Human Pathology (2021) **107**, 58–68

ELSEVIER

Human PATHOLOGY

www.elsevier.com/locate/humpath

Original contribution

Artificial intelligence identifies inflammation and confirms fibroblast foci as prognostic tissue biomarkers in idiopathic pulmonary fibrosis☆

Kati Mäkelä MD [a,*], Mikko I. Mäyränpää MD, PhD [b], Hanna-Kaisa Sihvo DVM, PhD, DECVP [c], Paula Bergman MSc [d], Eva Sutinen MSc [a], Hely Ollila BM [a], Riitta Kaarteenaho MD, PhD [e], Marjukka Myllärniemi MD, PhD [a]

[a] *Individualized Drug Therapy Research Program, Faculty of Medicine, University of Helsinki and Heart and Lung Center, Helsinki University Hospital, FI-00290, Helsinki, Finland*
[b] *Pathology, University of Helsinki and Helsinki University Hospital, FI-00290, Helsinki, Finland*
[c] *Aiforia Technologies, FI-00290, Helsinki, Finland*
[d] *Biostatistics Consulting, Department of Public Health, University of Helsinki and Helsinki University Hospital, FI-00290, Helsinki, Finland*
[e] *Research Unit of Internal Medicine, University of Oulu and Medical Research Center Oulu, Oulu University Hospital, FI-90014, Oulu, Finland*

Received 24 August 2020; revised 27 October 2020; accepted 30 October 2020

Available online 5 November 2020

Keywords:
Idiopathic pulmonary fibrosis;
Usual interstitial

Summary A large number of fibroblast foci (FF) predict mortality in idiopathic pulmonary fibrosis (IPF). Other prognostic histological markers have not been identified. Artificial intelligence (AI) offers a possibility to quantitate possible prognostic histological features in IPF. We aimed to test the use of AI in IPF lung tissue samples by quantitating FF, interstitial mononuclear inflammation, and intra-

☆ competing interests: The use of Aiforia® platform was funded with a grant from Orion Pharma. K.M. reports grants from the Foundation of the Finnish Anti-Tuberculosis Association, the Research Foundation of the Pulmonary Diseases, the Väinö and Laina Kivi Foundation, the University of Helsinki, the Biomedicum Helsinki Foundation, and the Finnish Medical Society and lecture fees from Roche and Boehringer Ingelheim. M.I.M. reports receiving a grant from the Finnish Medical Foundation, remuneration for congress and meeting travel costs from Pfizer, Roche, Bristol-Myers Squibb, and MSD, and advisory board and lecture fees from Boehringer Ingelheim, Bristol-Myers Squibb, MSD, and Takeda. H.-K.S. reports being an employee of Aiforia Technologies. H.O. reports receiving grants from the Väinö and Laina Kivi Foundation and the Tampere Tuberculosis Foundation. R.K. reports receiving grants for the research group from the Foundation of the Finnish Anti-Tuberculosis Association, the Research Foundation of the Pulmonary Diseases, the Research Foundation of North Finland, and the Jalmari and Rauha Ahokas Foundation, a state subsidy from Oulu University Hospital, remuneration for a congress travel cost from Orion, lecture fees from Roche, Boehringer Ingelheim, and Ratiopharm, and an advisory board fee from Boehringer Ingelheim. M.M. reports receiving grants for the FinnishIPF study from the Sigrid Jusélius Foundation, the Helsinki University Hospital funds, the Nummela Sanatorium Foundation, and Boehringer Ingelheim. The other authors declare that they have no duality of interest. The funding sources of the study had no role in study design, in the collection, analysis, or interpretation of the data, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

* Corresponding author. Transplantation Laboratory, Pulmonary Diseases, B410b, Haartman Institute, Haartmaninkatu 3, FI-00290, Helsinki, Finland.
E-mail address: kati.makela@helsinki.fi (K. Mäkelä).

https://doi.org/10.1016/j.humpath.2020.10.008
0046-8177/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

pneumonia;
Inflammation;
Fibroblast focus;
Artificial intelligence;
Deep neural network

alveolar macrophages with a deep convolutional neural network (CNN). Lung tissue samples of 71 patients with IPF from the FinnishIPF registry were analyzed by an AI model developed in the Aiforia® platform. The model was trained to detect tissue, air spaces, FF, interstitial mononuclear inflammation, and intra-alveolar macrophages with 20 samples. For survival analysis, cut-point values for high and low values of histological parameters were determined with maximally selected rank statistics. Survival was analyzed using the Kaplan-Meier method. A large area of FF predicted poor prognosis in IPF ($p = 0.01$). High numbers of interstitial mononuclear inflammatory cells and intra-alveolar macrophages were associated with prolonged survival ($p = 0.01$ and $p = 0.01$, respectively). Of lung function values, low diffusing capacity for carbon monoxide was connected to a high density of FF ($p = 0.03$) and a high forced vital capacity of predicted was associated with a high intra-alveolar macrophage density ($p = 0.03$). The deep CNN detected histological features that are difficult to quantitate manually. Interstitial mononuclear inflammation and intra-alveolar macrophages were novel prognostic histological biomarkers in IPF. Evaluating histological features with AI provides novel information on the prognostic estimation of IPF.
© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Despite the recent update of the histological criteria of idiopathic pulmonary fibrosis (IPF) [1], few prognostic histological factors have been identified thus far. Fibroblast foci (FF) are key histological features in IPF, which manifests histologically as the usual interstitial pneumonia (UIP) pattern. High numbers of FF have been associated with worse outcomes for patients with IPF in several studies [2–7], but some controversial results have also been published [8–11].

Previously, inflammatory cells such as T lymphocytes and intra-alveolar macrophages were considered essential for the pathogenesis of IPF [12]. Currently, repetitive alveolar injury and fibrotic repair of damaged tissue are suggested to be hallmarks of IPF pathogenesis, whereas inflammation has been considered an epiphenomenon [13]. Mild inflammation can exist in the UIP pattern, but it is deemed an atypical feature [1]. The evidence for inflammation in IPF pathogenesis is, however, controversial [13]. We have previously noted that abundant interstitial inflammation is a common finding in lung tissue samples of carefully re-evaluated patients with IPF [14]. In addition, interobserver variation is substantial in the detection of inflammatory cells [14,15]. Besides interstitial inflammation, intra-alveolar macrophages have been detected in IPF samples [9,16], but very little is known of their clinical or biological significance. Quantitating inflammatory cells manually from lung tissue samples is time consuming, inaccurate, and subject to intraobserver and interobserver variation, providing an explanation for the scant histological studies on inflammation in IPF.

