Internal and Emergency Medicine (2021) 16:1457–1465
https://doi.org/10.1007/s11739-020-02583-x

IM - ORIGINAL

A convolutional neural network with transfer learning for automatic discrimination between low and high-grade synovitis: a pilot study

Vincenzo Venerito[1] · Orazio Angelini[2,3] · Gerardo Cazzato[4] · Giuseppe Lopalco[1] · Eugenio Maiorano[4] · Antonietta Cimmino[4] · Florenzo Iannone[1]

Received: 25 September 2020 / Accepted: 20 November 2020 / Published online: 2 January 2021

Abstract

Ultrasound-guided synovial tissue biopsy (USSB) may allow personalizing the treatment for patients with inflammatory arthritis. To this end, the quantification of tissue inflammation in synovial specimens can be crucial to adopt proper therapeutic strategies. This study aimed at investigating whether computer vision may be of aid in discriminating the grade of synovitis in patients undergoing USSB. We used a database of 150 photomicrographs of synovium from patients who underwent USSB. For each hematoxylin and eosin (H&E)-stained slide, Krenn's score was calculated. After proper data pre-processing and fine-tuning, transfer learning on a ResNet34 convolutional neural network (CNN) was employed to discriminate between low and high-grade synovitis (Krenn's score <5 or ≥ 5). We computed test phase metrics, accuracy, precision (true positive/actual results), and recall (true positive/predicted results). The Grad-Cam algorithm was used to highlight the regions in the image used by the model for prediction. We analyzed photomicrographs of specimens from 12 patients with arthritis. The training dataset included n.90 images (n.42 with high-grade synovitis). Validation and test datasets included n.30 (n.14 high-grade synovitis) and n.30 items (n.16 with high-grade synovitis). An accuracy of 100% (precision = 1, recall = 1) was scored in the test phase. Cellularity in the synovial lining and sublining layers was the salient determinant of CNN prediction. This study provides a proof of concept that computer vision with transfer learning is suitable for scoring synovitis. Integrating CNN-based approach into real-life patient management may improve the workflow between rheumatologists and pathologists.

Keyword Synovitis · Machine learning · Convolutional neural network · Ultrasound-guided synovial biopsy

Abbreviations

AI	Artificial Intelligence
bDMARDs	Biological disease-modifying anti-rheumatic drugs
CNN	Convolutional Neural Network
csDMARDs	Conventional synthetic disease-modifying anti-rheumatic drugs
DAPSA	Disease Activity in Psoriatic Arthritis
DAS28-ESR	Disease Activity Score on 28 joints with Erythrocyte Sedimentation Rate
MLP	Multi-layer Perceptron
MTX	Methotrexate
PDN	Prednisone
TNFi	Tumour Necrosis Factor inhibitors
SSZ	Sulfasalazine
USSB	Ultrasound-guided synovial tissue biopsy

Vincenzo Venerito and Orazio Angelini joint first authors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11739-020-02583-x.

✉ Giuseppe Lopalco
glopalco@hotmail.it

Vincenzo Venerito
vincenzo.venerito@gmail.com

Eugenio Maiorano
eugenio.maiorano@uniba.it

Florenzo Iannone
florenzo.iannone@uniba.it

[1] Department of Emergency and Organ Transplantations-Rheumatology Unit, University of Bari "Aldo Moro", Bari, Italy

[2] King's College London, London, UK

[3] Amazon Research Cambridge, Cambridge, UK

[4] Department of Emergency and Organ Transplantations-Pathology Unit, University of Bari "Aldo Moro", Bari, Italy

Introduction

Ultrasound-guided synovial tissue biopsy (USSB) provides a better understanding of the pathophysiology of inflammatory arthritis, facilitating the discovery of new biomarkers for diagnostic and/or prognostic purposes and the identification of new therapeutic targets [1, 2]. As recently shown, USSB may allow treatment personalization in the very next future [3] and perfectly suits everyday clinical practice as it is a reliable, safe, and relatively fast procedure, also in outpatient settings. A critical issue in the diagnostic work-up is the standardization of biopsy reading, therefore validated technological support may help pathologists in providing more informative and faster biopsy reports. In an increasing number of social and clinical scenarios, artificial intelligence (AI) is proving to be a valuable tool for the generation and implementation of complex multi-parametric decision algorithms. In particular, there is a rising interest in the application of image analysis and machine learning techniques to histopathology. Whatever the image acquisition method is, either traditional small fields capture on microscope or modern whole slide image digitalization, computer vision may yield accurate diagnostic interpretations [4]. In 2012, convolutional neural networks (CNNs) outperformed previous machine learning approaches by classifying 1.2 million high-resolution images in the ImageNet LSVRC2010 contest into 1000 different classes. At the same time, CNNs were found to be superior to other methods in segmenting nerves in electron microscopy images and detecting mitotic cells in histopathology images. Since then, methods based on CNNs have consistently outperformed other handcrafted methods in a variety of classification tasks in digital pathology [5]. The ability of CNNs to learn features directly from the raw data without the need for inputs from pathologists and the availability of marked histopathology datasets has also fuelled the explosion of interest in deep learning applied to histopathology [5]. Unlike previous deep learning architectures such as the *multilayer perceptron*, where each neuron in one layer connects to every neuron in the following layers, CNNs are supervised multidimensional algorithms consisting of a neural network with several hidden neuron layers, where connection and activation only take place between spatially close neurons, echoing the organization of the animal visual cortex in which neurons selectively respond to different stimuli. When presented with sufficient annotated training image data, CNNs can learn complex histological patterns from images through deconvolution of the image content into thousands of salient features, followed by selection and aggregation of the most meaningful ones [5]. These patterns may be then identified in yet unseen images [6].

