Deep Learning for Classification of Non-Small Cell Lung Cancer histologic subtypes

Elvis Murina[1], Ruben Casanova[1-2], Hanna Honcharova-Biletska[2], Bart Vrugt[2], Alex Soltermann[2], Oliver Dürr[3*], Beate Sick[1*]

1. Zurich University of Applied Sciences, School of Engineering, Winterthur, Switzerland
2. Institute of Pathology and Molecular Pathology, University Hospital Zurich, Zurich, Switzerland
3. HTWG Konstanz University of Applied Sciences, Konstanz, Germany

Motivation

- Non-Small Cell Lung Cancer (NSCLC) is the main lung cancer type and comprises two main histologic classes: adenocarcinoma (ADC) and squamous cell carcinoma (LSCC).
- Tumor subtyping is routinely done by trained pathologists using conventional microscopy. Automated image analysis can potentially speed-up routine diagnostic.
- In clinical practice, histologic subtyping is an important factor for treatment decisions, as different therapies are proposed to non-squamous NSCLC.
- Deep learning methods have become increasingly popular in recent years, mostly because they can outperform traditional machine learning algorithms, without requiring extraction of handcrafted features.
- In this work we show the potential of deep learning with convolutional neural networks (CNN) in the classification of histologic images of lung cancer.

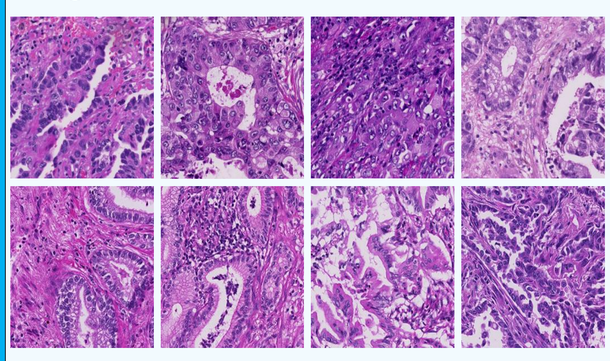
Sample preparation workflow

1. Tumor resection
2. FFPE block preparation
3. Whole section H&E staining
4. Whole section scanning
5. Digital library preparation
6. Convolutional neural network training

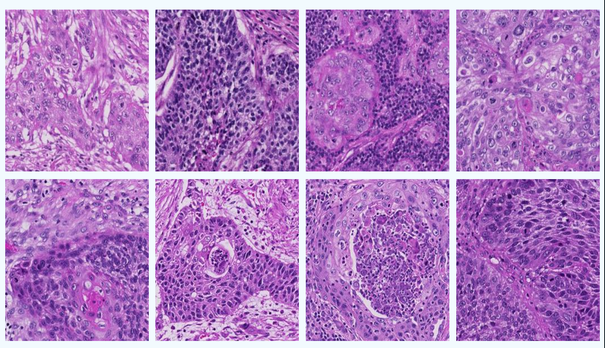
Probability for ADC and LSCC

Legend: FFPE=Formalin-Fixed, Paraffin-Embedded (tissue); H&E=Hematoxylin and eosin stain