The development of artificial intelligence (AI) enables new approaches to image analysis. AI models have been shown to recognize histological UIP pattern by using genomic data from lung biopsies [17–19]. Radiological findings can also be quantitated using automated image analysis and have been associated with pulmonary function, survival, and response to antifibrotic medication [20–22]. In a manner comparable to radiologists, an AI model can classify fibrotic lung diseases according to high-resolution computed tomography images [23]. AI models have been used in the histology of experimental mouse models of pulmonary fibrosis [24–26]. To our knowledge, histological features of IPF samples have not been previously studied using automated image analysis. Before developing diagnostic AI models for the UIP pattern, the ability of AI to identify specific histological features should be tested.

We hypothesize that automated image analysis can count both interstitial and intra-alveolar inflammatory cells in IPF lung tissue and that the numbers of inflammatory cells have a prognostic value in IPF. We also aimed to test the previous association between FF and prognosis of patients with IPF using the automated image analysis. Our approach was to pilot an AI model with a small data set and test its generalizability in slides that were not included in the training data set. Using lung tissue samples of thoroughly characterized patients from the FinnishIPF registry patients [27], we developed the AI model with a deep convolutional neural network (CNN) in the Aiforia® platform (Aiforia Technologies, Helsinki, Finland). Of the data produced by the AI model, we analyzed the prognostic significance of FF, interstitial mononuclear inflammation, and intra-alveolar macrophages.

2. Materials and methods

2.1. Study population

The study population originated from the FinnishIPF registry, which is a prospective, multicentre study of patients with IPF [27]. Respiratory medicine specialists or multidisciplinary teams have re-evaluated diagnoses according to the 2011 international diagnostic guidelines for

Table 1 Patient characteristics.

	Mean ± SD or %	*n*
Age at diagnosis (years)	61.5 ± 10.4	71
Age at sample collection (years)	62.3 ± 10.0	71
Age at death (years)	70.5 ± 8.2	37
Age at transplantation (years)	56.8 ± 8.3	17
Deaths (%)	52.1	37
Lung transplant recipients (%)	23.9	17
Follow-up time (months)	72.5 ± 42.7	71
Sex		
Male (%)	69.0	49
Female (%)	31.0	22
Smoking at diagnosis		
Never (%)	36.6	26
Ex-smoker (%)	47.9	34
Current smoker (%)	15.5	11
Pack-years of smoking	22.3 ± 12.4	40
BMI (kg/m^2)	29.1 ± 4.8	66
FVC%	75.6 ± 16.7	66
DLCO%	56.3 ± 16.0	67
6MWT (m)	424.7 ± 170.5	17

The values were not available for all patients, and they were from the time of diagnosis. For follow-up time, death or lung transplantation was used as an endpoint event. Follow-up time for patients having no endpoints was defined as the time interval between IPF diagnosis date and 29 April 2019. Six-minute walk test (6MWT) was performed without extra oxygen. BMI, body mass index; FVC%, forced vital capacity of predicted; DLCO%, diffusing capacity for carbon monoxide of predicted; 6MWT, 6-min walk test.

IPF [27,28]. Participants have given written informed consent, and the study has been reviewed and approved by the Ethics Committees of the Finnish University Hospitals. In January 2017, all patients with hematoxylin and eosin (HE)−stained histopathological samples from all five university hospital districts were collected, resulting in 71 patients with IPF. The most representative slide revealing typical histopathological features for UIP was selected for each patient; of the 71 representative samples of patients, 62 were surgical lung biopsies (SLB, 87.3%), six explant samples (8.5%), and three autopsy samples (4.2%). Patient characteristics are shown in Table 1. This study included slides of the patients that have been analyzed in our previous studies by the 2011 IPF diagnostic criteria [14,28,29]. All available lung tissue samples of 60 patients of this study have been previously thoroughly analyzed by four pathologists [14]. The other eleven cases were collected after the study [14]; however, pathologists have evaluated all available lung tissue samples in the clinical setting. For 60 cases, the most representative slide was chosen by a consensus of two pathologists [14]. For the rest of the cases, K.M. selected the most representative slide.

2.2. Automated image analysis with the AI model

All slides were digitally scanned with bright field using Pannoramic 250 Flash II (3DHistech, Budapest, Hungary) at 40× magnification, 0.12 µm/pixel resolution, and 40X/0.95 NA objective. The scanned whole-slide images were uploaded to Aiforia® image management and analysis platform (Aiforia Technologies, Helsinki, Finland) and then analyzed with an AI model developed with a deep CNN and supervised learning. The AI model development followed the previously described workflow [30,31]. In short, after uploading the whole-slide digital images of the scanned histological slides to Aiforia® platform, K.M. reviewed all images and chose 20 images (28.2%, N = 71) for the AI model development. In the 20 images, K.M. manually annotated representative morphological areas in the images that were used for training the AI model. Examples of training areas and target feature annotations are shown in Appendix 1. In practice, K.M. freely panned and zoomed the whole-slide image and chose the representative areas of morphological features, where the training areas and feature labels were manually drawn. After each training round, the AI model results were evaluated visually, old annotations were edited, and new annotations were created. After that, a new training round was performed. After five rounds of adding and modifying the annotations, the final AI model was trained with 10 000 iterations. All 71 images were then analyzed with the final AI model.

For training the model, we chose the most representative 20 slides (28.2%, N = 71). By choosing only part of the slides as training data, we wanted to test if the AI model could be able to analyze slides that it has not previously encountered (51/71, 71.8%). Focusing on the sufficient level of variation and the quality of the training annotations with the iterative workflow of the model enabled the training of the AI model with a small training data set. We aimed at diverse training data; therefore, we chose slides representing the morphologic variation, different qualities, and staining intensities and from different laboratories. In the training data, we also included slides with artefacts, for example, blur, tissue folds, dust, and pathologist's color markers. We trained the AI model to separate artefacts from the wanted histological features. We included explants (n = 4) and autopsy samples (n = 2) for testing lung tissue recognition, but for the quantitative analysis of FF, interstitial mononuclear inflammation, and intra-alveolar macrophages, we focused on SLBs (n = 14).