In rheumatology, AI may enable a further step towards precision medicine, leading to the improvement of patient profiling and treatment personalization. AI has proven to be effective in predicting treatment responses to TNF inhibitors by driving treatments based on the clinical and genetic features of analysed patients [7]. Computerized digital analysis based on RGB video signal acquisition through a microscope had been used in the last two decades to quantify cells infiltrating the synovium, although this procedure has never been adopted in real-life patient management as it was time-consuming and led to a non-negligible level of disagreement between centres [8, 9] The potential benefit of using computer vision for histopathologic analyses of synovitis is largely unexplored so far. The aim of our study was to investigate whether CNNs have the potential to accurately identify synovitis' grade according to Krenn's synovitis score [10].

Materials and methods

Data acquisition

For training, validation and testing, we used a dataset of 150 photomicrographs of different synovitis slides, originally taken at 1280 × 1080 pixels (SONY® Sensor IMX185) at diverse magnifications (4 × to 20x), obtained from rheumatic patients with arthritis of knees who underwent USSB for routine clinical practice in a single tertiary centre, from January 2019 to January 2020. This was necessary because of the heterogeneity of inflammatory changes in the synovial membrane requires sample analyses at different magnifications, as originally proposed by Krenn et al. [10].

For each patient, demographic and clinical characteristics, including ultrasound features of synovitis were recorded for routine clinical practice, moreover informed consent was obtained from all patients and the local ethical committee approved this study as part of the Biopure registry (IRB approval n 5277/2017). According to OMERACT standardization of synovial tissue biopsy procedure, for each patient, at least 6 specimens were obtained from an involved knee joint using 16G Tru-Cut needles and then embedded in paraffin [11]. In total, 78 specimens (mean length ± SD 0.83 ± 0.08 cm) were processed. For all slides, biopsy surface area was greater than 2.5 mm^2. Haematoxylin and eosin (H&E) staining was used for histopathological analysis and, for each of them, Krenn's score was calculated by a pathologist with 20 year experience in synovial histopathology (AC). The latter score is obtained by semiquantitatively evaluating three features of chronic synovitis (enlargement of the lining cell layer, the cellular density of synovial stroma, leukocytic infiltrate) (from 0, absent to 3, strong), allowing discrimination between low grade (i.e. Krenn's Score < 5)

and high-grade synovitis (i.e. Krenn's Score ≥ 5) [10]. Lining layer was evaluable in all 150 images.

The whole dataset was split and photomicrographs were randomly allocated to either the training, validation and test datasets according to a 3:1:1 ratio [12].

Theory

CNNs require large labelled image datasets to attain a high level of classification accuracy [13]. In several fields though, the acquisition of such an image dataset is challenging and their annotation is expensive.

Transfer learning (TL) has emerged as a powerful tool to mitigate these issues; TL consists of a process where a model trained on one problem is exploited to predict labels related to a second, similar, problem [14]. The first few layers in CNNs extract very general information such as colors, dots, lines and edge information, while subsequent layers aggregate these features into complex patterns. The initial layers of a CNN trained on a large and varied dataset can be hence treated as general image feature extractors [13]. Because of this, it is possible to "freeze" the weights of a CNN's initial layers and fine-tune or re-train the last layers only, which generally encode higher-level features. This allows the model to recognize features specific, in our case, to histopathologic images. On one hand pre-trained models substantially ease the development of new models, also leading to lower generalization error. On the other hand, there is robust evidence that, with suitable fine-tuning, pre-trained CNNs may outperform a CNN trained from scratch for biomedical applications [14–16].

We fine-tuned a specific CNN architecture called ResNet34 [17] (He et al. 2015). For reproducibility, we used a Python 3.6 environment with PyTorch 1.4.0 and fast.ai [18] 1.06 libraries. ResNet34 is a particular architecture which was pre-trained with ImageNet database [19] containing approximately 1.2 million of images of about 22.000 categories. ResNet34 consists of 5 convolutional layer groups ending with a pooling layer group for prediction (for details see Supplementary Materials and Fig. 1 [17, 18]. Briefly, as the input flows through the ResNet34, less complex feature maps in the former layers apply filters for the above mentioned basic visual elements, whereas the latter layers provide more complex features. For this reason, the initial layers of ResNet34 can be considered as general image feature extractors. The architecture of ResNet34 has been briefly represented in Fig. 1, where we annotated the layers that we fine-tuned during transfer learning (see Supplementary Materials). To improve the robustness and ability of our model to generalize, and to further decrease the risk of overfitting [6], image augmentation (i.e. up to n.1680 different versions of the same item) was performed following the default augmentation protocol in fast.ai 1.06 [18].

