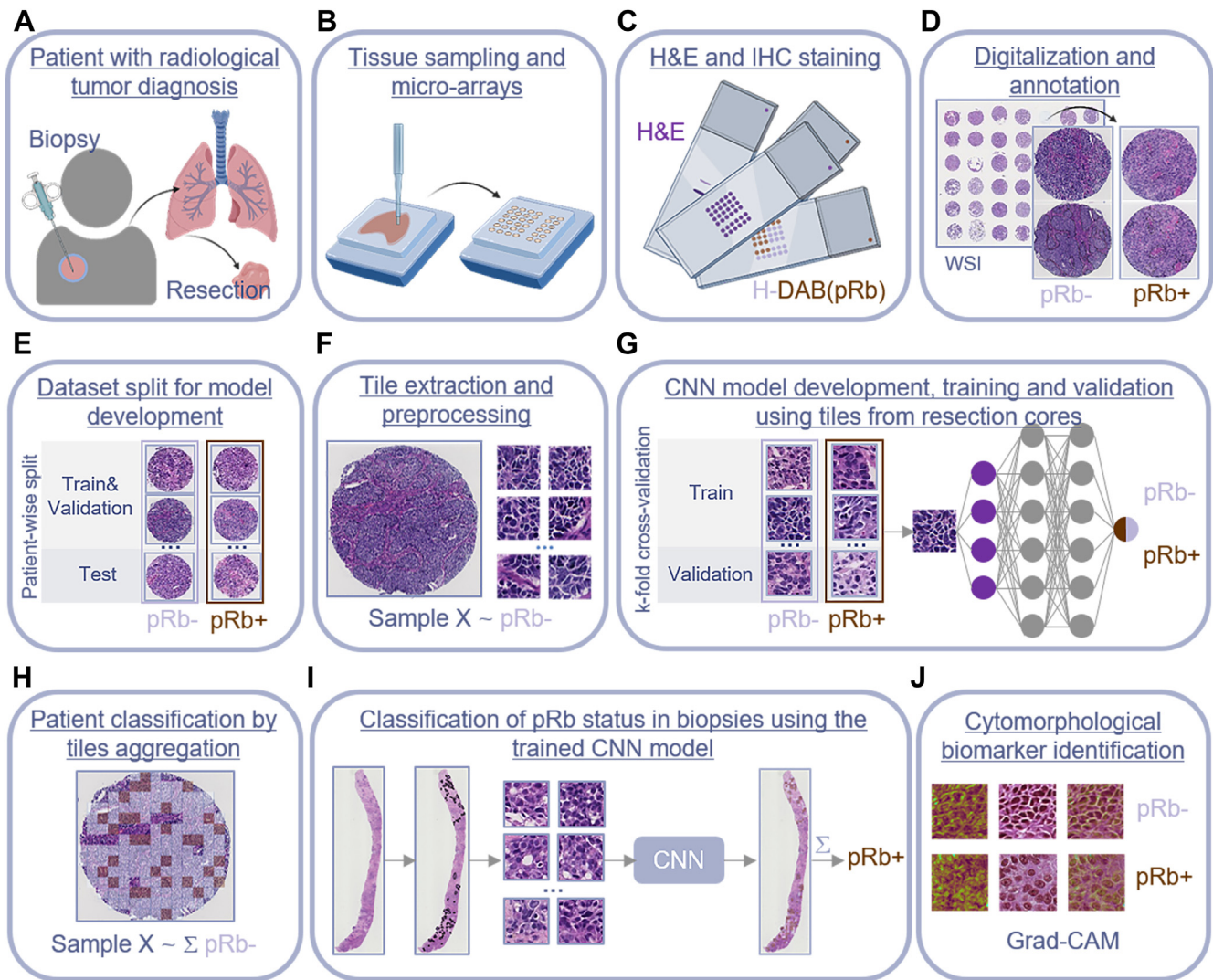
Feature Visualization and Exploration

Gradient-weighted class activation mapping (Grad-CAM) is an explainable AI technique providing insights into the decision-making process of CNNs by determining which series of neurons are activated in the forward pass during inference.²⁹ Grad-CAM uses the gradients of each target class that flow into the last convolutional layer to produce a coarse localization map highlighting the relevant features of an input image for predicting the respective class (Fig. 1J).

Results

Extensive Immunohistochemical Evaluation

In addition to pRb evaluation, the expression of p16, p40, cyclin D1, and Napsin A was IHC evaluated on the TMA set. The results addressing the functionality of pRb are presented in [Supplementary Tables S1 to S6](#). The expression of p16 is strongly linked to pRb loss, consistent with compensatory feedback regulation ([Supplementary Table S1](#)).²⁴ Furthermore, cyclin D1 overexpression occurs in both pRb retained and pRb lost tumors, suggesting it may contribute to tumor progression in some cases ([Supplementary Table S2](#)). The predominant expression pattern is pRb loss with p16 expression and without cyclin D1 expression (53 cases), aligning with the findings of Papaxoinis et al.²⁵ ([Supplementary Table S3](#)). This is consistent with classical RB pathway disruption, namely loss of pRb, leading to compensatory

**Figure 1.**

An overview of the workflow enabling the classification of retinoblastoma protein (pRb)-based large-cell neuroendocrine carcinoma subtypes using deep learning. “pRb+” refers to the pRb retained class, whereas “pRb-” refers to the pRb lost class, as assessed using immunohistochemistry. CNN, convolutional neural network; GRAD-CAM, gradient-weighted class activation mapping; H-DAB, hematoxylin-3, 3'-diaminobenzidine; H&E, hematoxylin and eosin; IHC, immunohistochemistry; WSI, whole-slide image.

p16 upregulation, but cell cycle progression may occur through other mechanisms than cyclin D1 overexpression. No significant correlation was identified between pRb and Napsin A expression (Supplementary Table S4). Combined pRb/p16/cyclin D1/Napsin A subtypes of LCNEC can be found in Supplementary Table S5. Furthermore, p40 evaluation showed that 6 cases may represent LCNEC with focal squamous differentiation or a combined LCNEC-squamous cell carcinoma (Supplementary Table S6).