We aimed to teach the AI model to recognize lung tissue, air spaces, FF, interstitial mononuclear inflammation, and intra-alveolar macrophages (Fig. 1). The receptive field size (field of view) of the CNN and the surface areas of the training annotations are shown in Table 2. As part of the

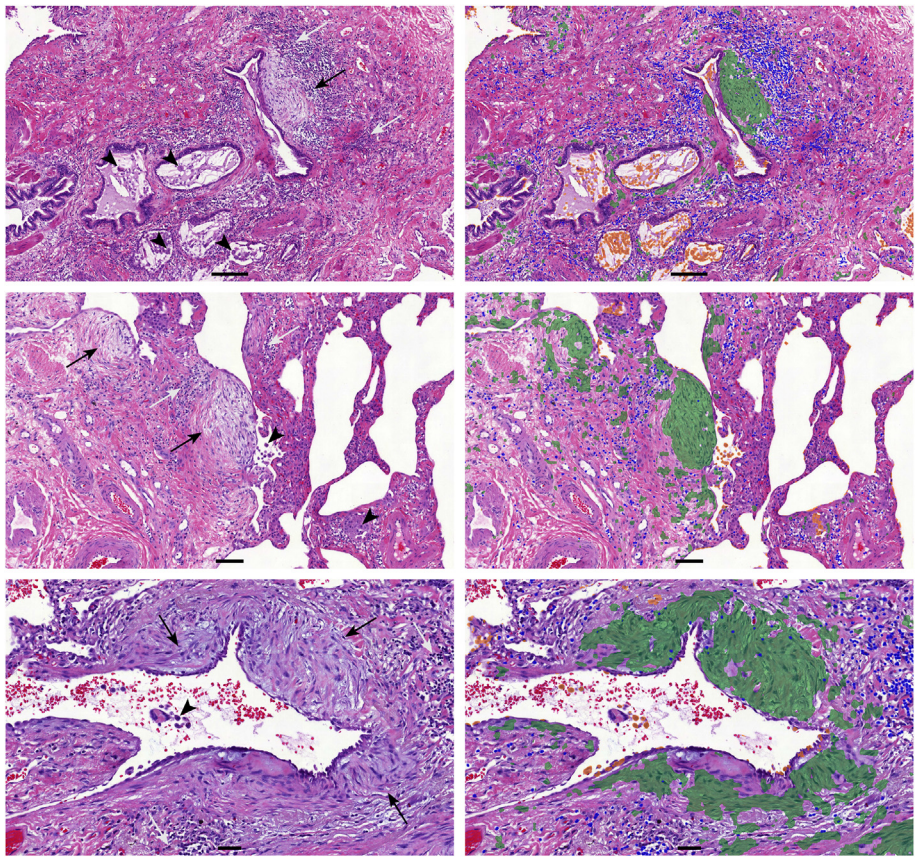

Fig. 1 Fibroblast foci (green mask), interstitial mononuclear inflammatory cells (blue mask), and intra-alveolar macrophages (orange mask) recognized by the artificial intelligence model. Black arrows point to examples of fibroblast foci, white arrows to interstitial mononuclear inflammatory cells, and black arrowheads to intra-alveolar macrophages. The scale bar in the first picture is 200 μm, in the second picture 100 μm, and in the third picture 50 μm. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

supervised learning, we annotated manually training areas of each feature (Appendix 1). The annotations created the model's ground truth, that is, the data of which the model learned to identify the target features. To avoid annotations conflicting with each other, we decided that only K.M. made all of the annotations. In the creation of the ground truth, K.M. consulted M.I.M., a pathologist experienced in pulmonary pathology, and H.-K.S., a veterinary pathologist experienced in the use of Aiforia®. The annotations were based purely on morphology. Consistency was the key element in the training of the AI model. We accepted only high-quality and high-confidence features as training data and let the model decide on borderline cases. For forming the ground truth, the model was taught target features by characteristic examples but also by numerous annotations that did not represent the feature. In training features with a characteristic morphology, we followed definitions of each feature described in Table 2. During the development of the model, we noticed that teaching the difference between the classical definition of FF and intraluminal fibrosis/organizing pneumonia (OP) was difficult. As FF and OP have quite a similar morphology, conflicting training, that is, similar features marking two different things, led to the malfunction of the model. As OP was a rare feature in our training data [14], teaching the difference with a sufficient amount of examples was not an option. Therefore, we accepted OP-like features as FF, as this enabled us the better function of the model. We also trained the AI model to exclude features that resembled our target features; for example, the model was trained that the loose fibrosis around the vessels, perivascular fibrosis, was not FF. After each iteration of the training, we visually evaluated the ability of the AI model to recognize wanted features from the training areas. Based on our observations on the visual results after each iteration of the training, we improved the training data by correcting annotation errors made earlier. We also expanded the training data by annotating features the model had not learned yet based on the visual feedback. K.M. reviewed the training results and improved the annotations with H.-K.S.

Table 2 The definition, the field size, and the sum of the surface area of training annotations of each histological feature taught to the artificial intelligence model.

	Definition	Field size (μm)	Training data (mm^2)
Lung tissue	The complete tissue area on the slide, including both the interstitium and the air spaces.	1000	836
Interstitium	The lung parenchyma without air spaces.	80	35
Air spaces	Areas inside bronchi, bronchioles, alveolar spaces, and honeycombing cysts.	80	17
Fibroblast foci	Myxoid, pale-staining subepithelial loose fibrosis also including intraluminal foci. Loose fibrosis away from alveolar areas and perivascular fibrosis were excluded.	20	1.6
Interstitial mononuclear inflammatory cells	Small cells with round, dark nuclei and scant cytoplasm assumed to be lymphocytes or plasma cells. Lymphocytes and plasma cells inside fibroblast foci were also counted. Interstitial inflammatory cells morphologically resembling neutrophils, eosinophils, macrophages, fibrocytes, or myocytes were not counted.	20	0.03
Intra-alveolar macrophages	Macrophages inside air spaces, including macrophages with both fine and coarse haemosiderin pigment, giant cells, and foamy macrophages. Intra-alveolar cells that were morphologically more suitable for red cells, lymphocytes, neutrophils, eosinophils, or epithelial cells were excluded.	10	0.07

The model consisted of four layers, each being an independent neural network, that were chained as an analysis pipeline. The first layer marked the lung tissue, and the second one separated air spaces from the interstitium. In the interstitium, the third layer recognized FF and interstitial mononuclear inflammation. In the air spaces, the fourth layer recognized intra-alveolar macrophages. The layer-based structure of the model enabled us to set the optimal field of view for each layer, based on the feature morphology. The total area of all whole-slide images in the training data set was 16 960 mm^2. Training areas covered 4.9% of that area in the layer of tissue, 0.3% in the layer of the interstitium and alveolar spaces, 0.06% in the layer of FF and interstitial mononuclear inflammatory cells, and 0.01% in the layer of intra-alveolar macrophages. As we had considerable variation represented in the training data and aimed to avoid the creation of conflicting data, we used only a minimal digital augmentation of the training data: size scaling between −1 and 1.01%, 1% aspect ratio change, 1% shear distortion, luminance change between −1 and 1%, contrast change between −1 and 1.01%, 1% white balance change, and noise level of 0 units.