Parameters and metrics

In the training and validation phases, we relied on the train and validation losses to investigate the model's goodness-of-fit.

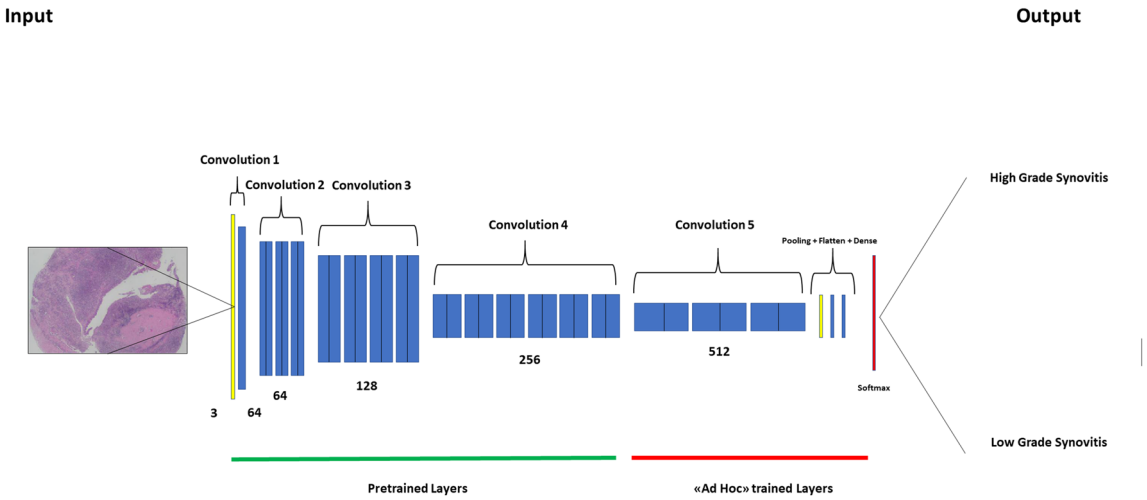

Fig. 1 We depict an outline of ResNet34's architecture, and show what parts were fine-tuned in this work. The model consists on one convolution and pooling step (in yellow) followed by n.4 convolution groups of similar structure. Its last layers are a global average pooling layer and a 1000-way fully-connected layer with softmax for prediction. Layers in distinct convolution groups follow the same pattern performing 3 × 3 convolution with a fixed feature map of ascending dimension (64, 128, 256, 512), with Rectified Linear Unit (ReLu) activation, bypassing the input every 2 layers. The initial layers of a CNN, trained on a large and varied dataset, can be treated as general image feature extractors. Because of this, it is possible to "freeze" the weights of a CNN's initial layers and fine-tune or re-train the last layers only, to enable the model to recognize features specific to histopathologic images

The loss represents a quantitative measure of how much the model's predictions differ from ground truth (i.e. the pathologist's Krenn's score). In general, losses are defined to be inversely proportional to the number of correct predictions of the model, so that the training procedure can be defined as a loss minimization problem. In other words, loss can be defined as an average of the errors made by the models on the images contained in a subset of the data. The loss is calculated on the training and validation sets, therefore it can be interpreted as a number describing how accurately the model is predicting these two sets.

To rule out overfitting and underfitting during training, both losses should typically show a trend towards decrease, with the training loss normally being the smallest.

An "epoch" indicates one forward pass and one backward pass of all the training images. The number of epochs after which to stop training is a metaparameter usually chosen to minimize the loss while avoiding overfit [14].

The test phase performance of our CNN was assessed with the following metrics:

$$\text{Accuracy} = \frac{\text{truepositives} + \text{truenegatives}}{\text{truepositives} + \text{truenegatives} + \text{falsepositives} + \text{falsenegatives}}$$

$$\text{Recall} = \frac{\text{truepositives}}{\text{truepositives} + \text{falsenegatives}}$$

$$\text{Precision} = \frac{\text{truepositives}}{\text{truepositives} + \text{falsepositives}}$$

Calculation

Images were scaled at 500 × 281 pixels and underwent pixel z-score normalization. Our training dataset included n.90 images (n.42/90 high-grade synovitis); the classification capability of CNNs is highly reliant on the size of the data used for the training; if the dataset is small, the CNN model starts overfitting already after a few epochs [20]. Given that, ResNet34 was trained for 4 epochs using fast.ai 1cycle policy on the aforementioned dataset, opting for a fine-tuning involving only the last convolutional layer, an approach that has been discussed and found to work well in cases similar to ours [21, 22] (see Supplementary Materials for details). Figure 1 graphically shows what parts of the network were fine-tuned.