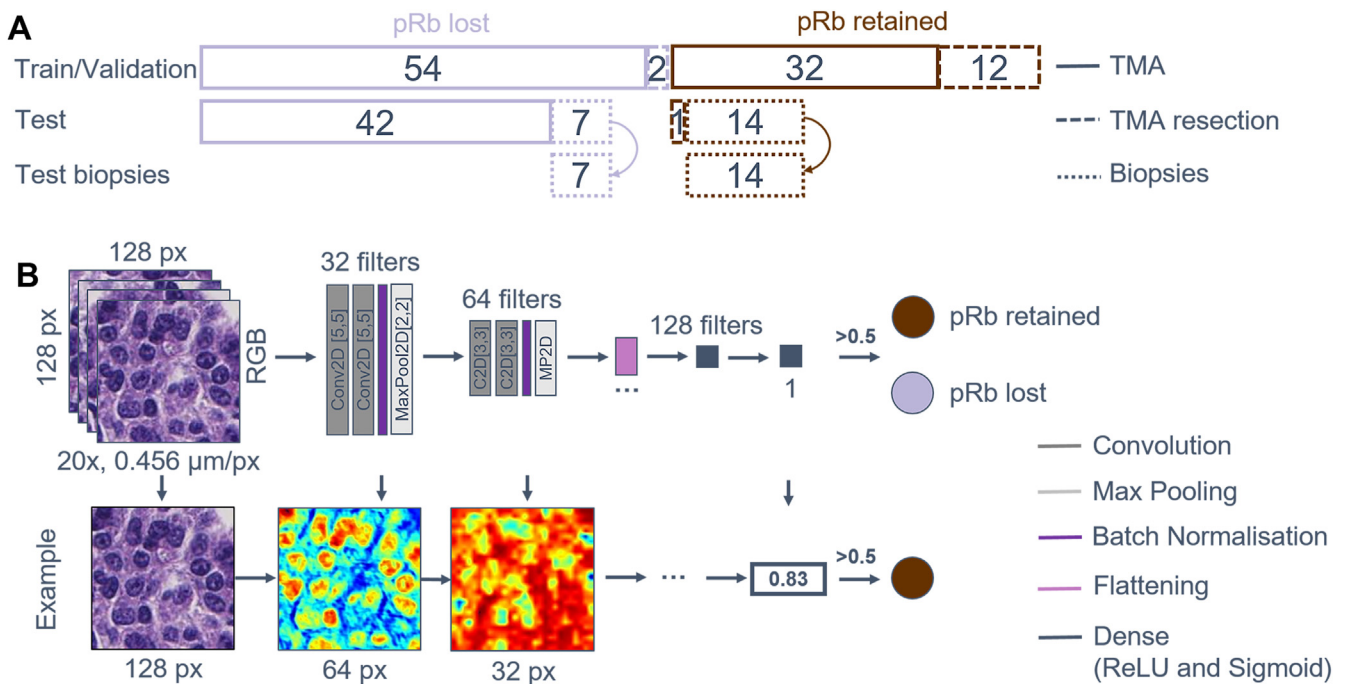
Performance Evaluation on Validation and Test Sets

Across all cross-validation folds, the model achieved an average patient-wise validation ROC-AUC value of 0.76 ± 0.08 (Supplementary Fig. S2) and an average patient-wise test ROC-AUC value of 0.67 ± 0.11 (Supplementary Fig. S3). Based solely on biopsies, the patient-wise ROC-AUC value was 0.62 ± 0.11 (Supplementary Fig. S4). The variable performance within the folds was probably caused by the limited training data and heterogeneity within the data set,

which limited the random split folds 1 and 4 to generalize to the test sets (Supplementary Table S7). However, the best-performing fold (fold 3), identified by the lowest validation loss, was evaluated for its performance at both the tile level and patient level (Fig. 3A-D). On the validation set, the model achieved a balanced accuracy value of 0.76 at both levels. It also demonstrated good generalizability to the test set, with a balanced accuracy value of 0.70 at the tile level and 0.76 at the patient level. Specifically, for biopsies, the model recorded a tile-level balanced accuracy value of 0.71 and a patient-level balanced accuracy value of 0.75.

Comparative Performance Analysis With Pathologists' Histopathological Examination

To evaluate the model's performance relative to that of pathologists, 2 experts (J.H.v.d.T. and L.M.H.) independently classified the H&E-stained biopsy specimens as either “SCLC” or “NSCLC-like LCNEC,” that is, as pRb lost or pRb retained, without knowledge of

**Figure 2.**

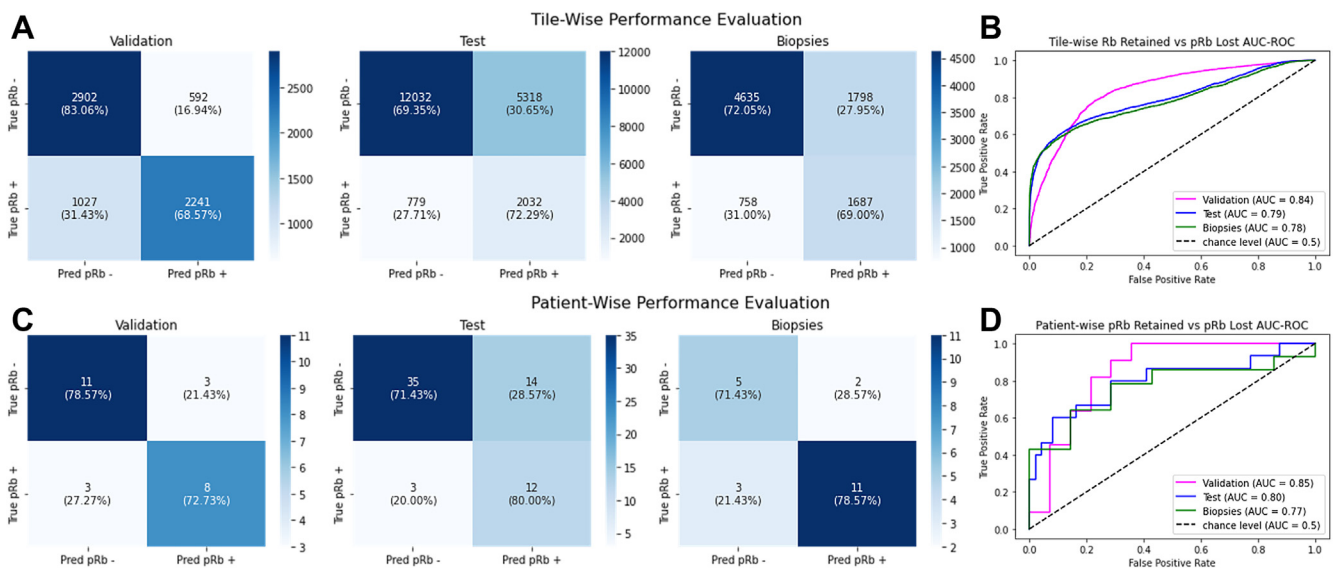
Data set structure and convolutional neural network architecture. (A) An overview of data set organization. (B) The architecture of the convolutional neural network, in which Conv2D (C2D) denotes 2-dimensional convolutional layers with either 5×5 -pixel or 3×3 -pixel kernels, and MaxPool2D (MP2D) indicates a layer with 2-dimensional max-pooling kernels. pRb, retinoblastoma protein; px, pixels; ReLU, rectified linear unit activation function; RGB, Red Green Blue; TMA, tissue microarray.

