UiT The Arctic University of Norway

Faculty of Health Sciences

Developing a machine learning model for tumor cell quantification in standard histology images of lung cancer

Siri Kristoffersen

Main supervisor: Mehrdad Rakaee, PhD, Department of Clinical Pathology, UNN

Co-supervisor: Elin Richardsen, MD, PhD, Department of Clinical Pathology, UNN

Co-supervisor: Lill-Tove Rasmussen Busund, MD, PhD, Department of Medical Biology, UiT

Master's thesis in medicine MED-3950 June 2023

Acknowledgement

This master thesis was realized thanks to a collaboration between the Translational Cancer Research Group, UiT and the Department of Clinical Pathology at UNN. Dr. Mehrdad Rakaee was my main supervisor, and Prof. Elin Richardsen and Prof. Lill-Tove Rasmussen Busund were my co-supervisors. I would like to express my great appreciation to my supervisors for the work and time that you have put in to guide me through my work in this master thesis. I would especially like to thank Elin Richardsen for introducing me to the research group, and Mehrdad Rakaee for teaching me about the processes of machine learning. The process of the writing of this master thesis has been a challenging, interesting and educational supplement to the course of my medical studies. It has inspired to further interest in research.

I will also like to say thank you to all the patients who have agreed to participate in the research project. None of the research done in this thesis would have been possible without their participation.

Table of content

Acknowledgement ..... **i**

Summary ..... **iii**

Abbreviations ..... **iv**

Introduction **1**

Lung cancer *1*

Epidemiology 1

Risk factors 2

Diagnostics 3

Treatment 4

Tumor cellularity *6*

Computational pathology *6*

Cell/object-based vs pixel-based ML classifier 8

The aim of this thesis *10*

Methods **11**

Material *11*

Manual tumor cell estimation *11*

Digital tumor cell quantification *11*

Statistical analysis *12*

Ethic clearance *13*

Results **14**

Patients *14*

Manual tumor purity estimation *15*

Developing ML classifier *17*

Digital tumor purity estimation *18*

Discussion **21**

Conclusion **25**

Bibliography **26**

Supplementary **30**

Summary

Background

Tumor purity estimation plays a crucial role in genomic profiling and is traditionally carried out manually by pathologists. This manual approach has several disadvantages, including potential inaccuracies due to human error, inconsistency in evaluation criteria among different pathologists, and the time-consuming nature of the process. These issues may be addressed by adopting a digital approach. In this thesis, we employ a machine learning (ML)-based, cell-based classifier to estimate tumor purity in lung cancer tissues.

Materials and methods

In this study, conducted as part of the subsequent clinical trial TNM-I, we incorporated 61 patients diagnosed with non-small cell lung cancer (NSCLC). Tumor purity was initially estimated manually by two pathologists. The digital estimation of tumor purity was executed using a ML-based classifier in QuPath. To determine the level of agreement and inter-rater reliability between the two pathologists, as well as between the manual and digital estimations, we computed Intraclass Correlation Coefficient (ICC) and Cohen's Kappa using SPSS.

Results

The ICC coefficient when comparing the tumor purity estimations done by the two pathologists was 0.833, indicating good reliability. According to Cohen's Kappa the inter-rater reliability between the pathologists was moderate with a value of 0.534. The ICC coefficient when comparing the manual and digital tumor purity estimation was 0.838, which indicates good reliability. When analyzing for Cohen's Kappa we got a value of 0.563, indicating moderate inter-rater reliability between the tumor purity estimations done manually and digitally. All the results were statistically significant.

Conclusion

In summary, we have successfully developed a ML classifier that estimates tumor purity in lung cancer tissue. Our findings align with previous research and demonstrate strong correlation with traditional detection methods. These results underscore the importance of continuing research in enhancing ML-based strategies for tumor purity estimation.