Validation and test datasets included n.30 (n.14/30 high-grade synovitis) and n.30 items (n.16/30 high-grade synovitis), respectively.Following the training phase, validation and test were carried out on the remaining datasets (See Supplementary Materials for details).

Grad-CAM algorithm

Similarly to other deep learning models, CNNs are considered "black box" methods, for which researchers cannot precisely explain what parts of the input image the network is "attending" to, or how the model arrived at its final output [23]. It is crucial to resolve these issues, with particular regards to biomedical contexts. To provide explainability we employed the Grad-CAM algorithm, which has been discussed and applied in recent literature for visually debugging CNNs and properly understanding which features or parts of the image are the most important for classification purposes [14, 23].

In brief, Grad-CAM uses the loss functions with respect to one specific test image to produce a heatmap highlighting the regions that are more relevant to the model for predicting the given label.

With Grad-CAM we checked where in each test image the CNN was looking when a histopathological slice is evaluated. This allowed us to further validate that the model works correctly, by verifying that it is indeed "attending" intuitively correct patterns in the image and activating around those patterns. Examples are discussed in the Results section.

Ablation study

In deep learning research, an ablation study typically refers to removing some features of the model or algorithm and seeing how that affects performance for the sake of explainability.

To properly observe the effect of the fine-tuning on the CNN, we also compared the performance of the model with the fine-tuning and the performance of it without the fine-tuning.

Inter-rater reliability study

Krenn's synovitis score was independently assessed on the aforementioned test dataset by a second pathologist (GC), who was not aware of its colleague's classification. Using Cohen's K method we measured inter-rater reliability, comparing ResNet34 outcomes with the latter pathologist's report.

Results

Twelve patients (6/12 female, 50%) with a mean (± SD) age 48.7 ± 12 years underwent USSB of knee synovium during routine clinical practice in the time frame of the study. In particular 6/12 patients (50%) had Psoriatic Arthritis, 5/12

(41.7%) had Rheumatoid Arthritis, 1/12 (8.3%) had Peripheral Spondyloarthritis. All patients met the current classification criteria for each disease [24–26]. Detailed patients' characteristics are illustrated in Table 1.

Validation phase

The learning curves in the validation phase showed that the CNN learned steadily with a rapid decrease in train loss and the concomitant increase of accuracy (Fig. 2). Fine-tuned (ft-) ResNet34 showed a good fit: train loss and validation loss, both plotted over epochs, displayed a trend towards a decrease with the former being the smallest (Validation accuracy 96.67%). That was not the case of the plain ResNet34, trained without fine-tuning for ablation purposes that indicated underfitting, with the training loss continuing to decrease until the end of the training, displaying higher values than validation loss throughout the whole process (Validation accuracy 86.67%, Fig. 2).

Test phase

Conversely deploying the ft-ResNet34 on the test dataset yielded an accuracy of 100% was shown (precision = 1, recall = 1, Fig. 3). As expected plain ResNet34 showed worse performance, scoring accuracy = 90%, precision = 0.90 and recall = 0.90 (Fig. 3).

Grad-CAM analysis

Grad-Cam (Fig. 4) shows that the activation map of our ft-CNN focused on the cellularity in the synovial lining and the sublining layers—two of Krenn's score fundamental items [8]. This is further confirmation that the model is working correctly. We conclude that it consistently focuses its activation map on areas of the image that are considered very informative for synovitis grading by human pathologists as well.

Reliability

Sixteen out of 30 high grade synovitis were identified by the second pathologist in the test dataset (ground truth n.16/30). Cohen's K for inter-rater reliability taking into account ft-CNN output and the latter pathologist's diagnosis was 1, indicating very good agreement.

Table 1 Clinical demographic and histopathological characteristics of our cohort

	Av. Obs	Rheumatoid arthritis (n.5 patients)	Psoriatic arthritis (n.6 patients)	Peripheral spondyloarthritis (n.1 patient)
Female, *n* (%)	12	4 (30)	2 (16.7)	0
Age at biopsy, years, mean (SD)	12	42.6 (13.3)	53 (11.7)	51
Disease duration at biopsy, months, mean (SD)	12	155 (126)	114 (98.9)	3
csDMARDs at biopsy, *n* (%)	12	4 (30)	1(8.3)	0
MTX		3 (25)	0	0
SSZ		1(8.3)	0	0
bDMARDs at biopsy, *n* (%)	12	2 (16.7)	3 (25)	0
TNFi		2 (16.7)	2 (16.7)	0
Secukinumab			1(8.3)	0
tsDMARDs at biopsy, *n* (%)	12	1 (8.3)	0	0
Baricitinib		1 (8.3)	0	0
Steroids at biopsy *n* (%)	12	2 (16.7)	1(8.3)	0
≤7.5 mg PDN		2 (16.7)	1(8.3)	0
>7.5 mg PDN		0	0	0
DAS28-ESR, mean (SD)	12	3.76 (1.2)		
DAPSA, mean (SD)	12		13.7 (2.1)	18.2
Power Doppler grade, median (IQR)	12	1 (1)	2 (0)	2
Krenn's score, median (IQR)	12	4 (1)	6 (3)	6