their pRb status from routine IHC (see [Tables 1 and 2](#)). This classification was made based on general cytonuclear and architectural features that are more reminiscent of SCLC (eg, paucity of cytoplasm, mild nuclear pleomorphism, and inconspicuous nucleoli) or NSCLC (eg, abundant cytoplasm, variable nuclear pleomorphism, and prominent nucleoli), respectively. When comparing the performance metrics, the model performed significantly better than both specialists, particularly at identifying pRb lost cases. Details on

the performance of specialist analysis can be found in [Supplementary Figures S5 and S6](#).

Comparative Analysis of Histomorphologic Features in Small Cell-Like Large-Cell Neuroendocrine Carcinoma and SCLC

To investigate whether the extracted histomorphologic features associated with loss of pRb status, indicative of “SCLC-like

**Figure 3.**

Evaluation of retinoblastoma protein (pRb)-based subtyping in large-cell neuroendocrine carcinoma. (A) Confusion matrices depicting tile-level performance across the validation set, test set, and the biopsy subgroup within the test set. (B) Corresponding ROC-AUC curves for tile-level classification. (C) Confusion matrices summarizing patient-level predictions, aggregated from tile-level classifications. (D) Corresponding ROC-AUC curves for patient-level analysis. “pRb+” and “pRb-” stand for “retained” and “lost,” whereas “pred” refers to “predictions.” ROC-AUC, area under the receiver operating characteristic curve.

Table 1

Comparison of the deep learning model's patient-level predictions on the biopsy data set with evaluations made by pathologists

Evaluator	Patient/WSI-level predictions				Patient/WSI-level performance evaluation						Pathologist 1 vs DL model	Pathologist 2 vs DL model
	TP	FP	TN	FN	Accuracy	Precision	Recall	F1 score	Specificity	AUC ^a		
Pathologist 1	11	5	2	3	0.62	0.69	0.79	0.73	0.29	0.54	$P = .043$ (Wilcoxon signed-rank test)	$P = .031$ (Wilcoxon signed-rank test)
DL model	11	2	5	3	0.76	0.85	0.79	0.81	0.71	0.75		
Pathologist 2	10	5	2	4	0.57	0.67	0.71	0.69	0.29	0.51		
Agreement (Cohen kappa score)	Pathologist 1 vs pathologist 2				DL model vs pathologist 1			DL model vs pathologist 2				
	0.631				0.021			−0.061				

TP, FP, TN, and FN describe various outcomes of classification predictions.

DL, deep learning; FN, false negative; FP, false positive; ROC-AUC, area under the receiver operating characteristic curve; TN, true negative; TP, true positive; WSI, whole-slide image.

^a ROC-AUC was computed based on the binary outcomes of the patient-level classifications.

LCNEC," resemble those of SCLC cases characterized by RB1 inactivation, we applied our classification model to the external cohort of SCLC specimens. Consistent with expectations, the best-performing model identified 89.57% of the SCLCs as having pRb lost histomorphologic features (Table 3). Furthermore, the model assigned the pRb lost class with higher certainty to the SCLC cases than the retained class. Further details on the performance of all folds can be found in Supplementary Table S8. By evaluating the example SCLC tiles in Figure 4, we can conclude that predicted pRb lost tiles have a more compact arrangement and smaller nuclei than the tiles classified as pRb retained.

Qualitative Evaluation of the Proposed Model's Prediction

For the qualitative assessment of cytomorphologic features using Grad-CAM, samples from both the validation and the test set

were analyzed. Figure 5 presents a selection of resection TMA cores, whereas Figure 6 highlights selected regions of interest from biopsy specimens. Alongside the Grad-CAM activation maps, we also indicate the predicted class for each tile, with color intensity reflecting the model's confidence in its prediction. At higher magnifications, predicted pRb lost cases were found to be characterized by condensed, sometimes elongated, nuclei, with absent or inconspicuous nucleoli and scant cytoplasm (Figs. 5A and 6A). The model primarily focused on the lack of cytoplasm and nuclear shape rather than on intranuclear features. In contrast, predicted pRb retained cases exhibited an open chromatin pattern with prominent nucleoli and more voluminous, eosinophilic cytoplasm, with the model concentrating on nuclear morphology (Figs. 5B and 6B).