After the development of the AI model, we analyzed all 71 samples with the AI model. The AI model produced data of the surface areas and counts of every histological feature. The area of each feature was quantified as a percentage in relation to the whole tissue area (area%), and the density of each feature was determined by dividing the counts by whole tissue area.

2.3. Validation of the AI model

Previous studies have demonstrated the prognostic value of FF [2−7] and the quantitation of FF using image analysis software [4,5,7]. Hence, we compared the FF detection by the AI model to the pathologist's manual annotations of FF in 30 validation areas. Of 71 slides included in the analysis, 51 were held out as a validation set and excluded from the training data. From these, we randomly selected an internal validation set of 30 slides. Of each selected slide, we identified areas that consisted of at least one FF. A rectangular area was created around the FF and surrounding tissue so that at least half of the area was tissue adjacent to FF. Blinded to the results analyzed by the AI model, a pathologist experienced in pulmonary pathology annotated FF in the validation areas. The pathologist's annotations that formed the ground truth were compared to the model's analysis results both statistically and visually. False positive, false negative, error, precision, sensitivity, and F1 score values were calculated for all 30 validation regions. For the values of false positive, false negative, and error, the model's analysis results per total area of all pathologist's annotations in 30 validation areas were calculated. Error was the sum of false positive and false negative. Precision was calculated as the model's analysis result area found within the pathologist's annotation area per total area of the model's analysis result area in a single validation area. Sensitivity was calculated as the pathologist's annotation area found by the model's analysis per total area of pathologist's annotation in a single validation area. F1 score represented the harmonic mean of precision and sensitivity.

Table 3 Minimum, maximum, median, and cut-point values of the areas and counts of fibroblast foci, interstitial mononuclear inflammatory cells, intra-alveolar macrophages, and air spaces and fibroblast focus/interstitial mononuclear inflammatory cell index values.

	Minimum (% or 1/mm^2)	Maximum (% or 1/mm^2)	Median (% or 1/mm^2)	Cut-point value (% or 1/mm^2)
FF area (mm^2)	0.05 (0.1 • 10^{-3})	29.5 (69.5 • 10^{-3})	3.0 (8.5 • 10^{-3})	4.3 (11.3 • 10^{-3})
FF count	733 (2)	114,110 (238)	21,355 (66)	31,240 (78)
Interstitial mononuclear inflammatory cell area (mm^2)	0.06 (0.3 • 10^{-3})	23.6 (73.9 • 10^{-3})	2.3 (7.7 • 10^{-3})	1.4 (3.2 • 10^{-3})
Interstitial mononuclear inflammatory cell count	3544 (20)	329,567 (640)	79,131 (253)	64,515 (115)
Intra-alveolar macrophage area (mm^2)	0.05 (0.1 • 10^{-3})	32.3 (95.4 • 10^{-3})	1.3 (5.0 • 10^{-3})	0.7 (2.6 • 10^{-3})
Intra-alveolar macrophage count	2736 (7)	233,392 (454)	37,498 (127)	18,385 (49)
Air space area (mm^2)	11.9 (5.5)	287.3 (58.2)	97.5 (33.0)	73.3 (23.6)
FF/interstitial mononuclear inflammatory cell index value	0.03	14.9	1.2	1.4

For areas, percentages (%) of tissue area are expressed in brackets, and for counts, densities (1/mm^2) are expressed in brackets. Cut-point values for high and low values were determined with maximally selected rank statistics. FF, fibroblast foci.

For validation areas, visual confusion matrix data are available in Appendix 2.

In previous studies, AI models have been shown to recognize inflammatory cells in cancer tissue samples [32–37]. We evaluated the performance of the AI model on inflammatory cell recognition visually.

We tested the reproducibility of the AI model's results by running the AI model three times in a subanalysis of five slides and compared the results of three separate analyses with each other.

2.4. Statistical analysis

We used IBM SPSS Statistics for Windows, version 25.0 (IBM Corp., Armonk, NY, USA). For determining whether the data were normally distributed, we used Kolmogorov-Smirnov and Shapiro-Wilk tests. For correlations, Spearman's correlations were used. For normally distributed data, a t-test was used in the comparison of two groups. For survival analysis, cut-point values for air spaces, FF, interstitial mononuclear inflammatory cells, and intra-alveolar macrophages were determined with maximally selected rank statistics [38]. For cut-point values, the R package *maxstat*, version 0.7–25 [39], in R software for Windows, version 3.5.3 (R Foundation for Statistical Computing, Vienna, Austria), was used. Survival was analyzed using the Kaplan-Meier method, and the log-rank test determined its significance. Survival was defined as the time between IPF diagnosis date and the date of an endpoint event, which was defined as death or lung transplantation. For patients alive without lung transplantation at the end of the follow-up, the follow-up time was defined as the time between IPF diagnosis date and 29 April 2019. Two-sided p-values ≤ 0.05 were considered statistically significant.

3. Results

3.1. Observations on the training of the AI model

In the training of the AI model, the correct recognition of air spaces was challenging in the samples that had large cystic cavities resembling the background of the slide. We solved the problem by increasing the field of view of the neural network. Training the model to detect inflammatory cells in both the interstitium and the air spaces was a relatively simple task, whereas FF, being more complex structures, required more training data (Table 2) due to the large variance of the HE staining intensity between slides. Although we had taught the model to separate loose fibrosis from FF, the model also recognized smaller areas of FF-like spots as FF.