Abbreviations

AI - Artificial intelligence

ALK - Anaplastic lymphoma kinase

BRAF - V-raf murine sarcoma viral oncogene homolog B1

CLIA - Clinical Laboratory Improve Amendments

CT - Computed tomography

DL - Deep learning

EGFR - Epidermal growth factor receptor

HER2 - Human epidermal growth factor receptor 2

H&E - Hematoxylin and eosin

ICC - Intraclass correlation coefficient

ICIs - Immune checkpoint inhibitors

KA - Krippendorff's alpha

KRAS - Kirsten rat sarcoma virus

LUAD - Lung adenocarcinoma

LUSC - Lung squamous cell carcinoma

MDT – Multidisciplinary team

MET - Mesenchymal epithelial transition

ML - Machine learning

MRI - Magnetic resonance imaging

NSCLC - Non-small cell lung cancer

NTRK - Neurotrophic tyrosine receptor kinase

OD – Optical density

PD-1 - Programmed Cell Death Protein 1

PD-L1 - Programmed Cell Death Ligand 1

PET-CT - Positron emission tomography–computed tomography

REDCap - Research electronic data capture

REK – Regional committee

RET - Rearranged during transfection

ROS1 - Proto-oncogene tyrosine-protein kinase

SCLC - Small cell lung cancer

sTILs - Stromal tumor-infiltrating lymphocytes

TNM-I - Tumor-node-metastasis-immunoscore

UNN - University Hospital of North Norway

Introduction

Lung cancer

Epidemiology

Lung cancer is the most common form of cancer worldwide (1, 2). In Norway, lung cancer is the second most common cancer in both men and women, as shown in **Figure 1** (2, 3). In the period 2017-2021, lung cancer accounted for approximately 10% of all newly diagnosed cancer cases. There were 3685 new cases of lung cancer registered in 2021 (4). The incidence is rising steadily, while the prevalence has tripled in the past 20 years (2). The increase in prevalence can be explained by better survival due to improved guidelines for examination, treatment, and follow-up of patients. In the last 20 years, the survival rate for patients with lung cancer has almost doubled. Even though the survival from lung cancer is increasing, lung cancer is still the form of cancer that takes the most lives. In 2021, 1190 men and 1053 women died of lung cancer in Norway (4). A study calculated that in 2012 almost 33,000 years of life were lost due to lung cancer, roughly the same number as breast cancer, prostate cancer and colon cancer combined (5). Worldwide, it is estimated that 1.8 million die from lung cancer annually (1, 2).

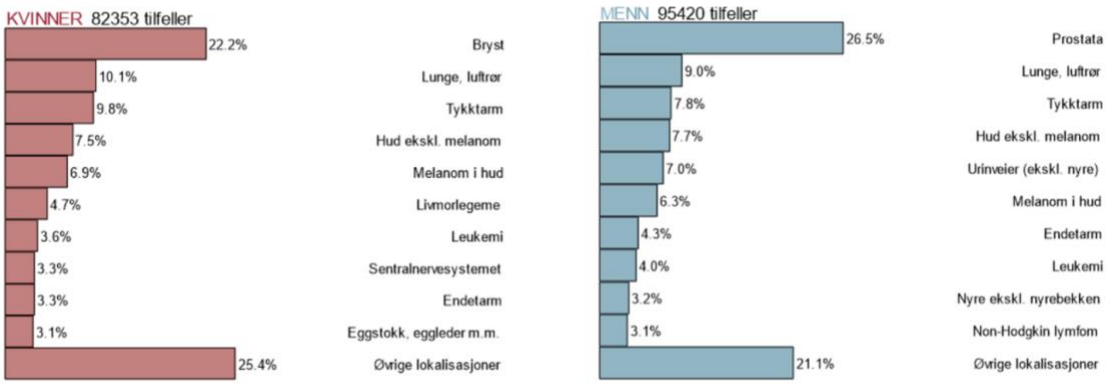


Figure 1: Illustrating that lung cancer is the second most common form of cancer in Norwegian men (blue) and women (red). The two figures are retrieved from the Norwegian Cancer Registry's website. They illustrate the most common forms of cancer in percentage for men and women in the period 2017-2021 (3)

Risk factors

Tobacco smoking is the major risk factor of all major histological types of lung cancer (6). As much as 8 out of 10 lung cancer cases are due to tobacco use (4). The increased risk of lung cancer is 20 to 50 times greater in smokers compared to never smokers (6). Duration of smoking is the strongest risk factor for lung cancer among smokers (6, 7). It has been shown that quitting tobacco use has a beneficial effect on relative risk, even if you quit smoking late in life. It is still important to mention that an excess risk throughout life probably persists even in long-term quitters (6, 8). There is a casual association between second-hand exposure to cigarette smoke and lung cancer risk in nonsmokers. A nonsmoker who is married to a smoker will have 20-30% excess risk for developing lung cancer themselves (6, 9, 10). When it comes to diet and lung cancer, case-control studies have shown that a rich diet in vegetables and fruits may exert some protective effect against lung cancer (6, 11, 12). In contrast, intake of meat may increase the risk of lung cancer. Especially fried or well-done red meat increases the risk (6, 13). It is difficult to assess the relationship between alcohol consumption and the risk of lung cancer, since the correlation between alcohol use and tobacco smoking is strong in many populations (6).

Studies have suggested that patients with chronic obstructive pulmonary disease, such as asthma, are at increased risk for lung cancer (6, 14-16). The same applies to patients with tuberculosis (6, 17). Whether the excess risk in tuberculosis patients is caused by the chronic inflammation in the lung parenchyma or by the Mycobacterium itself is unclear (6). When it comes to indoor air pollution it has been reported positive association between indoor air pollution and lung cancer risk in Europe, but also in several regions of Asia (6, 18). Indoor air pollution is created by poorly ventilated houses, burning of wood and fumes from high-temperature cooking using unrefined vegetable oils (6, 19).

Ionizing radiation increases the risk of lung cancer, which can be seen in studies done on atomic bomb survivors and patients treated with radiotherapy (6, 20). Exposure to radioactive radon and its decay products have been found to increase risk of lung cancer in underground miners (6, 21). However, today the main concern about lung cancer risk due to radon exposure comes from residential rather than occupational exposure (6). In Norway around 12% of cases of lung cancer can probably be attributed to exposure to radioactive radon gas in the bedrock in several areas (4, 22). For smokers who are also exposed to radon gas, the risk of lung cancer is particularly high. This can be explained by the two factors working in

different ways to promote lung cancer, but also because radon gas binds to particles in the tobacco smoke (4, 22).

