

Using deep learning to predict a gene expression-based breast cancer proliferation score from H&E WSIs

Andreas Ekholm, Yinxu Wang, Johan Hartman, Mattias Rantalainen (Karolinska Institute, Sweden)

Introduction

- Breast cancer (BC) is the most common cancer in women.
- Routine BC pathology classification only provides limited risk stratification of patients.
- Gene expression profiling can be utilised as a potential improvement to estimate proliferation scores for patients, but requires lab-based assays that are costly and time-consuming.

Aim

- The aim in this study was to determine if deep convolutional neural networks (CNNs) can be applied to predict a gene expression-based proliferation score (1) directly from routine H&E whole slide images (WSIs).

Material and Methods

- WSIs and RNA-sequencing data from N=991 surgically treated invasive BC patients were included from Clinseq (2) and TCGA (3) datasets.
- A deep CNN model (Inception V3) was optimised to predict the continuous proliferation score (4) from WSIs using the RNA-seq proliferation score as ground truth.
- Prediction performance was evaluated in independent test data by Spearman correlation between the predicted and the RNA-seq proliferation score.
- Prognostic performance was assessed through Cox proportional hazard modelling, estimating hazard ratios (HR) for recurrence-free survival.

Results

- The RNA-seq estimated proliferation score could be predicted from WSIs using the CNN model with a Spearman correlation of 0.691 ($p < 0.001$) (Figure 1).
- The RNA-seq estimated proliferation score was found to be associated with tumour grade (NHG) and BC subtype.
- A high proliferation score was significantly associated with higher risk of recurrence or death in univariate analysis HR=1.5 (95 % CI 1.0-2.1) $p=0.028$ (RNA-seq) and HR=1.5 (95 % CI 1.0-2.1) $p=0.039$ (CNN predicted).

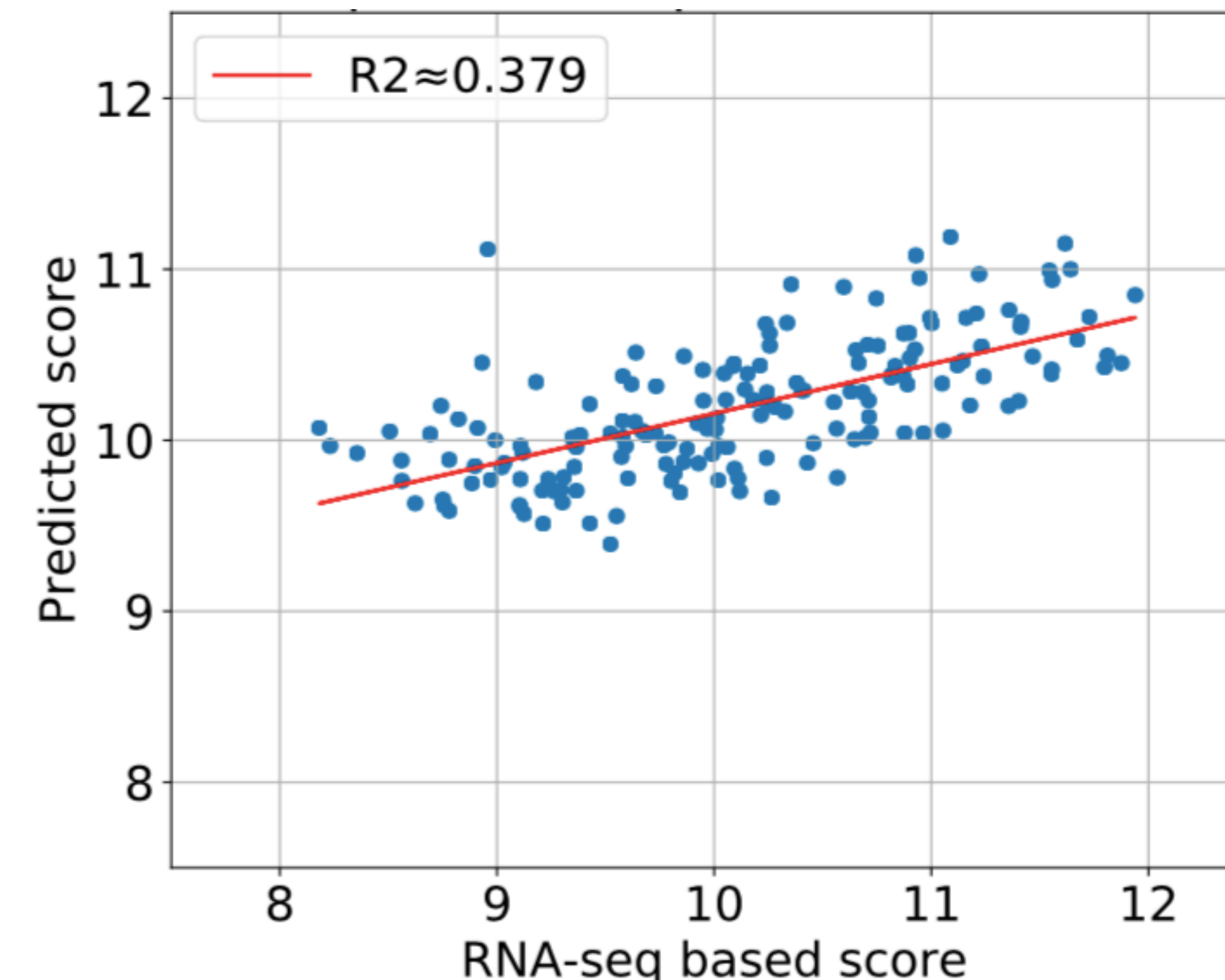


Figure 1. Proliferation score prediction in an independent test dataset of 172 patients, using RNA-seq as ground truth.