bDMARDs biological disease-modifying anti-rheumatic drugs, *csDMARDs* conventional synthetic disease-modifying anti-rheumatic drugs, *DAPSA* Disease Activity in Psoriatic Arthritis, *DAS28-ESR* Disease Activity Score on 28 joints with Erythrocyte Sedimentation Rate, *MTX* Methotrexate, *PDN* Prednisone, *TNFi* Tumour Necrosis Factor inhibitors; *SSZ* sulfasalazine

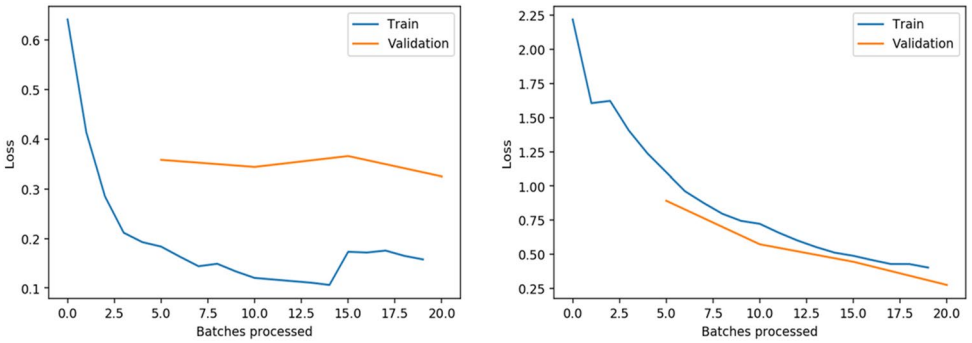

Fig. 2 Training and validation metrics. Train loss and validation loss have been plotted. In the fine-tuned ResNet34 model (left panel) a good fit was observed: both train loss and validation loss showed a trend towards a decrease with the former being the smallest. The plain ResNet34 model (right panel) indicated underfitting, with the training loss continuing to decrease until the end of training, displaying higher values than validation loss throughout the whole process

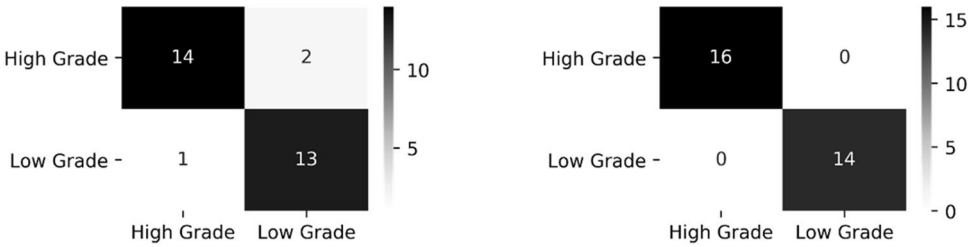

Fig. 3 Confusion matrix for test phase metrics (high VS low grade synovitis) for fine-tuned (right panel) and plain (left panel) ResNet34 models. Upper-left square shows true positive predictions (high grade synovitis), lower-right square shows true negative ones (low grade synovitis). Right and Lower-left squares and Upper-right squares show false positive and false negative predictions, respectively. For fine-tuned model (left panel): accuracy = 100%, precision (true positive/actual results) = 1, recall (true positive/predicted results) = 1. For plain model (right panel): accuracy = 90%, precision (true positive/actual results) = 0.90, recall (true positive/predicted results) = 0.90