Interestingly, in Figure 6C, which shows a pRb retained case correctly predicted by the DL model but not by the pathologists,

Table 2

Overview of the pRb-based classification of LCNEC in H&E-stained biopsy specimens by the proposed deep learning model and 2 pathology specialists

Sample	Label	Tile wise			Patient wise			Pathologist evaluation	
		DL model evaluation						Pathologist 1	Pathologist 2
		Total no.	Predicted 1	Predicted 0	Probability 1	Probability 0	DL model		
LCNEC1	0	1439	524	915	0.36	0.64	0	1	1
LCNEC2	0	2983	509	2474	0.17	0.83	0	0	0
LCNEC3	0	727	608	119	0.84	0.16	1	1	1
LCNEC4	0	1064	62	1002	0.06	0.94	0	1	1
LCNEC5	0	91	45	46	0.49	0.51	0	0	0
LCNEC6	0	68	16	52	0.24	0.76	0	1	1
LCNEC7	0	61	34	27	0.56	0.44	1	1	1
LCNEC8	1	105	54	51	0.51	0.49	1	1	1
LCNEC9	1	65	54	11	0.83	0.17	1	0	1
LCNEC10	1	221	215	6	0.97	0.03	1	0	0
LCNEC11	1	94	63	31	0.67	0.33	1	0	0
LCNEC12	1	600	600	0	1.00	0.00	1	1	1
LCNEC13	1	251	220	31	0.88	0.12	1	1	1
LCNEC14	1	526	10	516	0.02	0.98	0	1	1
LCNEC15	1	18	15	3	0.83	0.17	1	1	0
LCNEC16	1	254	253	1	1.00	0.00	1	1	1
LCNEC17	1	31	2	29	0.06	0.94	0	1	1
LCNEC18	1	44	23	21	0.52	0.48	1	1	1
LCNEC19	1	127	127	0	1.00	0.00	1	1	0
LCNEC20	1	102	44	58	0.43	0.57	0	1	1
LCNEC21	1	7	7	0	1.00	0.00	1	1	1

0 represents pRb lost/SCLC-like LCNEC and 1 represents pRb retained/NSCLC-like LCNEC.

DL, deep learning; H&E, hematoxylin and eosin; LCNEC, large-cell neuroendocrine carcinoma; NSCLC, non-small cell lung carcinoma; pRb, retinoblastoma protein; SCLC, small cell lung carcinoma.

Table 3

Quantitative evaluation of the pRb classification of SCLC cases by the deep learning model

SCLC data set	Tiles (n = 7988)		Patients (n = 115)	
	pRb lost	pRb retained	pRb lost	pRb retained
No.	6980 (87.38%)	1008 (12.62%)	103 (89.57%)	12 (10.43%)
Mean certainty	0.91 $\pm$ 0.19	0.69 $\pm$ 0.32	0.87 $\pm$ 0.23	0.52 $\pm$ 0.35
Median certainty	1.00	0.84	0.99	0.51

pRb, retinoblastoma protein; SCLC, small cell lung carcinoma.

the exact features that led the model to assign a retained or lost class are unclear. However, we hypothesize that nuclear size and variability in nuclear shapes and textures played a role in the model's decision. This aligns with our broader observations that SCLC-like LCNEC exhibits more compact cellular arrangements, whereas NSCLC-like LCNEC tends to contain pleomorphic cells. This reasoning may also explain the misclassification of the pRb lost case in [Figure 5C](#), in which a variety of cellular shapes are evident.

Discussion

The molecular subtyping of LCNEC through loss or retention of pRb expression, distinguished as SCLC-like and NSCLC-like LCNEC, can guide appropriate therapeutic choices. In medical practice, pRb expression is typically assessed through IHC. However, the diagnosis of patients with LCNEC with inoperable disease often relies on small tumor biopsies, which may not always provide conclusive histologic and molecular subtyping, sometimes necessitating larger or repeated biopsies.^{30,31} This work represents a pioneering effort to explore alternative methods for subclassifying LCNEC in small specimens using DL. In validation experiments using resection TMA cores, the proposed model reached a patient-wise balanced accuracy value of 0.76 and an ROC-AUC value of 0.85. Moreover, applied solely on biopsies, our

model achieved a patient-wise balanced accuracy value of 0.75 and an ROC-AUC value of 0.77. It is important to note that, in this study, the model significantly outperformed both pathologists in subclassifying LCNEC cases based solely on H&E histomorphology (P value < .05). The pathologists achieved ROC-AUC values of 0.54 and 0.51. This suggests that the DL model may have identified discriminative features that are not easily detectable by the human eye, particularly when distinguishing pRb lost cases.

In addition, our study aims to provide explainable AI insights to pathologists to assist them during biomarker identification. By predicting the 2 major molecular subtypes of LCNEC and analyzing their relevant features, we sought to understand the morphologic differences that define these therapeutically significant subtypes, potentially gaining deeper biological insights. This approach could eventually enable morphology-based subtyping, reducing the reliance on IHC or molecular diagnostics, thus decreasing the turnaround time and costs in routine diagnostics. Based on features highlighted using Grad-CAM, SCLC-like LCNEC appears to be typified by compact cells with condensed nuclei and scant cytoplasm, whereas NSCLC-like LCNEC may feature pleomorphic cells with open chromatin, prominent nucleoli, and more cytoplasm. However, it is possible that the DL model detects specific features that are imperceptible to the human eye, which may not directly correspond to recognizable morphologic traits. This assumption is supported by [Figure 6C](#), in which the correct class prediction for the sample