In Norway, asbestos and radon are considered the most important additional factors for the development of lung cancer in addition to smoking (4). In many low- and medium- resource countries the occupational exposure to asbestos remains widespread. Asbestos together with silica, polycyclic aromatic hydrocarbons, diesel exhaust and metals like arsenic, chromium and nickel make up the substances which increases lung cancer risk in occupational exposure (6).

Diagnostics

In Norway, we have a package procedure so called “pakkeforløp” when it comes to diagnostics, treatment and follow-up of cancer patients. These package procedures contain deadlines so that patients with cancer receive the most optimal treatment possible. In order for a patient to be referred to the lung cancer package procedure, certain criteria must be met. One criteria is that if a chest X-ray or chest computed tomography (CT) results in suspicion of lung cancer, the patient must be referred to the lung cancer package procedure. A patient may also be referred to a lung cancer package procedure despite a normal chest X-ray if there is still a strong clinical suspicion of lung cancer, such as in the case of unexplained dyspnea and hemoptysis lasting more than a week. In two cases, a patient should be referred to the lung cancer package procedure without waiting for the results on the chest X-ray first. This applies to persistent hemoptysis in smokers or ex-smokers over the age of 40 or to signs of obstruction of the superior vena cava (23).

When the patient has been referred in the package procedure, the diagnostics starts. First by clarifying whether it is a malignant disease or not. If it proves to be cancer, an early decision must be made as to whether the condition is limited or widespread. This is done to decide if the treatment should be curative or palliative. In the case of a potentially curative condition, the investigation must be comprehensive, so that the histological diagnosis, distribution and state of health can be clarified so that further treatment can be adapted. In contrast, the examinations should be limited if it is an obvious palliative condition (24).

The further course of investigation can be divided into two main categories based on whether the tumor is located centrally or peripherally in the lung or whether there are distant

metastases. The main goal in both categories is to get the best possible biopsy from the tumor. Centrally located tumors can be biopsied by using bronchoscopy or endobronchial ultrasound examination. While one can use ultrasound or CT guided needle biopsy to take samples from peripheral located tumors. Spread to mediastinal lymph nodes can be investigated using needle biopsy guided by endobronchial or endoscopic ultrasound examination, or by doing mediastinoscopy. Assessment of any metastases to the adrenal gland, liver, brain and other organs is done using CT, magnetic resonance imaging (MRI) or scintigraphy, depending on location. Patients considered for curative treatment undergo further investigations, such as Extended Pulmonary Function Test, gait test, cardiopulmonary stress test, cardiology assessment and positron emission tomography–computed tomography (PET-CT). In general when diagnosing lung cancer, histological diagnosis is important for treatment selection and it forms the basis for further individualized treatment. For this reason, a definitive histological diagnosis is sought in all patients before treatment (24).

Treatment

Determination of the diagnosis and treatment decision is made in multidisciplinary team (MDT) meetings, which is an interdisciplinary meeting where, for example, a pulmonologist, pathologist, radiologist, surgeon and oncologist meet and discuss each individual patient (24, 25). Lung cancer is treated with surgery, radiotherapy or chemotherapy, either alone or in various combinations. Disease extent, tissue type as well as the patient's general condition and any other concurrent illness determine the choice of treatment. When choosing treatment, a distinction is made between non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), where curative treatment is rarely relevant for SCLC (26, 27). NSCLC is further divided into stages which in turn help to determine the treatment. At stage I, the lung tumor is less than 5 centimeters without spread to lymph nodes, while at stage II the primary tumor is larger or there is spread to lymph nodes near the primary tumor. At stage III, there is either spread to lymph nodes in the breast septum or the tumor has grown into surrounding structures. At stage IV, distant metastases have been detected (27).

For NSCLC in stage I or II, the primary treatment is surgery, given that it is technically and medically possible. After the operation, patients under the age of 70 and in stage II will benefit from chemotherapy. Radiotherapy is recommended postoperatively if there is a possibility of remaining cancer tissue. Curative radiation therapy may be appropriate as

primary treatment for stage I and II if the patient has concurrent heart disease or other lung disease that may cause the patient to be unable to tolerate the surgery (27).

In stage III NSCLC, surgery is usually not an option. When choosing treatment for patients with stage III, prognostic factors are decisive. If the patient has good prognostic factors, curative radiation therapy should be given, usually in combination with chemotherapy. In cases with poor prognostic factors, the patient is offered palliative treatment, in the form of radiotherapy or chemotherapy. In some cases, at stage III, chemotherapy and radiotherapy can be given in an attempt to make the tumor operable. In the case of stage IV NSCLC, that is metastatic disease, the patient must be offered palliative treatment with radiotherapy or chemotherapy. Targeted treatment with tablets may be relevant for specific mutations (27).