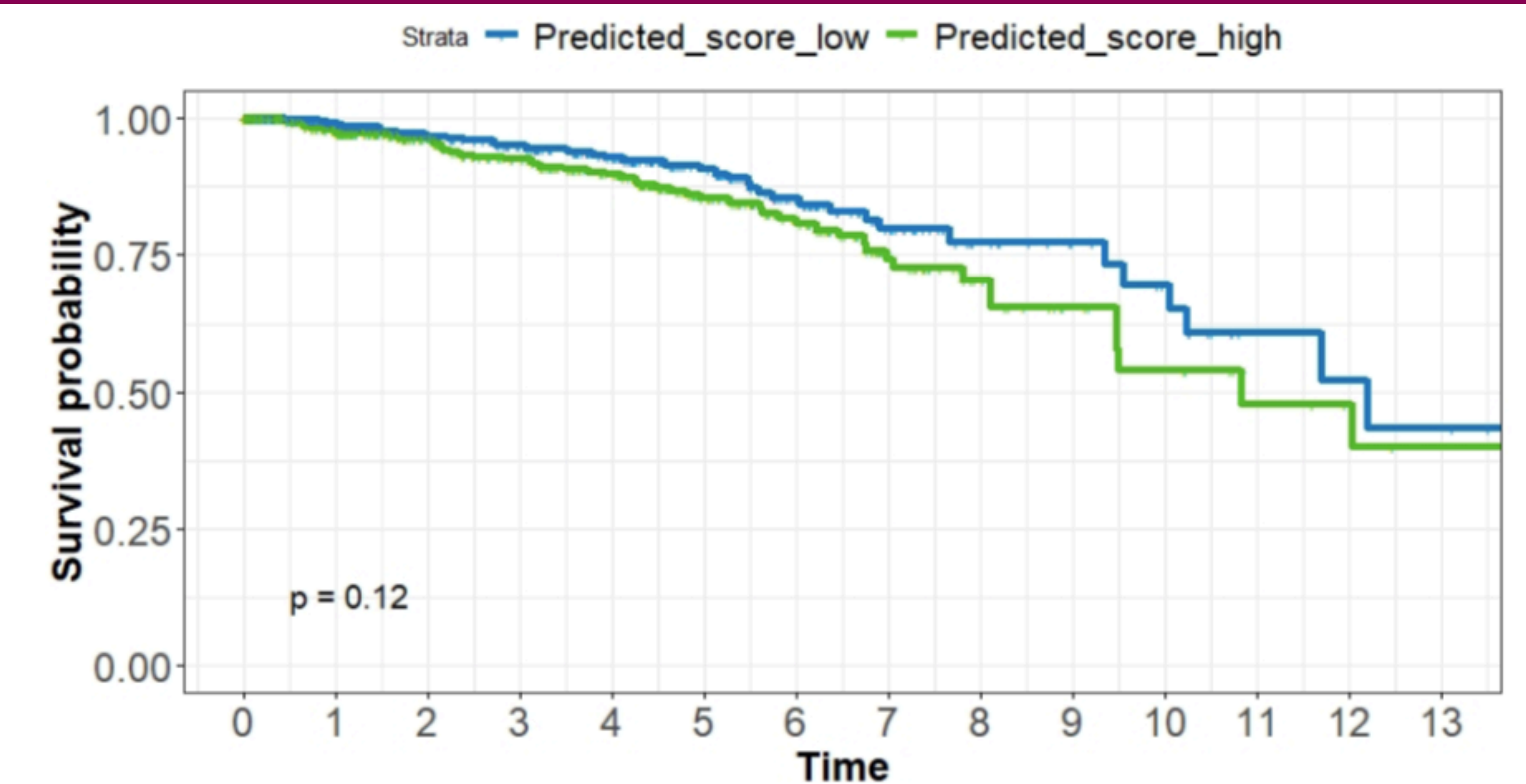


Figure 2. Kaplan-Meier curve (time-to-event analysis) based on the predicted proliferation scores – dichotomised into high and low score group by the median score.

Conclusions

The results suggest that a gene expression-based proliferation score for breast cancer can be predicted from morphological patterns in routine H&E stained WSIs using a deep CNN model, with comparable prognostic performance to the RNA-seq estimated proliferation score.

References

- Veta M, Heng YJ, Stathonikos N, Bejnordi BE, Beca F, Wollmann T, et al. Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge. *Med Image Anal.* 2019;54:111-21.
- Wang M, Klevebring D, Lindberg J, Czene K, Grönberg H, Rantalainen M. Determining breast cancer histological grade from RNA-sequencing data. *Breast Cancer Res.* 2016;18(1):48.
- The Cancer Genome Atlas Network., Genome sequencing centres: Washington University in St Louis., Koboldt, D. et al. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490:61–70.
- Munir K, Elahi H, Ayub A, Frezza F, Rizzi A. Cancer Diagnosis Using Deep Learning: A Bibliographic Review. *Cancers (Basel).* 2019;11(9).



Karolinska Institutet
Andreas Ekholm (Medical Student)
E-mail: andreas.ekholm@stud.ki.se
Department of Medical Epidemiology and Biostatistics
Karolinska Institute
Stockholm, Sweden



**Karolinska
Institutet**