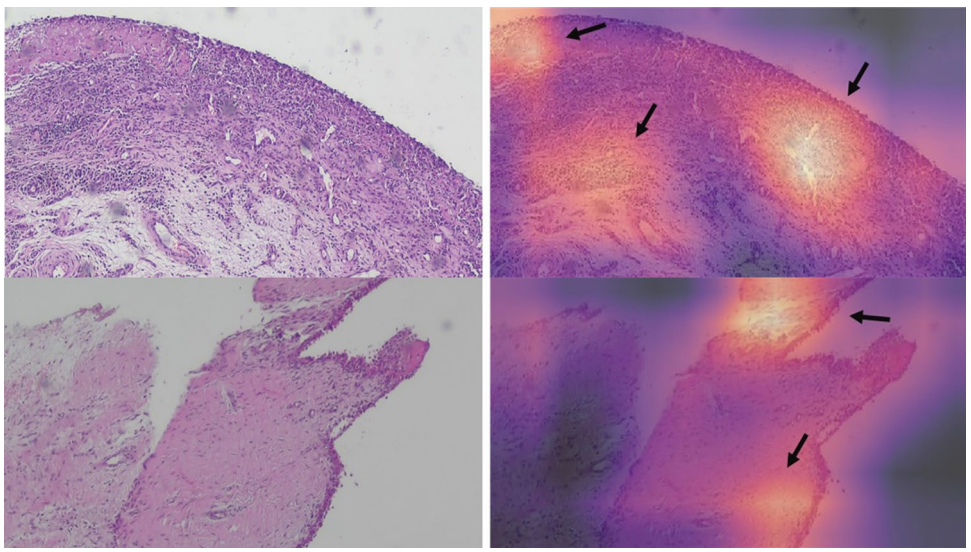

Fig. 4 Two examples of Grad-CAM algorithm output on our fine-tuned model. The algorithm provides a heatmap highlighting the regions that are more relevant to the model for predicting the given label. Here we show each test example on the left, and the same image with the heatmap superimposed on the right. Red-yellow zones (highlighted by black arrows) are the most informative areas for the model's classification of both high grade synovitis (upper panels, 20× magnification) and low grade synovitis (lower panels, 20× magnification). It can be noticed that cellularity in lining and sublining layers is a salient image characteristic. Areas that are less informative for the model are darkened by the heatmap

Discussion

To the best of our knowledge, this is the first report showing that a CNN trained with a TL approach can accurately discriminate between high and low-grade synovitis in USSB specimens. H&E stained slides, the basic tool of precision-based diagnostics, also represent the basis of personalized care in rheumatology. USSB is now poised to revolutionize management of patients with rheumatic diseases, by helping rheumatologists to extract large amounts of objective and multiparametric information about disease pathogenesis, prognosis and treatments outcome [3, 27]-With the clear perspective of USSB-driven therapy, pathologists will play a key role in such revolution, being progressively involved in rheumatologists' everyday clinical practice.

In the past years, interactive measurement of synovial layers and computer-assisted image analysis based on RGB signal nuances recognition had been used to count cells in lining and synovial stroma [8]. Nevertheless, this process was time-consuming as it required pathologists to mark the margins of synovial lining, whereas the final score showed only an adequate correlation with ground truth ($r = 0.725$) [8].

Moreover, RGB signal recognition deployed to quantify CD3 + lymphocytes and CD68 + macrophages sublining infiltration across different European centers showed unsatisfactory agreement (Intraclass correlation coefficients: 0.79 and 0.58 respectively) [9].

Conversely, the application of CNN-based computer vision may help to smooth and globally improve the workflow in the future. A benefit of CNNs is to provide a fast, reproducible and standardized tool to assist diagnosis. Indeed, the overall agreement between this model and pathologists is encouraging and suggests that CNNs have the potential to be employed in the clinical management of patients with rheumatic diseases when a USSB is needed.

Undoubtedly, this new technology needs to be integrated with pathologist's expertise to oversee and approve machine-based interpretations. Note that, at this stage, this algorithm does not answer all the questions rheumatologists may ask about synovial histological samples. Our model cannot yet identify "pathotypes" that are claimed to be informative for treatment personalization. Indeed, the implementation of this feature is conditional on the availability of larger datasets to train the algorithm. To accurately train the multi-class classifier that would be needed for this task, a researcher would conceivably need plenty of slices for "lympho-myeloid", "myeloid" and "pauci-immune" pathotypes [28]. Such data cannot in any way be substituted by image augmentation, due to the peculiar features of individual pathotypes. This, once more, poses the necessity to improve data sharing for visual contents and the creation of public image databases for machine-learning research, as it currently happens for melanoma detection. Given that CNNs seem to spontaneously attend to classical Krenn's score items found in the image—cells in the synovial lining, synovial stroma and inflammatory infiltrate—it is conceivable that, with adequate training and fine-tuning, CNNs could discriminate pathotypes based on distinctive histopathological patterns.

The ablation study indicated that the fine-tuning was of utmost importance for maximizing testing metrics. We observed that the fine-tuned ResNet34 seemed to rely on lining and sublining cellularity for discrimination. To explain the usefulness of fine-tuning, we hypothesize that the higher level feature maps found in the fine-tuned layers, which were learned during pretraining on the ImageNet dataset, needed to be adjusted specifically to improve classification accuracy on our dataset.

Finally, we must acknowledge that this was a pilot study with a small image dataset obtained from a low number of patients. Although promising, our model still needs to be prospectively validated in real-life cohorts.

Conclusions

This study shows that CNNs have the potential to accurately discriminate between low grade and high-grade synovitis. The application of CNN-based computer vision may help to smooth and globally improve the workflow practice between rheumatologists and pathologists. Further research is necessary to evaluate performance in a real-world clinical setting, to test this technique across the sample distribution and spectrum of synovitis patterns detected in daily practice. But such potential developments are conditional on the creation of large and open-access datasets, which historically have driven the development of machine learning.