Over the past decade, a greater understanding of lung cancer biology at the molecular level has led to the development of targeted therapy and immunotherapy in the treatment of NSCLC (28). The principle behind immunotherapy is to stimulate the body's own immune system to target the cancer cells, whereas targeted therapy is treatment that targets specific genes and proteins found in cancer cells that are responsible for driving the cancer cells to grow and spread (29). Immune checkpoint inhibitors (ICIs) have shown enormous value in the treatment of NSCLC and is now a standard part of the treatment for patients with metastatic, locally advanced, and resectable NSCLC. Programmed Cell Death Protein 1/Programmed Cell Death Ligand 1 (PD-1/PD-L1) and cytotoxic T-lymphocyte associated protein 4 (CTLA-4) are the most common targets for ICIs, as they are utilized by the tumor to evade the host's immune system. ICIs is an integral part of the treatment of NSCLC without a driver mutation, whereas targeted treatment is an option for patients with driver mutation positive NSCLC (28). Currently there are eight targets in cancer cells that can be attacked by targeted treatment; epidermal growth factor receptor (EGFR), kirsten rat sarcoma virus (KRAS), anaplastic lymphoma kinase (ALK), mesenchymal epithelial transition (MET), proto-oncogene tyrosine-protein kinase (ROS1), v-raf murine sarcoma viral oncogene homolog B1 (BRAF), rearranged during transfection (RET), and neurotrophic tyrosine receptor kinase (NTRK) (30). A problem with immunotherapy, including ICIs, is that a substantial number of patients with NSCLC do not benefit from the treatment and rather experience only toxicity. To solve this problem new predictive immunotherapy biomarkers needs to be found in order to predict who will respond to ICIs (28).

Tumor cellularity

Manual microscopic histopathological evaluation of tumor tissue has been the gold standard for subtyping malignancies and deciding on treatment strategies for cancer patients (31). However, this approach has limitations, as the tumor tissue and microenvironment consist of neoplastic and various non-neoplastic cells, such as normal epithelial cells, immune cells, fibroblasts, endothelial cells, and others (32). When molecular genomic profiling is performed on bulk tissue samples, the resulting output is often influenced by the presence of non-neoplastic cells (33, 34). Therefore, tumor purity, which accounts for the fraction of neoplastic cells in the tissue, needs to be estimated accurately (34, 35). Pathologists typically estimate tumor purity by reviewing tumors or by using DNA-based tools (34). In addition, tumor cellularity, or the level of neoplastic cells in a tissue sample, can affect the inclusion or exclusion of patients for genomic profiling, as well as downstream variant analysis (36). For example, sequencing platforms require a sufficient fraction of neoplastic cells (>20%) in the tumor tissue to identify genomic variants accurately (37). It is crucial to determine the tumor cellularity accurately, as it can impact treatment decisions and patient outcomes (36).

In other hand, manual evaluation of histological sections by an expert pathologist is a time-consuming and subjective process, which can lead to low interobserver agreement among pathologists for tumor purity estimation (38). Therefore, efforts should have been made to develop automated methods for accurate and reproducible estimation of tumor purity and cellularity. These methods include machine learning (ML)-based algorithms that use morphological features to classify neoplastic and non-neoplastic cells and DNA-based methods that quantify the proportion of neoplastic cells in a sample.

Computational pathology

Computational pathology has emerged as a promising field in recent years, revolutionizing the way we analyze and interpret biomedical images. Digital pathology technologies have greatly improved the image quality, making it easier to analyze large volumes of data at scale. In particular, artificial intelligence (AI) based image analysis platforms have the potential to enable rapid, reproducible, and quantitative analysis of complex pathology images (**Figure 2**) (39).