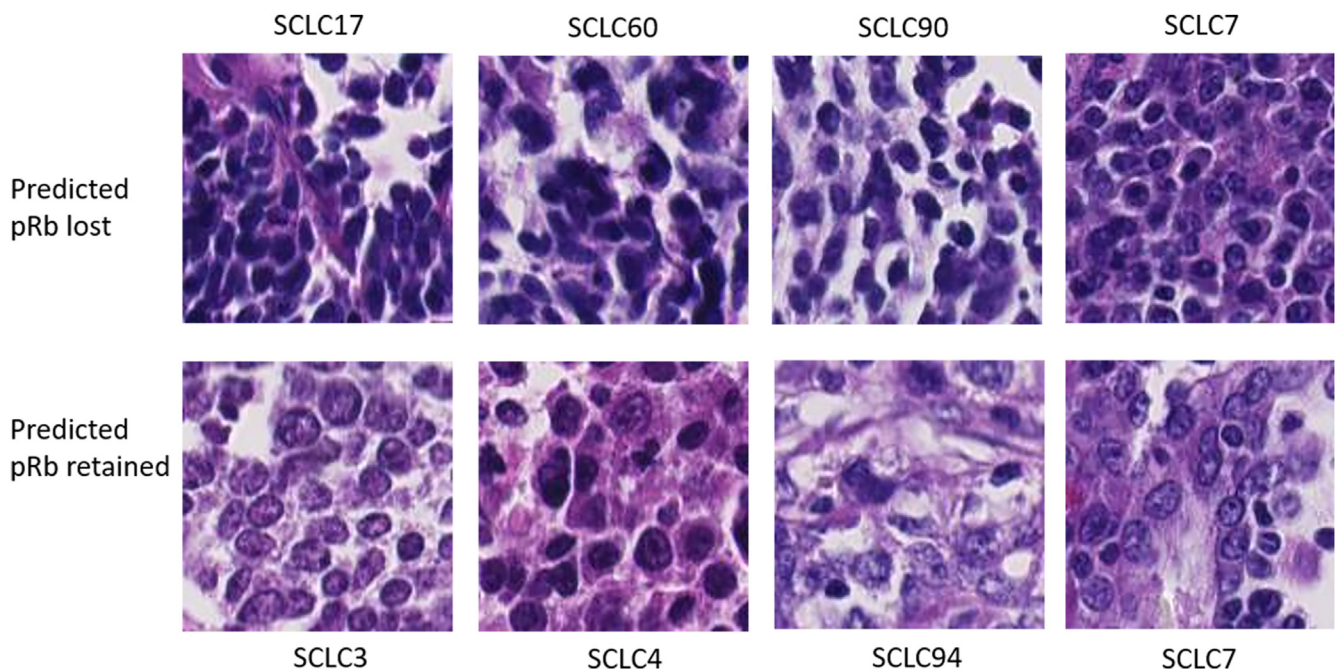
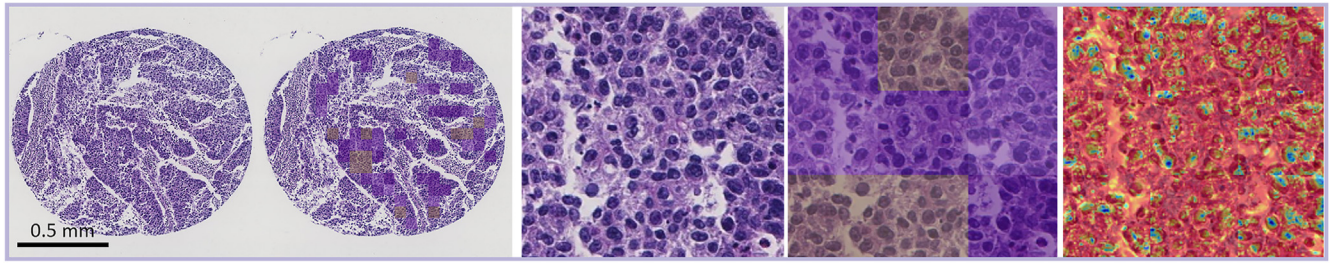
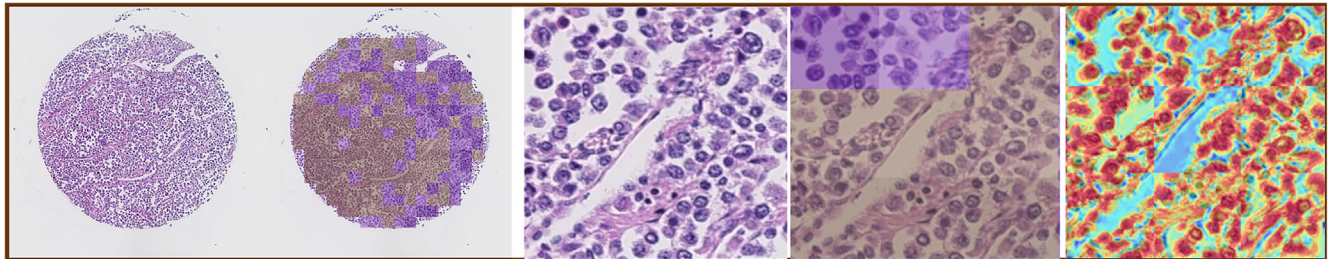
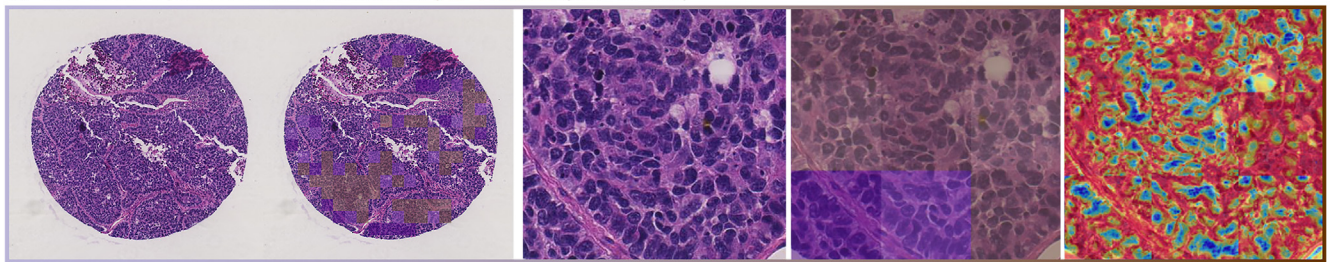


Figure 4. retinoblastoma protein (pRb) classification of small cell lung carcinoma (SCLC) cases using the deep learning model. Qualitative view on tiles assigned to the pRb lost and the pRb retained classes.

A IHC-defined pRb lost LCNEC classified as pRb lost (81% of tiles)**B** IHC-defined pRb retained LCNEC classified as pRb retained (61% of tiles)**C** IHC-defined pRb lost LCNEC classified as pRb retained (58% of tiles)

Probability predictions

pRb lost

0.5

1

pRb retained

0.5

1

Activation intensity

0

1

Figure 5.