Author contributions VV: Conceptualization; VV, GC, GL, AC: Data curation; VV, OA: Formal analysis; No Funding acquisition; VV, OA, EM, FI: Investigation; VV, OA: Methodology; VV: Project administration; VV: Resources; VV: Software; EM, FI; Supervision; VV, OA: Validation; VV, OA: Visualization; VV, OA, FI: Roles/Writing—original draft; VV, OA, FI: Writing—review and editing. VV and OA are joint first authors. This work is not related to OA's Amazon employment.

Funding Authors declare no financial supports for this work.

Availability of data and materials Image and patient database not available due to local ethical committee privacy issues.

Code availability Model available upon request.

Compliance with ethical standards

Competing interests The authors declare that they have no conflict of interest.

Statement of human and animal rights This study was performed according to the Declaration of Helsinki and study reached the approval by the ethical committee at the University of Bari, Italy as part of the Biopure registry (IRB approval n 5277/2017). The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation as required by the applicable law.

Informed consent Patients signed informed consent regarding publishing their data andhistopathological imagess.

References

1. Najm A, Le Goff B, Orr C, Thurlings R, Canete JD, Humby F, Alivernini S, Manzo A, Just SA, Romao VC, Krenn V, Muller-Ladner U, Addimanda O, Tas SW, Stoenoiu M, Meric de Bellefon L, Durez P, Strand V, Wechalekar MD, Fonseca JE, Lauwerys B, Fearon U, Veale DJ, Group ESS, Group OSTSI (2018) Standardisation of synovial biopsy analyses in rheumatic diseases: a consensus of the EULAR synovitis and OMERACT synovial tissue biopsy groups. Arthritis Res Ther 20(1):265. https://doi.org/10.1186/s13075-018-1762-1
2. Venerito V, Cazzato G, Lopalco G, Fornaro M, Righetti G, Urso L, Cimmino A, Iannone F (2019) Histopathologic features of fibrotic knee synovitis in a young adult with seronegative rheumatoid arthritis. J Clin Rheumatol. https://doi.org/10.1097/RHU.0000000000001247
3. Humby F, Buch M, Durez P, Lewis M, Bombardieri M, Rizvi H, Kelly S, Fosatti L, Hands R, Giorli G, Mahto A, Montecucco C, Lauwerys B, Romao V, Pratt A, Bugatti S, Ng N, Rivellese F, Ho P, Bellan M, Congia M, Verschueren P, Sainaghi P, Gendi N, Dasgupta B, Cauli A, Reynolds P, Cañete J, Moots R, Taylor P, Edwards C, Isaacs J, Sasieni P, Eurico Fonseca J, Choy E, Pitzalis C (2019) A Randomised, Open Labelled Clinical Trial to Investigate Synovial Mechanisms Determining Response – Resistance to Rituximab versus Tocilizumab in Rheumatoid Arthritis Patients Failing TNF Inhibitor Therapy. Arthritis Rheumatol 71 (suppl 10). https://acrabstracts.org/abstract/a-randomised-open-labelled-clinical-trial-to-investigate-synovial-mechanisms-determining-response-resistance-to-rituximab-versus-tocilizumab-in-rheumatoid-arthritis-patients-failing-tnf-inhibitor-t/. Accessed 4 Dec 2020
4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118. https://doi.org/10.1038/nature21056
5. Srinidhi CL, Ciga O, Martel AL (2019) Deep neural network models for computational histopathology: A survey.
6. Gertych A, Swiderska-Chadaj Z, Ma Z, Ing N, Markiewicz T, Cierniak S, Salemi H, Guzman S, Walts AE, Knudsen BS (2019) Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. Scientific Reports 9(1):1483. https://doi.org/10.1038/s41598-018-37638-9
7. Pandit A, Radstake T (2020) Machine learning in rheumatology approaches the clinic. Nat Rev Rheumatol 16(2):69–70. https://doi.org/10.1038/s41584-019-0361-0
8. Morawietz L, Schaeper F, Schroeder JH, Gansukh T, Baasanjav N, Krukemeyer MG, Gehrke T, Krenn V (2008) Computer-assisted validation of the synovitis score. Virchows Arch 452(6):667–673. https://doi.org/10.1007/s00428-008-0587-8
9. Rooney T, Bresnihan B, Andersson U, Gogarty M, Kraan M, Schumacher HR, Ulfgren AK, Veale DJ, Youssef PP, Tak PP (2007) Microscopic measurement of inflammation in synovial tissue: inter-observer agreement for manual quantitative, semi-quantitative and computerised digital image analysis. Ann Rheum Dis 66(12):1656–1660. https://doi.org/10.1136/ard.2006.061143
10. Krenn V, Morawietz L, Burmester GR, Kinne RW, Mueller-Ladner U, Muller B, Haupl T (2006) Synovitis score: discrimination between chronic low-grade and high-grade synovitis. Histopathology 49(4):358–364. https://doi.org/10.1111/j.1365-2559.2006.02508.x
11. Wechalekar MD, Najm A, Veale DJ, Strand V (2019) The 2018 OMERACT synovial tissue biopsy special interest group report on standardization of synovial biopsy analysis. J Rheumatol. https://doi.org/10.3899/jrheum.181062
12. Shahin M, Maier H, Jaksa M (2004) Data division for developing neural networks applied to geotechnical engineering. J Comput Civil Eng. https://doi.org/10.1061/(ASCE)0887-3801(2004)18:2(105)
13. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press
14. Brunese L, Mercaldo F, Reginelli A, Santone A (2020) Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. Comput Methods Programs Biomed 196:105608. https://doi.org/10.1016/j.cmpb.2020.105608
15. Kensert A, Harrison PJ, Spjuth O (2019) Transfer learning with deep convolutional neural networks for classifying cellular morphological changes. SLAS Discov 24(4):466–475. https://doi.org/10.1177/2472555218818756
16. Vu CC, Siddiqui ZA, Zamdborg L, Thompson AB, Quinn TJ, Castillo E, Guerrero TM (2020) Deep convolutional neural networks for automatic segmentation of thoracic organs-at-risk in radiation oncology—use of non-domain transfer learning. J Appl Clin Med Phys 21(6):108–113. https://doi.org/10.1002/acm2.12871
17. He K, Zhang X, Ren S, Sun J (2015) Deep Residual Learning for Image Recognition.
18. Howard J, Gugger S (2020) fastai: A Layered API for Deep Learning.
19. Deng J, Dong W, Socher R, Li L-J, Li K, Li FF (2009) ImageNet: a large-scale hierarchical image database. IEEE. https://doi.org/10.1109/CVPR.2009.5206848
20. Kaur T, Gandhi TK (2020) Deep convolutional neural networks with transfer learning for automated brain image classification. Mach Vis Appl 31(3):20. https://doi.org/10.1007/s00138-020-01069-2
21. Monshi M, Poon J, Chung V (2019) Convolutional neural network to detect thorax diseases from multi-view chest X-rays. In: Gedeon T, Wong K, Lee M (eds) Neural Information Processing. ICONIP 2019. Communications in computer and information science. Springer, Cham, pp 148–158. https://doi.org/10.1007/978-3-030-36808-1_17
22. Smith LN (2018) A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay.
23. He T, Guo J, Chen N, Xu X, Wang Z, Fu K, Liu L, Yi Z (2020) MediMLP: using Grad-CAM to extract crucial variables for lung cancer postoperative complication prediction. IEEE J Biomed Health Inform 24(6):1762–1771. https://doi.org/10.1109/JBHI.2019.2949601
24. Iannone F, Nivuori M, Fornaro M, Venerito V, Cacciapaglia F, Lopalco G (2019) Comorbid fibromyalgia impairs the effectiveness of biologic drugs in patients with psoriatic arthritis.