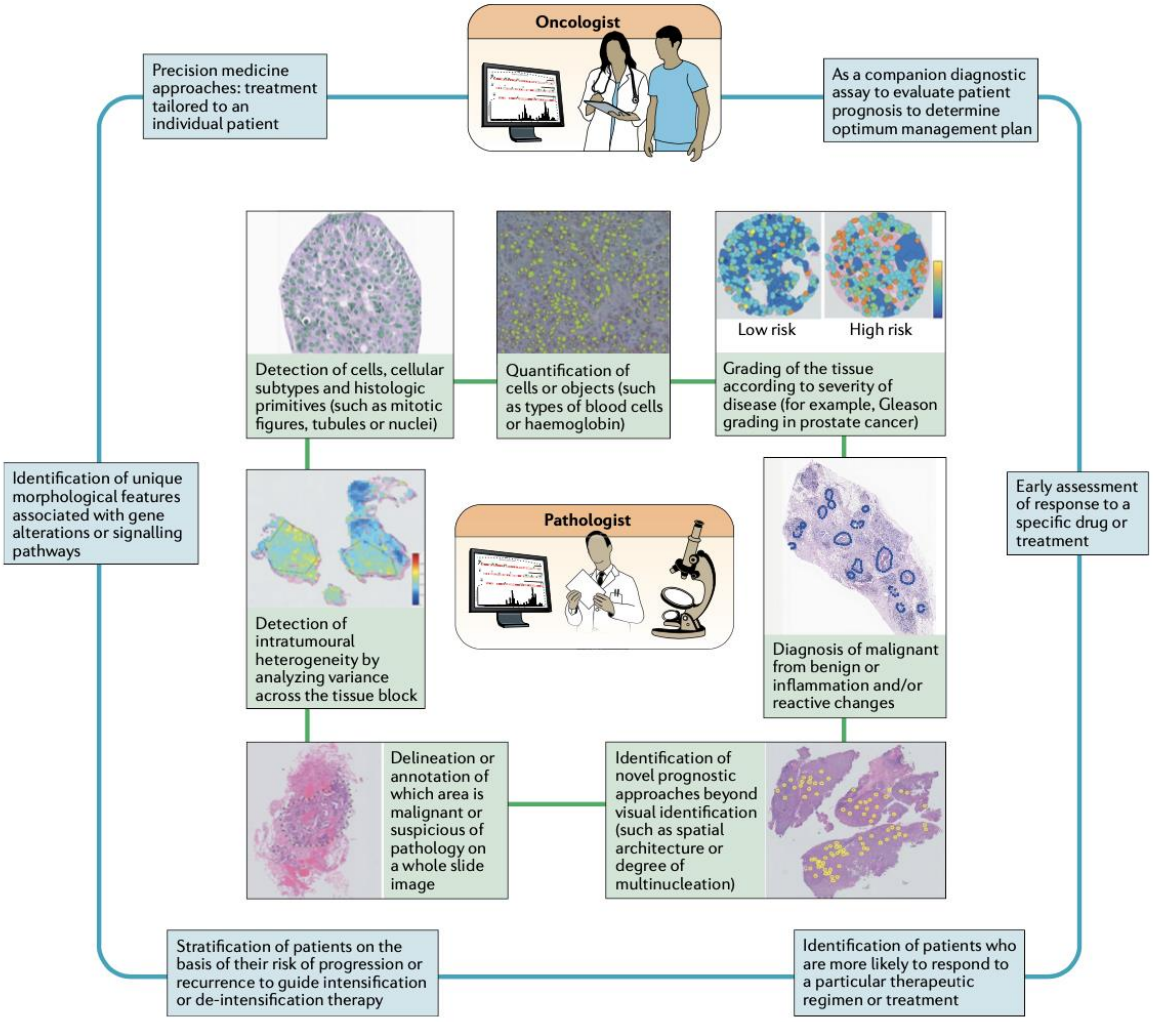


Figure 2: Some AI and ML approaches which are used by pathologists today. The figure is used with permission from the writes of "Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology" (39).

Machine/deep learning models have recently shown great promise in translational medicine for predicting the tissue of origin for unknown primary tumors (39), identifying mutational subtypes (40), and for tumor grading (41). These models can learn from vast amounts of data and can identify patterns and features that may not be visible to the human eye. As a result, these models can help pathologists make more accurate and informed diagnoses (39). One exciting application of machine/deep learning models in computational pathology is the development of algorithms for quantification of immune cell subsets on hematoxylin and eosin (H&E) stained sections in multiple malignancies (42). These models can accurately identify and quantify different immune cell types, which can have a significant impact on the effectiveness of immunotherapy. The predictive impact of these immunological markers can

be tested in relation to immunotherapy clinical outcomes, helping clinicians make more informed treatment decisions (43-45).

Furthermore, there are opportunities to integrate these models for the purpose of estimating tumor purity as well. AI-based quantification models have the potential to substitute classical microscopic estimations of tumor purity with improved reliability and reduced subjectivity. By accurately estimating tumor purity, clinicians can make better decisions about which patients are suitable for molecular genomic profiling and personalized treatment plans.

Cell/object-based vs pixel-based ML classifier

The segmentation of tumor tissues and cells is an essential step in cancer diagnosis and treatment. Accurate segmentation can aid in the identification of cancerous cells, determine the extent of tumor invasion, and assess the effectiveness of treatment. In recent years, ML techniques have been employed for tumor segmentation (39). Two popular approaches to tumor segmentation using ML are cell/object-based and pixel-based classification (46).

Cell/Object-Based Classification:

Cell/object-based classification methods rely on the identification of individual cells or objects in the tissue image. These methods segment the image by detecting individual cells or objects, such as nuclei or cytoplasm, and then classifying them based on their features. Features used in cell/object-based classification can include shape, size, texture, and intensity. This approach has been shown to achieve high accuracy in tumor segmentation (39, 46).

Pixel-Based Classification:

Pixel-based classification methods, on the other hand, classify individual pixels based on their features. These methods segment the image by dividing it into smaller regions, or pixels, and then classifying each pixel as tumor or non-tumor based on its features. Features used in pixel-based classification can include color, texture, and intensity. This approach has the advantage of being computationally efficient and can be used with a variety of imaging modalities (46).

While both cell/object-based and pixel-based classification methods have their advantages and disadvantages, studies have shown that cell/object-based classification methods tend to be

more accurate in tumor segmentation (39). Additionally, cell/object-based classification has the potential to provide more detailed information on individual cells, which can aid in the identification of different subtypes of cancer cells (46).

Overall, both cell/object-based and pixel-based classification methods have their strengths and weaknesses in tumor tissue and cell segmentation. While pixel-based classification methods have the advantage of being computationally efficient and can be used with a variety of imaging modalities, cell/object-based classification methods tend to be more accurate and can provide more detailed information on individual cells (47). The choice of which approach to use may depend on the specific needs of the study and the resources available.

The aim of this thesis

The ultimate objective of this master's thesis is to improve genomic profiling and enhance the stratification of patients for different treatment strategies by:

1) developing a ML-based cell-based classifier for estimating tumor purity in lung cancer tissues. This will involve training the ML algorithm to accurately identify neoplastic cells and assess their percentage in the tissue sample.