3.2. Histological features and survival

For cut-point values used in the survival analysis, refer Table 3. First, we aimed to validate previous results of the number and area of FF and patient survival in our model. A high area% of FF (30/71) and a high density of FF (29/71) were associated with shortened survival ($p = 0.01$, Fig. 2A, and $p = 0.02$, respectively). Second, we analyzed the number of interstitial mononuclear inflammatory cells. A high area% of interstitial mononuclear inflammatory cells (60/71) and a high interstitial mononuclear inflammatory cell density (62/71) were associated with prolonged survival ($p = 0.01$, Fig. 2B, and $p = 0.04$, respectively). A high area% of intra-alveolar macrophages (51/71) was associated with better survival than low intra-alveolar macrophage values ($p = 0.01$, Fig. 2C). To evaluate the relationship of FF and interstitial mononuclear inflammatory cells in individual samples, the area of FF was divided

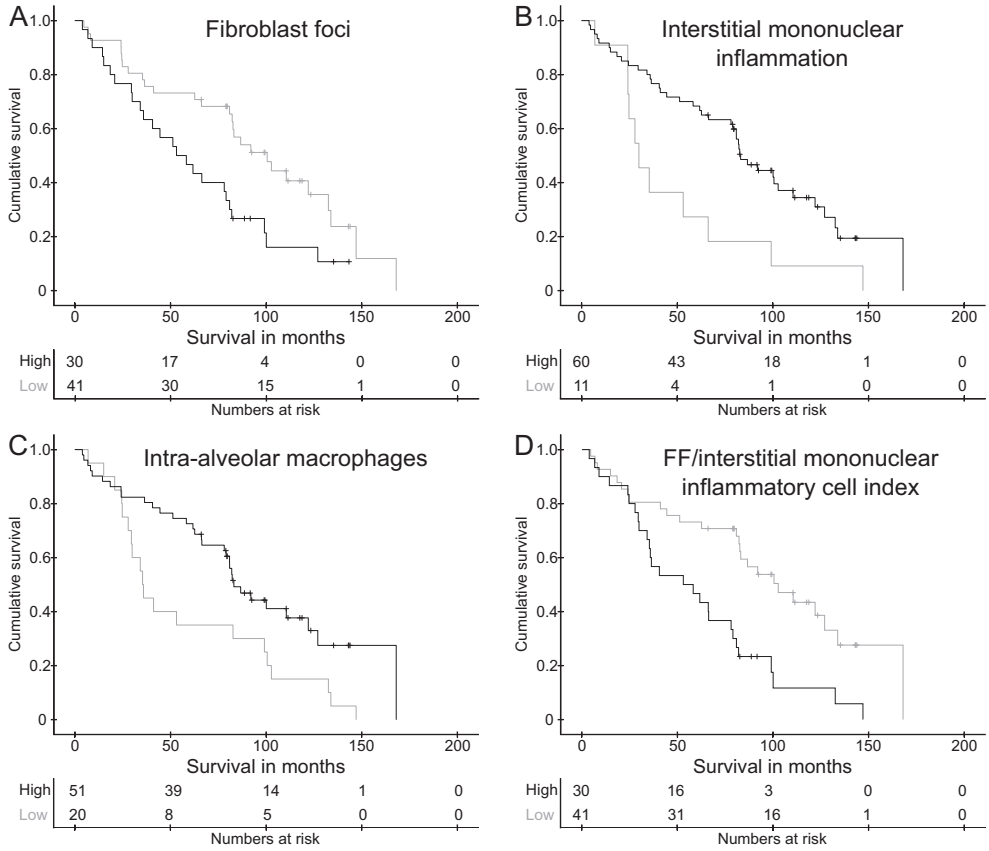

Fig. 2 Kaplan-Meier survival estimates for patients with idiopathic pulmonary fibrosis according to the areas of fibroblast foci (FF) (A), interstitial mononuclear inflammation (B), intra-alveolar macrophages (C), and FF/interstitial mononuclear inflammatory cell index values (D). The surface areas of each feature were adjusted to the whole tissue surface areas.

by the area of interstitial mononuclear inflammatory cells to create an FF/interstitial mononuclear inflammatory cell index value. A high index value (30/71) was associated with shorter survival ($p = 0.001$, Fig. 2D). The results were similar when only SLBs ($n = 62$) were taken into account.

3.3. The connection between clinical parameters and histological features

The area% of alveolar spaces was higher in SLBs than in explant and autopsy samples (62/71, median of 33.0%, range of 7.6%−58.2% vs 9/71, 24.7%, 5.5%−39.0%, Mann-Whitney U test, $p = 0.04$). Regarding the other histological parameters, no significant differences between sample types existed.

A high density of FF in the lung tissue associated with a low diffusing capacity of predicted (DLCO%) at the time of diagnosis compared with a low FF density (29/67, mean of 51.5% ± 15.8% vs 38/67, 59.9% ± 15.3%, t-test, $p = 0.03$). Patients with a high area% of intra-alveolar macrophages in lung tissue samples had a higher forced vital capacity of predicted (FVC%) at the time of diagnosis than patients with a sample with a low intra-alveolar macrophage area (46/66, mean of 78.6% ± 16.8% vs 20/66, 68.8% ± 14.7%, t-test, $p = 0.03$). The results were similar when only SLBs ($n = 62$) were taken into account.

All current smokers had a high density of intra-alveolar macrophages (11/11, 100.0%) compared with ever-smokers (24/34, 70.6%) and never-smokers (23/26, 88.5%, Fisher's exact test, $p = 0.06$). FF or interstitial mononuclear inflammatory cell amounts did not correlate with smoking status ($p > 0.05$).

3.4. Quantitated histological features

Refer Table 3 for minimum, maximum, and median values of FF, interstitial mononuclear inflammatory cells, and intra-alveolar macrophages.

3.5. Correlations between histological features

Refer Table 4 for correlations between the histological features analyzed by the AI model.

3.6. Validation of the AI model

The 30 selected validation areas, the pathologist's annotations of FF, and the results analyzed by the AI model

Table 4 Spearman's correlation coefficients between the densities of fibroblast foci, interstitial mononuclear inflammation, and intra-alveolar macrophages and alveolar space area in relation to the whole tissue area.

	FF density (1/mm^2)	Interstitial mononuclear inflammation density (1/mm^2)	Intra-alveolar macrophage density (1/mm^2)
FF density (1/mm^2)		0.460, $p < 0.001$	0.025, $p = 0.833$
Mononuclear inflammation density (1/mm^2)			0.320, $p = 0.007$
Alveolar space area (%)	−0.413, $p < 0.001$	−0.347, $p = 0.003$	0.368, $p = 0.002$

Densities were counted by the absolute count of the feature in relation to the whole tissue area. Alveolar space area was counted in relation to the whole tissue. Correlations were similar when only surgical lung biopsies were taken into account. FF, fibroblast foci.