Lung adenocarcinoma

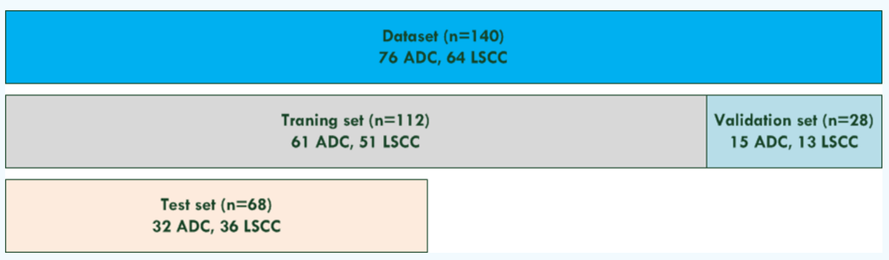


Lung squamous cell carcinoma



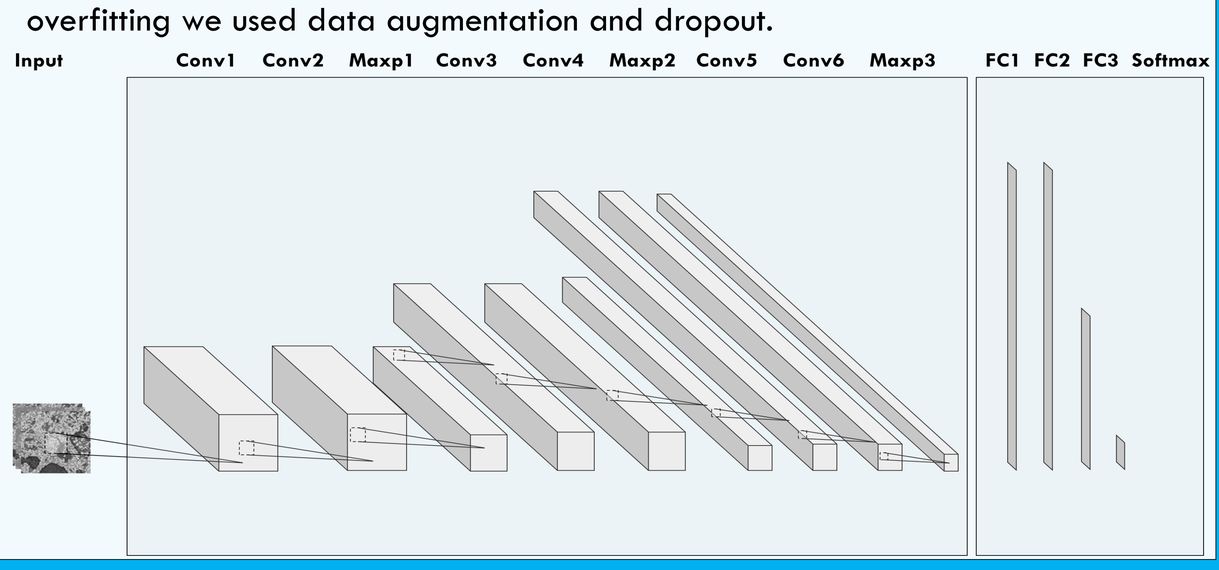
Examples of NSCLCs, stained by Hematoxylin and eosin (H&E)

Data and Network architecture

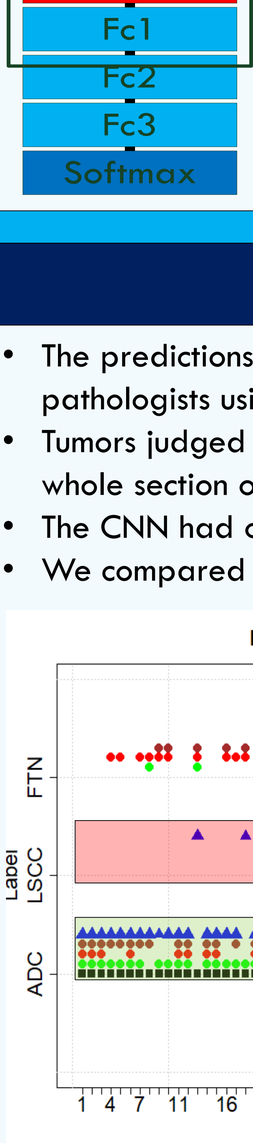


Original	(256x256x3)
Input	(128x128x3)
Conv1	(128x128x32)
Conv2	(128x128x32)
Maxp1	(64x64x32)
Conv3	(64x64x64)
Conv4	(64x64x64)
Maxp2	(32x32x64)
Conv5	(32x32x128)
Conv6	(32x32x128)
Maxp3	(16x16x128)
FC1	(200)
FC2	(200)
FC3	(50)
Softmax	(2)

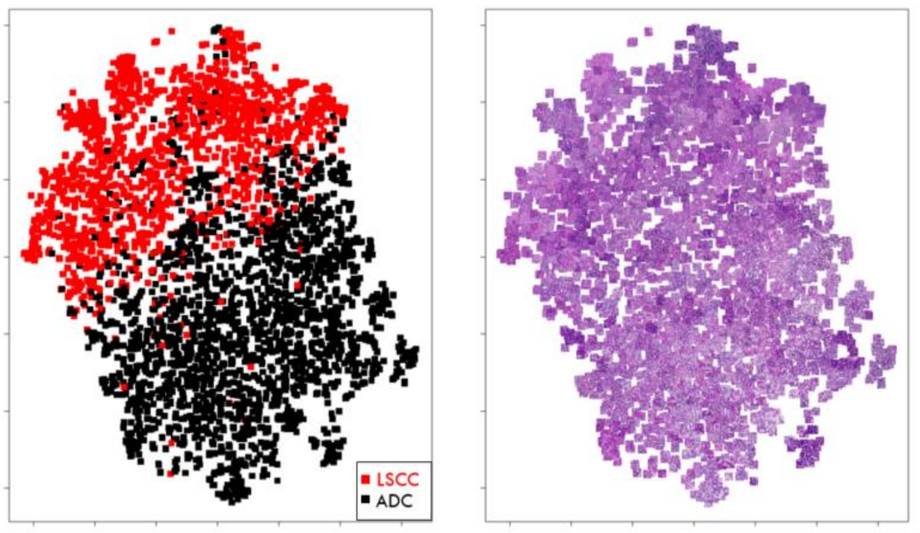
- We developed and trained a CNN to differentiate ADC from LSCC. Patients were split into a training and validation set. Each patient contains 50 histologic images.
- The performances of the CNN were evaluated on the test set and are compared with the results of three experienced pathologists.
- We merged the 50 predictions for each patient via majority vote to a final prediction with a confidence score.
- As activation function ReLu was used, and to prevent overfitting we used data augmentation and dropout.