2) in addition, the thesis will compare manual neoplastic cellularity reads with those obtained through ML-based classification. This comparison aims to assess the accuracy and reproducibility of the ML-based method and its potential to replace or complement the traditional manual method.

Overall, the thesis aims to contribute to the advancement of precision medicine in lung cancer by providing a reliable and objective method for tumor purity estimation and neoplastic cellularity reads. The results of this study may have implications for personalized treatment decision-making and may lead to improved outcomes for lung cancer patients.

Methods

Material

In our study, which is a part of the subsequent clinical trial (tumor-node-metastasis-immunoscore (TNM-I)), we included 61 patients diagnosed with NSCLC at pathological stages I-IIIA at the University Hospital of North Norway (UNN), Tromsø, between 2017 and 2018 (ClinicalTrials.gov ID: NCT03299478). Tumor samples were collected from the patients and de-identified patient demographic and clinicopathological data were compiled and stored in a secure database using the Research Electronic Data Capture (REDCap) tool.

Manual tumor cell estimation

Tumor cellularity refers to the proportion or percentage of neoplastic cells present in the tumor tissue sample and is typically estimated through manual microscopic evaluation by a trained pathologist. In the TNM-I trial, tumor content was manually estimated by Pathologist 1 (Lill-Tove Busund) prior to next-generation sequencing for the patients included in genomic profiling (48). For the purpose of this study and to assess inter-observer variability, the tumor content of the tissue was then reviewed by a second observer, Pathologist 2 (Elin Richardsen).

Digital tumor cell quantification

Digital slide:

Using a Panoramic 250 Flash III scanner (3DHistech, Hungary), the surgically resected slides were digitized at a resolution of 0.24 microns per pixel. To automate tumor cell assessment, supervised ML algorithms were employed sequentially using the open-access program QuPath v.0.3.0 from Queen's University in Northern Ireland. A single slide was used for each patient in the analysis.

Color normalization and cell detection:

The staining vectors were estimated and RGB channels were normalized per slide using color deconvolution due to the variation in H&E intensity on different slides. Cells were identified based on size, shape, and optical density (OD) of nuclei in the hematoxylin layer using watershed segmentation, with 33 features calculated for each cell. Parameters were set for hematoxylin OD, including pixel size of 0.25 μm, background radius of 10 μm, median filter radius of 1 μm, sigma of 1.5 μm, minimum area of 7 μm2, maximum area of 500 μm2,

intensity threshold of 0.1, background intensity of 2, cell expansion of 2 µm, watershed post process with nuclei inclusion and smooth boundaries, and measurements taken for each cell. Additional intensity and smoothed object features were added, and Haralick texture features were calculated with a Haralick distance of 1 and 32 Haralick bins, along with gaussian-weighted averages per cell.

Training a cell-based classifier:

The study developed random decision forest cell classifiers, guided by pathologist annotations, to identify tumor and stroma cells. The cohort was randomly divided into training and test data sets, and 0.25 mm^2 ROIs were randomly harvested to build training images. The cell labeling process improved as more cells were annotated and curated until the classifiers achieved performance comparable to that of a pathologist. Quality control for cell detection and classification was carried out by pathologists on training images and randomly on the full cohort. Finally, the locked classifier was deployed on the remaining cases (test set). All the steps are shown in **Figure 3**. The digital tumor purity estimated by following equation: (number of tumor cells/ number of detected cells) x 100 %).

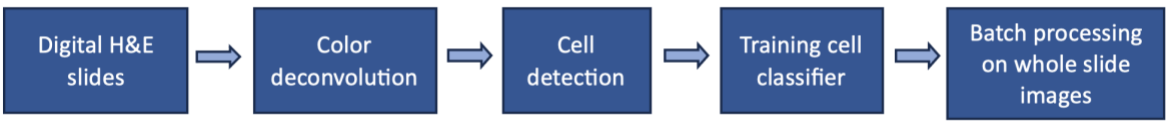


Figure 3: Flowchart showing the processes used in creating the ML-based classifier.

Statistical analysis

Statistics and data visualization were performed using SPSS version 29.0.0.0., with a p-value <0.05 considered statistically significant. Inter-observer agreement between the two pathologists was evaluated using weighted Cohen's kappa and Intraclass Correlation Coefficient (ICC) (two-way random-effects model with absolute agreement definition). The performance of automated digital scores was compared to pathologist manual scores also using weighted Cohen's kappa and ICC.

Ethic clearance

The Regional Committee (REK) and the Norwegian Data Protection Organization granted ethical approval for the entire clinical trial (NCT03299478) with IRB # REK2016/2

Results

Patients

As shown in **Table 1** the patients selected for the study are 51% male and 49% female. The ages range between 51 and 83 years old with a median of 71 years old. Most of them were former smokers with a percentage of 56%, 36% were current smokers and only 8% were never smokers. When it comes to types of lung cancer included in the study, most of the cases were lung adenocarcinoma (LUAD) as it accounted for 66%, while 30% of the histology included were lung squamous cell carcinoma (LUSC) and only 5% were other types of lung cancer. 59% of the cases included were stage I, 30% were stage II and only 11% were stage IIIA.