Table 5 The results of the confusion matrix for 30 selected validation areas. The results of the artificial intelligence model were compared against the annotations of a pathologist.

	Validation annotation positive	Validation annotation negative	Total
AI model result positive	True positive 0.480 mm^2	False positive 0.516 mm^2	0.996 mm^2
AI model result negatives	False negative 0.371 mm^2	True negative 9.709 mm^2	10.08 mm^2
Total	0.851 mm^2	10.225 mm^2	11.076 mm^2

AI, artificial intelligence.

are shown in Appendix 3. The median area used in the validation was 0.03 mm^2 (range of 0.002−0.1 mm^2). The median values for false positive, false negative, error, precision, sensitivity, and F1 score were 1.4% (range of 0%−6.7%), 1.0% (range of 0.1%−5.2%), 2.9% (range of 0.6%−9.9%), 54.5% (range of 7.3%−98.2%), 65.2% (range of 7.0%−87.3%), and 55.7% (range of 7.4%−85.5%), respectively. The results of the confusion matrix for all 30 validation areas are shown in Table 5. The visual confusion matrix data of individual validation areas are shown in Appendix 2. By a visual evaluation, the AI model's performance in most validation areas seemed acceptable, although the model did not properly function in a minority of the slides (Appendix 3). Most of the small cells with round, dark nuclei and scant cytoplasm were recognized as interstitial mononuclear inflammatory cells in all of the validation areas (Appendix 3).

In a subanalysis of five slides, the values of each feature's measurements were consistent between three separate analyses of the AI model (Appendix 4).

4. Discussion

We analyzed lung tissue samples of patients with IPF using a deep CNN aiming to evaluate AI's ability to find histological features that could have a prognostic value. In a well-characterized patient population with IPF and re-evaluated diagnoses, the AI model identified potential novel lung tissue biomarkers to estimate prognosis and disease severity. Although the role of SLBs in the diagnosis of IPF has decreased due to the evolution of radiology, SLB is still recommended in unclear cases [1]. In addition to diagnostic values, more prognostic information from SLBs could be used with the use of AI.

4.1. Interstitial mononuclear inflammation and survival

We found that low amounts of interstitial mononuclear inflammation are associated with shorter survival of patients with IPF. The finding contradicts the current perception of inflammatory cells having a minimal role in the pathogenesis of IPF. Some evidence against our result exists. A high amount of T lymphocytes has been associated with poor survival [40]. In the explant samples of patients with a rapidly progressing form of IPF, all types of inflammatory cells have been seen to increase [41]. In two other larger patient cohorts, no such associations were reported between interstitial lymphocytic inflammation and survival [3,9]. Nicholson et al. [3] demonstrated an association between high interstitial mononuclear inflammation and a decline in FVC%, whereas Collard et al. [9] showed an association between high interstitial mononuclear inflammation and improvement in FVC%. Regarding the distribution of interstitial mononuclear inflammation, a high inflammatory cell density has been seen in both areas of dense fibrosis [40] and areas of loose fibrosis associated with preserved alveolar epithelium [42]. In a recent study, the genes associated with mononuclear cell migration were

upregulated in the areas of the preserved alveolar epithelium, whereas in the areas of dense fibrosis, a downregulation was noted [43]. Our novel finding indicates that the use of AI in precise quantitation of interstitial inflammation could have an impact on determining the prognosis of patients with IPF.

4.2. FF, intraluminal fibrosis, and survival

We confirmed that a high amount of FF is a marker for poor prognosis in patients with IPF [2–7]. The association of FF with shorter survival, however, has not been shown in all studies [8–11]. Our AI model also identified intraluminal fibrosis/OP as FF. We accepted this due to technical reasons, and in this study population, OP was a rare finding [14]. The model also recognized FF-like spots as FF. The field of view that was used in the AI model's FF and interstitial mononuclear inflammation layer was much smaller in comparison with the visual analysis of the human eye. This partly explains the counting of FF-like spots toward the amount of FF and reflects the difference in the visual analysis between the human eye and AI. Methodological differences in the counting of the OP-like pattern and other FF-like areas toward the FF score could partly explain the controversial results among studies. Similarly to our method, King et al. [2] did not differentiate intraluminal fibrosis from FF, whereas some studies without an association between FF and survival deliberately excluded OP from the FF score [9,10]. Although OP is not considered a typical finding for UIP [1], it has been reported to exist in IPF samples [9,14,44]. The acute exacerbation of IPF can manifest histologically as OP or extensive FF superimposed on the UIP pattern [45]. Compared with the samples from patients with cryptogenic OP, which is characterized by reversible disease course, decreased vascularity and apoptosis of intraluminal fibrosis have been reported in the samples from patients with IPF [46,47]. Both OP and FF have been associated with a decline in FVC% at six-month follow-up [9]. In IPF samples, the use of pirfenidone and nintedanib has been reported to decrease OP compared with untreated patients [48]. In SLBs of stable patients with IPF, minute lesions of alveolar damage have been associated with later mortality and acute exacerbations [49]. Nodular granulation tissue and large FF coexisted with these lesions [49]. As our AI model also detected intraluminal fibrosis and FF-like spots besides FF, our result could support the theory of all fibrosing processes worsening the prognosis of patients with IPF.

4.3. Intra-alveolar macrophages and survival

Intra-alveolar macrophages are a common finding in lung tissue samples, and they are known to have both profibrotic and antifibrotic properties in fibrotic lungs [13]. In UIP samples, proliferative activity of intra-alveolar macrophages has been higher relative to controls [50]. We observed that high amounts of intra-alveolar macrophages were associated with prolonged survival and higher FVC% predicted at the time of diagnosis. The prognostic effect of intra-alveolar macrophages has not been previously shown [3,9]. Recently, monocytes in the peripheral blood, which are progenitor cells for intra-alveolar macrophages, have been associated with poor outcomes for patients with IPF [51]. One reason for sparse histological studies on intra-alveolar macrophages in the lungs of patients with IPF could be the difficulty of differentiating intra-alveolar macrophages from interstitial ones by immunohistochemistry-based methods. The ability of our model to separate alveolar spaces from the interstitium could be especially useful in the evaluation of alveolar pathology in IPF.