Unsupervised tsne visualization



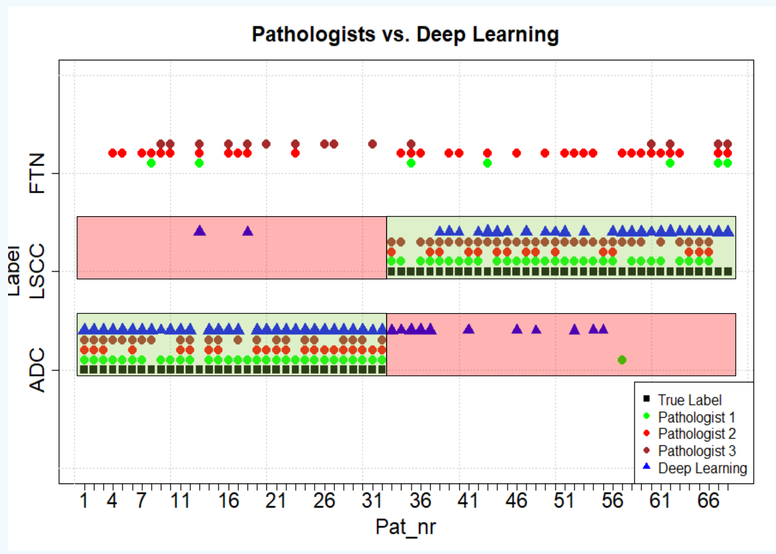
- A pretrained VGG16 network, trained on ImageNet data, was used as a feature extractor for the histological images.
- 4096 features of the third last layer ("Fc1") before the activation function were extracted for each image.
- Here we show results of a 2D tsne representation of all images in the training set. If you take a close look, you see that similar images are close to each other.



- There is a very good separation between ADC and LSCC. The labels were only used to color the tsne plot, this is a completely unsupervised approach.

Results

- The predictions of our network were compared to the predictions of three pathologists using the same image set.
- Tumors judged too difficult to evaluate by the pathologists, without access to the whole section or additional staining, were scored as FTN (further tests needed).
- The CNN had only two choices: ADC or LSCC.
- We compared the number of evaluated patients with the accuracy on those.



Deep Learning	#evaluated	%	Acc.
Cs>0.5	68 / 68	100%	80.8%
Cs>0.6	60 / 68	88.2%	86.6%
Cs>0.7	58 / 68	85.3%	87.9%
Cs>0.8	43 / 68	63.2%	93.0%

Cs: Confidence score

Pathologists	#evaluated	%	Acc.
patho 1	61 / 68	89.7%	98.4%
patho 2	36 / 68	52.9%	100.0%
patho 3	53 / 68	77.9%	100.0%

Conclusion / Outlook

- In this work we showed the potential of CNNs in that routine classification tasks with accuracies almost approaching those from trained pathologist with years of experience.
- The performances of the network could be improved by increasing the training set, or using images with different magnifications.
- Further improvements can be expected by choosing other network architecture or using batch normalization layers. In addition, dropout in the forward pass could also be used to get a confidence score for every image.
- An important future step includes classification visualization, to understand which parts of the image contribute to the class decision.

References

- Single-Cell Phenotype Classification Using Deep Convolutional Neural Networks (Oliver Dürr, Beate Sick, 2016)
- Very Deep Convolutional Networks for Large-Scale Image Recognition (Karen Simonyan, Andrew Zisserman, 2015)
- Visualizing Data using t-SNE (Laurens van der Maaten, Geoffrey Hinton, 2008)