Clinical characteristics	**N (%)**
Gender	
Male Female	31 (51) 30 (49)
Age (years)	
<71 ≥ 71	30 (49) 31 (51)
Histology	
LUAD LUSC Other	40 (66) 18 (30) 3 (5)
Smoking	
Former Current Never	34 (56) 22 (36) 5 (8)
pStage	
I II IIIA	36 (59) 18 (30) 7 (11)
Who performance	
0 1 2	46 (75) 13 (21) 2 (3)
Surgery procedure	
Lobectomy Pulmectomy	60 (98) 1 (2)

Table 1: Clinical characteristics of the patients included in the study, shown in numbers and percentage.

Manual tumor purity estimation

The manual tumor purity estimation done by Pathologist 1 and Pathologist 2 is shown graphically in **Figure 4-6.** The bar plots (Figure 5 and 6) display the score distribution for each observer. Both observers presented a mean score of 60. However, there was minimal variability in the data, as demonstrated by a standard deviation of 22 for Pathologist 1 and 23 for Pathologist 2, respectively.

We aimed to compare the consistency of tumor purity estimations made manually by Pathologist 1 and Pathologist 2. To achieve this, we utilized the ICC, a statistical measure typically used to establish the consistency of repeated measurements. Our objective was to assess the degree of agreement between the two pathologists - essentially, to determine how consistent their tumor purity estimations were in comparison with each other. The ICC calculated was 0.833, with a 95% confidence interval of 0.736-0.896. Considering that an ICC coefficient between 0.75 and 0.90 indicates good reliability, the calculated value of 0.833 suggests that Pathologist 1 and Pathologist 2 show good consistency in their estimations of tumor purity (49). The results were also statistically significant with a P-value of < 0.001.

Cohen's Kappa is a statistical measure of inter-rater reliability, employed to evaluate how reliably two raters assess the same parameter. In our study, we utilized Cohen’s Kappa to examine the level of agreement between Pathologist 1 and Pathologist 2 in their tumor purity estimations. Our analysis yielded a Cohen’s Kappa value of 0.534 (P-value < 0.001). Given that a Cohen’s Kappa value ranging between 0.4 and 0.6 represents moderate inter-rater reliability, our results suggest a moderate level of agreement between the two pathologists in their estimations of tumor purity (50).

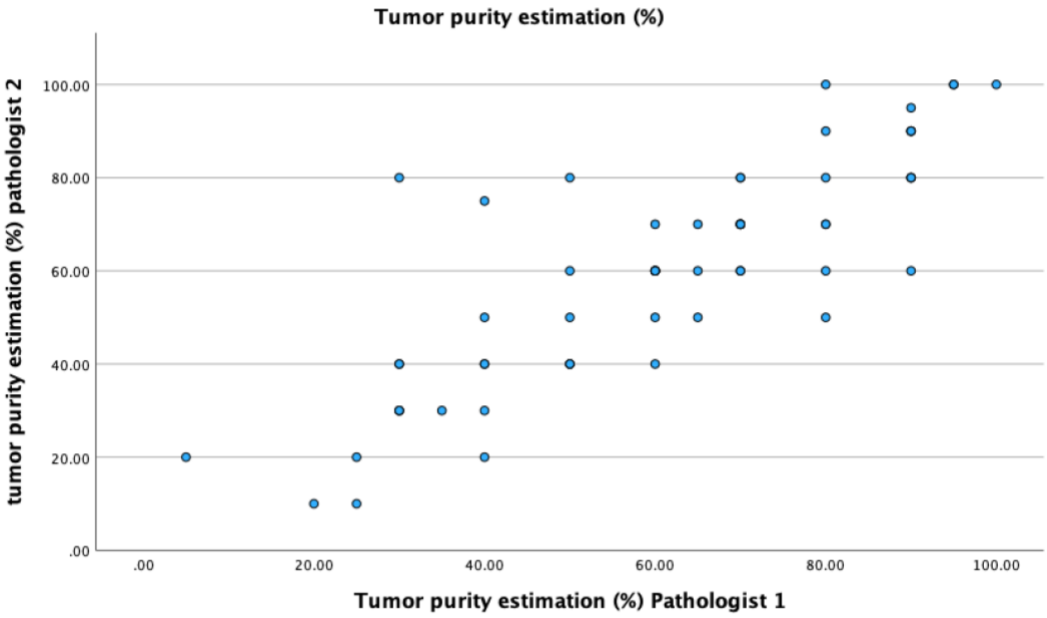


Figure 4: Scatter plot showing the tumor purity estimation done by Pathologist 1 and Pathologist 2.