4.4. Strengths and limitations

The most significant benefit of our AI model was the capability to accurately count the inflammatory cells in both intra-alveolar spaces and the interstitium, which has been practically impossible in manual methods. In the subanalysis of five slides, the measurements made by the AI model were consistent, which is a benefit in comparison with manual methods that are often prone to intraobserver variation. Technically, inflammatory cells were the easiest feature to train for the AI model. In learning some features, the model had occasional difficulties: especially in recognition of air spaces, which is a simple task for the human eye, and in recognition of FF. One explanation is that the size and conformation of air spaces and FF are more variable than those of inflammatory cells, creating a challenge in the adjustment of the field size. Besides, the morphology of FF is more variable and complex than that of inflammatory cells; FF is a structural element of lung tissue composing of myofibroblasts, endothelial, epithelial, and inflammatory cells, as well as abundant extracellular matrix. Visually, the AI model functioned the best when analyzing slides having a similar intensity of HE staining than the training data. Before implementing AI across different laboratories, taking into account all variations in intensities of HE staining and artefacts of real-life slides is also mandatory. Owing to the preliminary nature of the study, we selected one representative slide for each patient for digitization, which may cause a selection bias. The small training set used in this study can lead to data overfitting, that is, the model functions well for training data but underperforms with the data not included in the training. Data overfitting is a common problem in the CNN and can partly explain why the model did not recognize a minority of FF. Data overfitting can often be managed with a bigger training set. In this study, we had a limited number of samples, and we chose to preserve an internal validation set (51 samples out of 71) as large as possible. In addition,

more extensive use of digital augmentation could help with data overfitting. Our study also lacks an external validation set, which is a significant limitation. Furthermore, a truly generalizable AI model for diagnostic purposes would require substantially larger data sets both for AI model training and validation and several pathologists as validators. Creating an AI model suitable for clinical use, however, was beyond the scope of our pilot study. Nevertheless, we believe that AI is an advantage in the analysis of inflammation in the lung tissue and that the results of our study gave a reason to assume that the amount of inflammatory cells has an impact on the prognosis of patients with IPF. The routine use of AI in the histopathological analysis of IPF samples, however, will require additional studies.

5. Conclusions

Even an AI model developed with a small sample size can detect specific histological features in the IPF lung tissue samples that it has not previously encountered. A low amount of interstitial mononuclear inflammation and intra-alveolar macrophages was associated with poor prognosis in a study population with confirmed IPF diagnoses, a finding supporting the theory that inflammation has a significant role in IPF pathogenesis. The AI model confirmed the connection between high FF amounts and poor prognosis for patients with IPF. Automated image analysis could provide new possibilities for investigating the relationships of IPF histology and clinical parameters. In the future, AI could be a novel tool for the pathologists in the histological diagnosis of IPF and other interstitial lung disorders.

Acknowledgments

The authors thank all patients who participated in this study. The authors are also grateful to the participants of the FinnishIPF consortium. The authors are grateful to the pulmonary physicians who contributed to the study by recruiting patients and obtaining their informed consent. The authors thank the staff of Aiforia Technologies for technical support. For language revision, the authors thank Carol Ann Pelli.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humpath.2020.10.008.

References

[1] Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Med Crit Care 2018;198:e44−68.

[2] King T, Tooze J, Schwarz M, Brown K, Cherniack R. Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. Am J Respir Crit Care Med 2001;164:1171−81.

[3] Nicholson A, Fulford L, Colby T, du Bois R, Hansell D, Wells A. The relationship between individual histologic features and disease progression in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2002;166:173−7.

[4] Enomoto N, Suda T, Kato M, et al. Quantitative analysis of fibroblastic foci in usual interstitial pneumonia. Chest 2006;130:22−9.

[5] Tiitto L, Bloigu R, Heiskanen U, Pääkkö P, Kinnula V, Kaarteenaho-Wiik R. Relationship between histopathological features and the course of idiopathic pulmonary fibrosis/usual interstitial pneumonia. Thorax 2006;61:1091−5.

[6] Lee SH, Shim HS, Cho SH, et al. Prognostic factors for idiopathic pulmonary fibrosis: clinical, physiologic, pathologic, and molecular aspects. Sarcoidosis Vasc Diffus Lung Dis 2011;28:102−12.

[7] Harada T, Watanabe K, Nabeshima K, Hamasaki M, Iwasaki H. Prognostic significance of fibroblastic foci in usual interstitial pneumonia and non-specific interstitial pneumonia. Respirology 2013;18:278−83.

[8] Flaherty K, Colby T, Travis W, et al. Fibroblastic foci in usual interstitial pneumonia: idiopathic versus collagen vascular disease. Am J Respir Crit Care Med 2003;167:1410−5.

[9] Collard H, Cool C, Leslie K, Curran-Everett D, Groshong S, Brown K. Organizing pneumonia and lymphoplasmacytic inflammation predict treatment response in idiopathic pulmonary fibrosis. Histopathology 2007;50:258−65.

[10] Hanak V, Ryu J, de Carvalho E, et al. Profusion of fibroblast foci in patients with idiopathic pulmonary fibrosis does not predict outcome. Respir Med 2008;102:852−6.

[11] Triantafillidou C, Manali E, Magkou C, et al. Medical Research Council dyspnea scale does not relate to fibroblast foci profusion in IPF. Diagn Pathol 2011;6:28.

[12] Cherniack R, Crystal R, Kalica A. NHLBI Workshop summary. Current concepts in idiopathic pulmonary fibrosis: a road map for the future. Am Rev Respir Dis 1991;143:680−3.

[13] Heukels P, Moor C, von der Thüsen J, Wijsenbeek M, Kool M. Inflammation and immunity in IPF pathogenesis and treatment. Respir Med 2019;147:79−91.

[14] Mäkelä K, Hodgson U, Piilonen A, et al. Analysis of the histologic features associated with interobserver variation in idiopathic pulmonary fibrosis. Am J Surg Pathol 2018;42:672−8.

[15] Yagihashi K, Huckleberry J, Colby T, et al. Radiologic-pathologic discordance in biopsy-proven usual interstitial pneumonia. Eur Respir J 2016;47:1189−97.

[16] Rabeyrin M, Thivolet F, Ferretti G, et al. Usual interstitial pneumonia end-stage features from explants with radiologic and pathological correlations. Ann Diagn Pathol 2015;19:269−76.

[17] Kim SY, Diggans J, Pankratz D, et al. Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. Lancet Respir Med 2015;3:473−81.

[18] Pankratz D, Choi Y, Imtiaz U, et al. Usual interstitial pneumonia can be detected in transbronchial biopsies using machine learning. Ann Am Thorac Soc 2017;14:1646−54.