Rheumatology (Oxford). https://doi.org/10.1093/rheumatology/kez505
25. Lopalco G, Venerito V, Cantarini L, Emmi G, Salaffi F, Di Carlo M, Tafuri S, Gentileschi S, Di Scala G, Nivuori M, Cacciapaglia F, Galeazzi M, Lapadula G, Iannone F (2019) Different drug survival of first line tumour necrosis factor inhibitors in radiographic and non-radiographic axial spondyloarthritis: a multicentre retrospective survey. Clin Exp Rheumatol 37(5):762–767
26. Venerito V, Lopalco G, Cacciapaglia F, Fornaro M, Iannone F (2019) A Bayesian mixed treatment comparison of efficacy of biologics and small molecules in early rheumatoid arthritis. Clin Rheumatol 38(5):1309–1317. https://doi.org/10.1007/s10067-018-04406-z
27. Orr C, Vieira-Sousa E, Boyle DL, Buch MH, Buckley CD, Canete JD, Catrina AI, Choy EHS, Emery P, Fearon U, Filer A, Gerlag D, Humby F, Isaacs JD, Just SA, Lauwerys BR, Le Goff B, Manzo A, McGarry T, McInnes IB, Najm A, Pitzalis C, Pratt A, Smith M, Tak PP, Thurlings R, Fonseca JE, Veale DJ (2017) Synovial tissue research: a state-of-the-art review. Nat Rev Rheumatol 13(10):630. https://doi.org/10.1038/nrrheum.2017.161
28. Humby F, Lewis M, Ramamoorthi N, Hackney JA, Barnes MR, Bombardieri M, Setiadi AF, Kelly S, Bene F, DiCicco M, Riahi S, Rocher V, Ng N, Lazarou I, Hands R, van der Heijde D, Landewe RBM, van der Helm-van MA, Cauli A, McInnes I, Buckley CD, Choy EH, Taylor PC, Townsend MJ, Pitzalis C (2019) Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. Ann Rheum Dis 78(6):761–772. https://doi.org/10.1136/annrheumdis-2018-214539