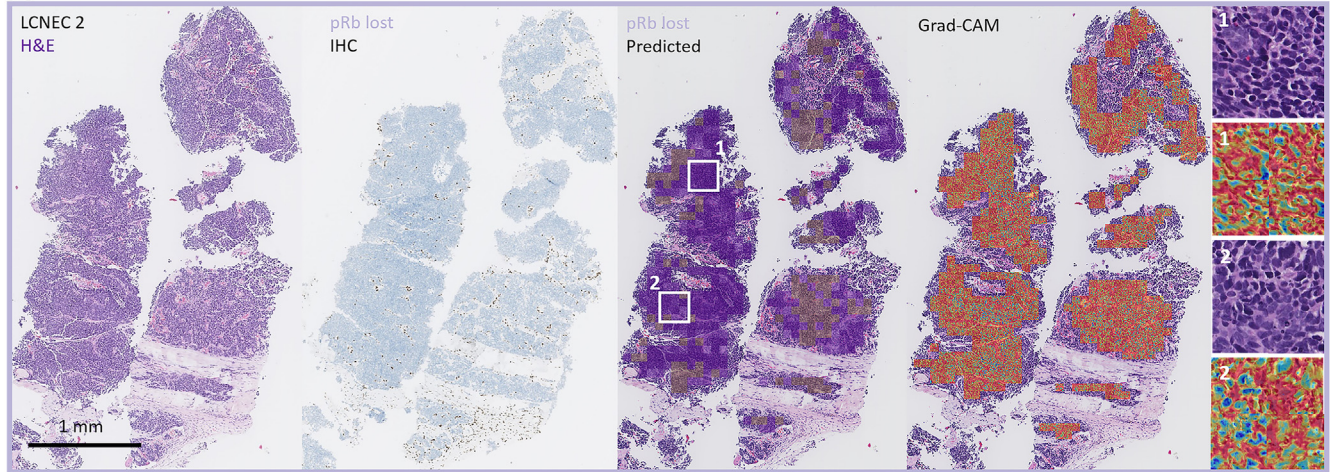
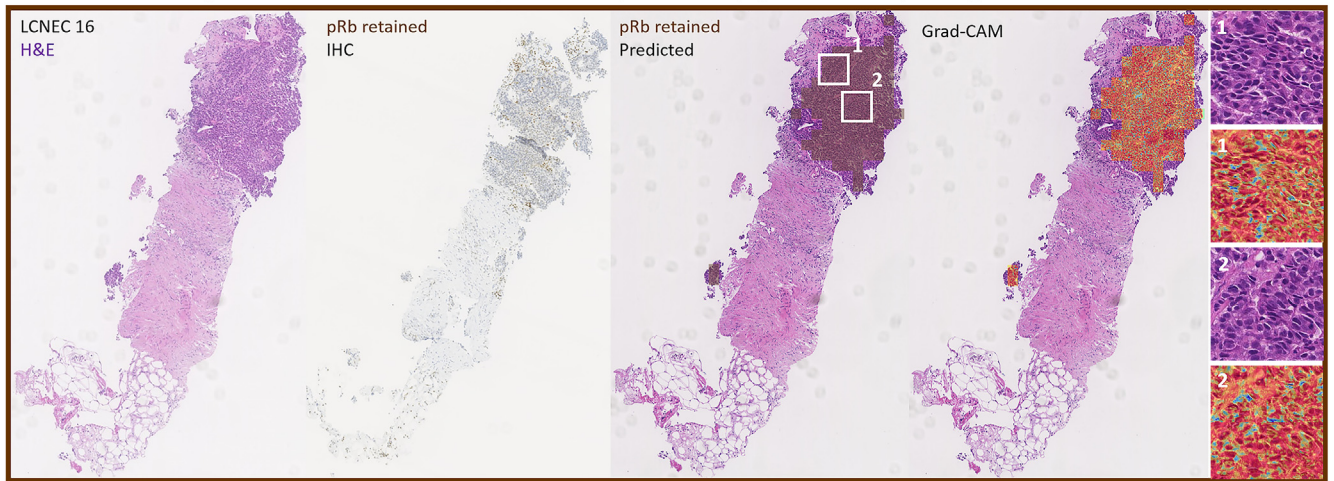
Qualitative evaluation of the deep learning predictions in 3 representative large-cell neuroendocrine carcinoma (LCNEC) specimens from resection tissue microarrays. (A) IHC-defined lost LCNEC correctly classified as pRb lost. (B) IHC-defined pRb retained LCNEC correctly classified as pRb retained. (C) IHC-defined pRb lost LCNEC incorrectly classified as pRb retained. The highlighted regions on the gradient-weighted class activation mapping heatmap indicate the image features that the model associated with the predicted class and contributed to the decision. Red represents “focus” areas, whereas blue represents noncritical features. IHC, immunohistochemistry; pRb, retinoblastoma protein.

lacks clear, attributable features, suggesting that the model may rely on subtler or less apparent characteristics.

To strengthen the conclusions of Grad-CAM evaluation, our framework was applied to a cohort of resections from 115 patients with SCLC. The goal of this analysis was to determine if the features identified as “SCLC-like” LCNEC, marked by lost pRb expression, are more closely associated with SCLC histology than “NSCLC-like” LCNEC. If this assumption were valid, applying the model to SCLC data would result in the majority of cases being classified as pRb lost. Our hypothesis was confirmed, with 89.57% of the SCLC tiles associated with the pRb lost class. This aligns not only with the confirmation by pRb provided by the data but also with the 6% to 10% RB1 positivity in SCLC.^{23,32} Furthermore, the model assigned the pRb lost class to the SCLC cases with higher certainty than the pRb retained class, indicating that the pRb lost features are more representative of the SCLC cohort (Fig. 4). Feature wise, the model consistently focused on the variability in nuclear detail, and the amount and color of the cytoplasm, which clearly differed between the 2 classes. Regarding the morphology of these pRb lost LCNEC cases, we could speculate that these cases represent a separate subclass with an intermediate morphology between LCNEC and SCLC, as observed in genomic evaluation of

LCNEC.⁸ These cases could potentially represent neuroendocrine carcinoma of intermediate-sized cells, a term used in the past.³³

Regarding the training process, the variability of the model performances across all folds highlights the intraclass heterogeneity and interclass similarities of the data set, which are exacerbated in a small data set such as ours (Supplementary Figs. S2-S4 and Supplementary Table S7). These limitations, which affect the robustness of the model, have led to an overall test ROC-AUC value of 0.67 ± 0.11 and an ROC-AUC value of 0.62 ± 0.11 on biopsy samples. Especially in the random split from the first fold, the train set did not identify sufficient features to enable the generalization to the test set. Furthermore, although the settings from the second, third, and fourth folds performed well on tissues derived from resections, their performance decreased when generalized to biopsy tissues. This is likely due to structural changes caused by needle-based tissue extraction. As the model was trained solely on resection data, it should be further trained on additional biopsy data to improve its generalizability and account for features specific to biopsies. Furthermore, increasing the training data and adapting the CNN architecture to the size and complexity of the extended cohort are expected to improve the predictive performance of the classification,