[19] Raghu G, Flaherty K, Lederer D, et al. Use of a molecular classifier to identify usual interstitial pneumonia in conventional transbronchial lung biopsy samples: a prospective validation study. Lancet Respir Med 2019;7:487−96.

[20] Maldonado F, Moua T, Rajagopalan S, et al. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. Eur Respir J 2014;43:204−12.

[21] Jacob J, Bartholmai B, Rajagopalan S, et al. Predicting outcomes in idiopathic pulmonary fibrosis using automated computed tomographic analysis. Am J Respir Crit Care Med 2018;198:767−76.

[22] Robbie H, Wells A, Jacob J, et al. Visual and automated CT measurements of lung volume loss in idiopathic pulmonary fibrosis. AJR Am J Roentgenol 2019;213:318−24.

[23] Walsh S, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. Lancet Respir Med 2018;6:837−45.
[24] Gilhodes J, Julé Y, Kreuz S, Stierstorfer B, Stiller D, Wollin L. Quantification of pulmonary fibrosis in a bleomycin mouse model using automated histological image analysis. PloS One 2017;12:e0170561.
[25] Heinemann F, Birk G, Schoenberger T, Stierstorfer B. Deep neural network based histological scoring of lung fibrosis and inflammation in the mouse model system. PloS One 2018;13:e0202708.
[26] Seger S, Stritt M, Vezzali E, et al. A fully automated image analysis method to quantify lung fibrosis in the bleomycininduced rat model. PloS One 2018;13:e0193057.
[27] Kaunisto J, Salomaa E-R, Hodgson U, et al. Demographics and survival of patients with idiopathic pulmonary fibrosis in the FinnishIPF registry. ERJ Open Res 2019;5. 00170-2018.
[28] Raghu G, Collard H, Egan J, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 2011;183:788−824.
[29] Mäkelä K, Ollila H, Sutinen E, et al. Inorganic particulate matter in the lung tissue of idiopathic pulmonary fibrosis patients reflects population density and fine particle levels. Ann Diagn Pathol 2019;40:136−42.
[30] Penttinen A-M, Parkkinen I, Blom S, et al. Implementation of deep neural networks to count dopamine neurons in substantia nigra. Eur J Neurosci 2018;48:2354−61.
[31] Sveen L, Timmerhaus G, Johansen L-H, Ytteborg E. Deep neural network analysis - a paradigm shift for histological examination of health and welfare of farmed fish. Aquaculture 2021;532:736024.
[32] Yuan Y, Failmezger H, Rueda OM, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. Sci Transl Med 2012;4:157ra143.
[33] Turkki R, Linder N, Kovanen PE, Pellinen T, Lundin J. Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. J Pathol Inform 2016;7:38.
[34] Saltz J, Gupta R, Hou L, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. Cell Rep 2018;23:181−93.
[35] Aprupe L, Litjens G, Brinker T, van der Laak J, Grabe N. Robust and accurate quantification of biomarkers of immune cells in lung cancer micro-environment using deep convolutional neural networks. PeerJ 2019;7:e6335.
[36] Linder N, Taylor JC, Colling R. Deep learning for detecting tumour-infiltrating lymphocytes in testicular germ cell tumours. J Clin Pathol 2019;72:157−64.
[37] Wang S, Wang T, Yang L, et al. ConvPath: a software tool for lung adenocarcinoma digital pathological image analysis aided by a convolutional neural network. EBio Med 2019;50:103−10.
[38] Lausen B, Schumacher M. Maximally selected rank statistics. Biometrics 1992;48:73−85.
[39] Hothorn T. maxstat: maximally selected rank statistics. 2017. https://CRAN.R-project.org/package=maxstat. [Accessed 5 May 2019].
[40] Parra E, Kairalla R, Ribeiro De Carvalho C, Eher E, Capelozzi V. Inflammatory cell phenotyping of the pulmonary interstitium in idiopathic interstitial pneumonia. Respiration 2007;74:159−69.
[41] Balestro E, Calabrese F, Turato G, et al. Immune inflammation and disease progression in idiopathic pulmonary fibrosis. PloS One 2016;11:e0154516.
[42] Nuovo G, Hagood J, Magro C, et al. The distribution of immunomodulatory cells in the lungs of patients with idiopathic pulmonary fibrosis. Mod Pathol 2012;25:416−33.
[43] Jonigk D, Stark H, Braubach P, et al. Morphological and molecular motifs of fibrosing pulmonary injury patterns. J Pathol Clin Res 2019;5:256−71.
[44] Takemura T, Akashi T, Kamiya H, et al. Pathological differentiation of chronic hypersensitivity pneumonitis from idiopathic pulmonary fibrosis/usual interstitial pneumonia. Histopathology 2012;61:1026−35.
[45] Churg A, Müller N, Silva C, Wright J. Acute exacerbation (acute lung injury of unknown cause) in UIP and other forms of fibrotic interstitial pneumonias. Am J Surg Pathol 2007;31:277−84.
[46] Lappi-Blanco E, Kaarteenaho-Wiik R, Soini Y, Risteli J, Pääkkö P. Intraluminal fibromyxoid lesions in bronchiolitis obliterans organizing pneumonia are highly capillarized. Hum Pathol 1999;30:1192−6.
[47] Lappi-Blanco E, Soini Y, Pääkkö P. Apoptotic activity is increased in the newly formed fibromyxoid connective tissue in bronchiolitis obliterans organizing pneumonia. Lung 1999;177:367−76.
[48] Zhang Y, Jones K, Achtar-Zadeh N, et al. Histopathological and molecular analysis of idiopathic pulmonary fibrosis lungs from patients treated with pirfenidone or nintedanib. Histopathology 2019;74:341−9.
[49] Emura I, Usuda H, Togashi K, Satou K. Minute lesions of alveolar damage in lungs of patients with stable idiopathic pulmonary fibrosis. Histopathology 2015;67:90−5.
[50] El-Zammar O, Rosenbaum P, Katzenstein A-L. Proliferative activity in fibrosing lung diseases: a comparative study of Ki-67 immunoreactivity in diffuse alveolar damage, bronchiolitis obliterans-organizing pneumonia, and usual interstitial pneumonia. Hum Pathol 2009;40:1182−8.
[51] Scott M, Quinn K, Li Q, et al. Increased monocyte count as a cellular biomarker for poor outcomes in fibrotic diseases: a retrospective, multicentre cohort study. Lancet Respir Med 2019;7:497−508.