A ROI of IHC-defined pRb lost LCNEC classified as pRb lost (83% of tiles) by DL model and pathologists**B** ROI of IHC-defined pRb retained LCNEC classified as pRb retained (100% of tiles) by DL model and pathologists**C** ROI of IHC-defined pRb retained LCNEC classified as pRb retained (67% of tiles) by DL model and as pRb lost by pathologists**Figure 6.**

Qualitative evaluation of the deep learning (DL) predictions in 3 representative regions of interest from large-cell neuroendocrine carcinoma (LCNEC) biopsy specimens. The highlighted regions on the gradient-weighted class activation mapping (Grad-CAM) heatmap indicate the image features that the model associated with the predicted class and contributed to the decision. (A) ROI of IHC-defined lost LCNEC correctly classified as pRb lost. (B) ROI of IHC-defined pRb retained LCNEC correctly classified as pRb retained. (C) ROI of IHC-defined pRb retained LCNEC correctly classified as pRb retained. Red represents "focus" areas, whereas blue represents noncritical features. H&E, hematoxylin and eosin; IHC, immunohistochemistry; pRb, retinoblastoma protein; ROI, region of interest.

promoting its clinical applicability as an alternative to the IHC pRB evaluation. Such tools could eventually contribute to updates in diagnostic frameworks involving biopsies, such as the WHO classification system, and provide accessible, cost-effective, and significantly faster solutions for clinics, by eliminating the need for antibodies and staining. This approach could prevent tissue scarcity and be particularly advantageous for resource-limited settings in which advanced IHC capabilities are unavailable.

Grad-CAM feature evaluations and the variability of each fold's performance highlight the data heterogeneity, aligning with clinical studies that emphasize the complexity of distinguishing between SCLC and LCNEC.^{34,35} Interobserver variability among pathologists is reported as only fair ($\kappa = 0.40$). In the study by den Bakker et al.,³⁴ which examined a mix of SCLC, LCNEC, and neuroendocrine lung carcinoma cases, unanimous diagnoses were reached in only 20 out of 170 cases, and a majority diagnosis was achieved in 115 cases, whereas no consensus was reached in 35 cases.

Future research should focus on improving the classifier's robustness by integrating data from international cohorts, which would help overcome current limitations related to the rarity of LCNEC and tissue heterogeneity. Expanding the training data set and thereby increasing the complexity of the captured morphologic features will likely allow for the development of deeper and more sophisticated CNN architectures, including semisupervised approaches. This, in turn, would enhance the model's ability to capture finer, localized histomorphologic features that are crucial for accurate subtyping, possibly pinpointing to more precise regions. These features may eventually inform updates to the WHO classification system using biopsies, enhancing its diagnostic and prognostic use for inoperable cases.

Author Contributions

T.E.T., F.A., J.L.D., A.-M.C.D., and J.H.v.d.T. conceived and designed the study. T.E.T., F.A., Y.L., and A.P.S. developed the methodology of the study. T.E.T., F.W.J.H., L.M.H., and J.H.v.d.T. performed data annotation. T.E.T., F.W.J.H., F.A., J.L.D., and J.H.v.d.T. performed writing, review, and revision of the article. T.E.T., F.W.J.H., F.A., L.M.H., Y.L., and J.H.v.d.T. performed acquisition, analysis, and interpretation of data, and statistical analysis. E.-J.M.S., A.-M.C.D., and J.H.v.d.T. supervised the study. Z.M. and B.D. provided material support. All authors read and approved the final article.

Data Availability

All data (including histologic images and annotations) are available upon reasonable request. Source codes are publicly available at ErasmusMC-Bioinformatics/DL_LCNEC_pRBsubtyping_HE (github.com). Relevant data information and figures can be found in the supplementary section.

Funding

This research was supported by an unrestricted grant of Stichting Hanarth Fonds, the Netherlands. B.D. was supported by the Austrian Science Fund (FWF I3522, FWF I3977, and I4677) and the "BIOSMALL" EU HORIZON-MSCA-2022-SE-01 project. B.D. and Z.M. were supported by funding from the Hungarian National Research, Development, and Innovation Office (2020-1.1.6-JÖVŐ, TKP2021-EGA-33, FK-143751, and FK-147045). Z.M. was supported

by the New National Excellence Program of the Ministry for Innovation and Technology of Hungary (UNKP-20-3, UNKP-21-3, and UNKP-23-5) and by the Bolyai Research Scholarship of the Hungarian Academy of Sciences. Z.M. is also the recipient of the International Association for the Study of Lung Cancer/International Lung Cancer Foundation Young Investigator Grant (2022).

Declaration of Competing Interest

E.-J.M.S. reports receiving institutional fees from Bayer, AstraZeneca, Janssen, GSK, Pfizer, and Merck, outside the current work. A.-M.C.D. reports receiving institutional fees from Roche, Eli Lilly, Boehringer Ingelheim, AstraZeneca, Janssen, Chiezi, Amgen, Pfizer, Bayer, Takeda, Pharmamar, Sanofi, and Daiichi, outside the current work. J.H.v.d.T. reports receiving advisory board and speakers' fees from Eli Lilly, BMS, MSD, AstraZeneca, Bayer, Janssen, Pfizer, outside the current work. All other authors report no disclosures.

Ethics Approval and Consent to Participate

Patients provided written informed consent to access tissue blocks/slides and medical records. The study protocol was approved by the medical ethical committee of the Erasmus Medical Center, Rotterdam, the Netherlands (MEC-2022-0130).

Supplementary Material

The online version contains supplementary material available at <https://doi.org/10.1016/j.labinv.2025.104192>

References

- Roesel C, Welter S, Kambartel KO, et al. Prognostic markers in resected large cell neuroendocrine carcinoma: a multicentre retrospective analysis. *J Thorac Dis.* 2020;12(3):466–476.
- Travis WD, Burke AP, Marx A, Nicholson AG. *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart.* 4th edition. World Health Organization; 2021.
- Lindsay CR, Shaw EC, Moore DA, et al. Large cell neuroendocrine lung carcinoma: consensus statement from The British Thoracic Oncology Group and the Association of Pulmonary Pathologists. *Br J Cancer.* 2021;125(9):1210–1216.
- Nicholson AG, Tsao MS, Beasley MB, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. *J Thorac Oncol.* 2022;17(3):362–387.
- Baine MK, Sinard JH, Cai G, Homer RJ. A semiquantitative scoring system may allow biopsy diagnosis of pulmonary large cell neuroendocrine carcinoma. *Am J Clin Pathol.* 2020;153(2):165–174.
- Derks JL, Dingemans AC, van Suylen RJ, et al. Is the sum of positive neuroendocrine immunohistochemical stains useful for diagnosis of large cell neuroendocrine carcinoma (LCNEC) on biopsy specimens? *Histopathology.* 2019;74(4):555–566.
- Rekhtman N, Pietanza MC, Hellmann MD, et al. Next-generation sequencing of pulmonary large cell neuroendocrine carcinoma reveals small cell carcinoma-like and non-small cell carcinoma-like subsets. *Clin Cancer Res.* 2016;22(14):3618–3629.
- George J, Walter V, Peifer M, et al. Integrative genomic profiling of large-cell neuroendocrine carcinomas reveals distinct subtypes of high-grade neuroendocrine lung tumors. *Nat Commun.* 2018;9(1):1048.
- George J, Lim JS, Jang SJ, et al. Comprehensive genomic profiles of small cell lung cancer. *Nature.* 2015;524(7563):47–53.
- Andrini E, Marchese PV, De Biase D, et al. Large cell neuroendocrine carcinoma of the lung: current understanding and challenges. *J Clin Med.* 2022;11(5):1461.
- Derks JL, Leblay N, Thunnissen E, et al. Molecular subtypes of pulmonary large-cell neuroendocrine carcinoma predict chemotherapy treatment outcome. *Clin Cancer Res.* 2018;24(1):33–42.
- Zhuo M, Guan Y, Yang X, et al. The prognostic and therapeutic role of genomic subtyping by sequencing tumor or cell-free DNA in pulmonary large-cell neuroendocrine carcinoma. *Clin Cancer Res.* 2020;26(4):892–901.

13. van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. *Nat Med*. 2021;27(5):775–784.
14. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med*. 2018;24(10):1559–1567.
15. Yang H, Chen L, Cheng Z, et al. Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: a retrospective study. *BMC Med*. 2021;19(1):80.
16. Tomita N, Tafe LJ, Suriawinata AA, et al. Predicting oncogene mutations of lung cancer using deep learning and histopathologic features on whole-slide images. *Transl Oncol*. 2022;24:101494.
17. Terada K, Yoshizawa A, Liu X, et al. Deep learning for predicting effect of neoadjuvant therapies in non-small cell lung carcinomas with histologic images. *Mod Pathol*. 2023;36(11):100302.
18. Zhang Y, Yang Z, Chen R, et al. Histopathology images-based deep learning prediction of prognosis and therapeutic response in small cell lung cancer. *NPJ Digit Med*. 2024;7(1):15.
19. Yang JW, Song DH, An HJ, Seo SB. Classification of subtypes including LCNEC in lung cancer biopsy slides using convolutional neural network from scratch. *Sci Rep*. 2022;12(1):1830.
20. Ilić M, Benzaquen J, Tourniaire P, et al. Deep learning facilitates distinguishing histologic subtypes of pulmonary neuroendocrine tumors on digital whole-slide images. *Cancers (Basel)*. 2022;14(7):1740.
21. Gonzalez D, Dietz RL, Pantanowitz L. Feasibility of a deep learning algorithm to distinguish large cell neuroendocrine from small cell lung carcinoma in cytology specimens. *Cytopathology*. 2020;31(5):426–431.
22. Megyesfalvi Z, Barany N, Lantos A, et al. Expression patterns and prognostic relevance of subtype-specific transcription factors in surgically resected small-cell lung cancer: an international multicenter study. *J Pathol*. 2022;257(5):674–686.
23. Febres-Aldana CA, Chang JC, Ptashkin R, et al. Rb tumor suppressor in small cell lung cancer: combined genomic and IHC analysis with a description of a distinct Rb-proficient subset. *Clin Cancer Res*. 2022;28(21):4702–4713.
24. Beasley MB, Lantuejoul S, Abbondanzo S, et al. The P16/cyclin D1/Rb pathway in neuroendocrine tumors of the lung. *Hum Pathol*. 2003;34(2):136–142.
25. Papaxoinis G, Bille A, McLean E, Nonaka D. Comparative study of Rb1, cyclin D1 and p16 immunohistochemistry expression to distinguish lung small-cell carcinoma and large-cell neuroendocrine carcinoma. *Histopathology*. 2022;81(2):205–214.
26. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):16878.
27. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. 1979;9(1):62–66.
28. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444.
29. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, editors. Grad-CAM: visual explanations from deep networks via gradient-based localization. Paper presented at: 2017 IEEE International Conference on Computer Vision (ICCV); October 22–29, 2017; Venice, Italy.
30. Derks JL, Hendriks LE, Buikhuisen WA, et al. Clinical features of large cell neuroendocrine carcinoma: a population-based overview. *Eur Respir J*. 2016;47(2):615–624.
31. Hermans BCM, Sanduleanu S, Derks JL, et al. Exploring imaging features of molecular subtypes of large cell neuroendocrine carcinoma (LCNEC). *Lung Cancer*. 2020;148:94–99.
32. McColl K, Wildey G, Sakre N, et al. Reciprocal expression of INSM1 and YAP1 defines subgroups in small cell lung cancer. *Oncotarget*. 2017;8(43):73745–73756.
33. Moran CA, Suster S, Coppola D, Wick MR. Neuroendocrine carcinomas of the lung: a critical analysis. *Am J Clin Pathol*. 2009;131(2):206–221.
34. den Bakker MA, Willemsen S, Grünberg K, et al. Small cell carcinoma of the lung and large cell neuroendocrine carcinoma interobserver variability. *Histopathology*. 2010;56(3):356–363.
35. Ha SY, Han J, Kim WS, Suh BS, Roh MS. Interobserver variability in diagnosing high-grade neuroendocrine carcinoma of the lung and comparing it with the morphometric analysis. *Korean J Pathol*. 2012;46(1):42–47.