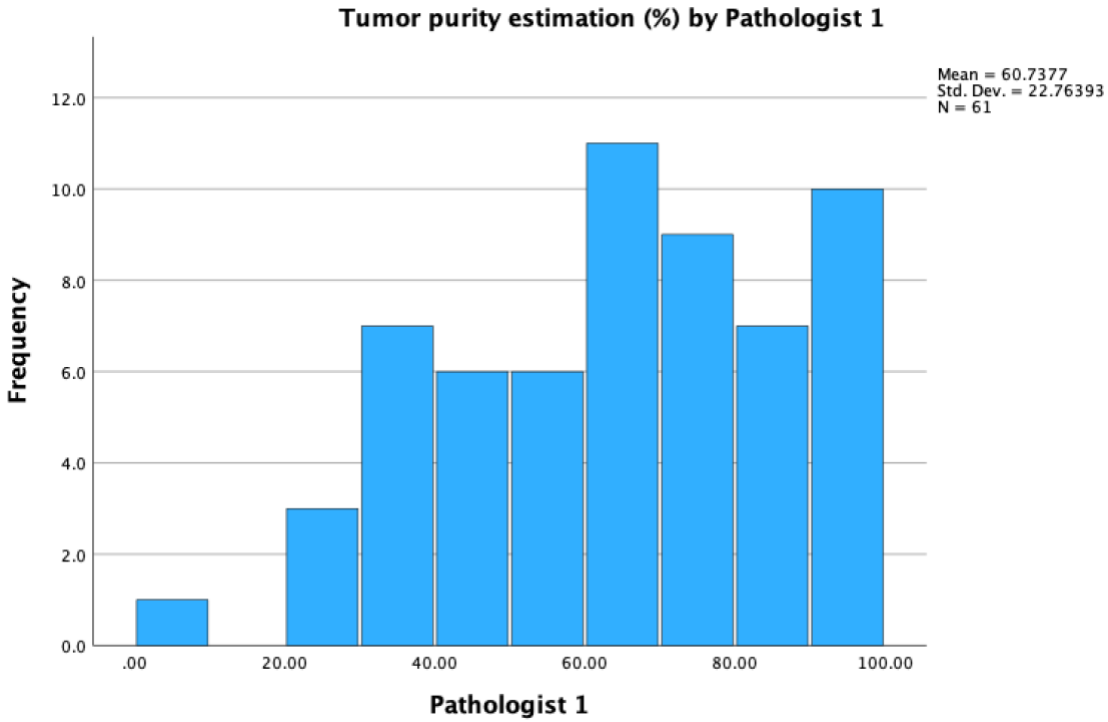


Figure 5: Bar plot showing the tumor purity estimations done by Pathologist 1.

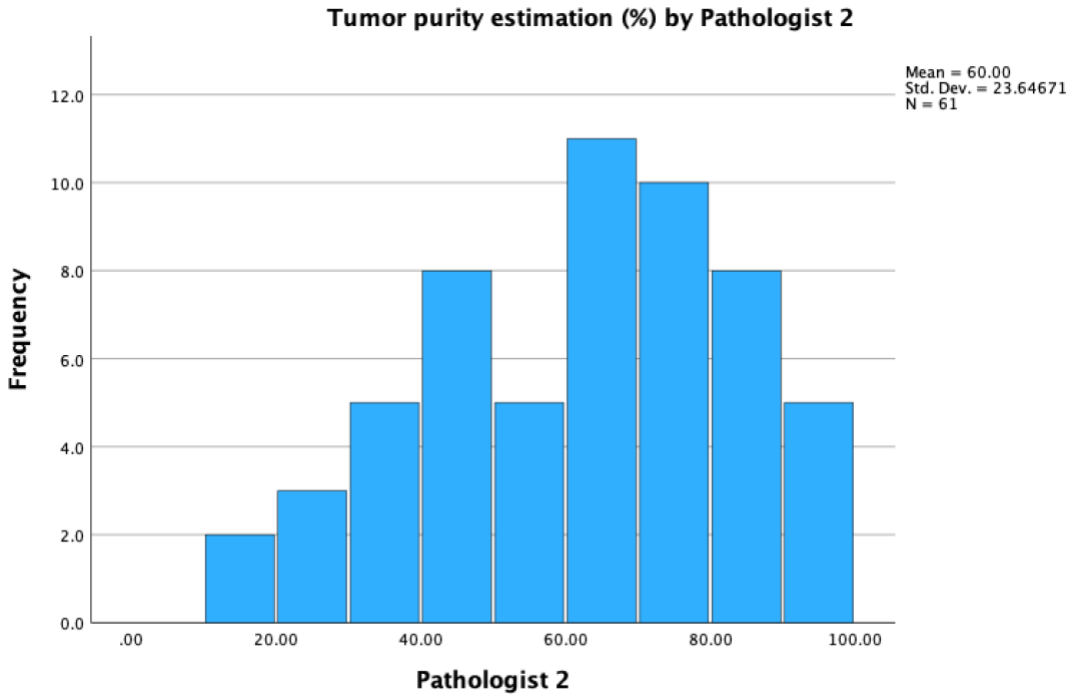


Figure 6: Bar plot showing the tumor purity estimations done by Pathologist 2.

Developing ML classifier

During the construction of our ML classifier, we prioritized determining the optimal cell detection settings in QuPath. We compared two settings; one suggested by QuPath and another used in a previous report (44). The QuPath settings included a pixel size of 0.5 µm, and the prior classifier settings featured a similar pixel size but with a background radius of 10 µm and a cell expansion of 2 µm. Upon testing, we found that the prior classifier settings consistently detected more cells, as shown in **supplementary Figures S1-S3**. Moreover, these settings didn't falsely identify non-cellular objects as cells, which motivated us to adopt these settings for this project.

We also modified the pixel size from 0.5 µm to 0.25 µm, which further enhanced cell detection. After finalizing these settings based on the accurate cell detection rate, a pathologist validated our choice. **Figure 7** showcases the cell classification model, and the performance of this cell classifier was verified by the pathologist.

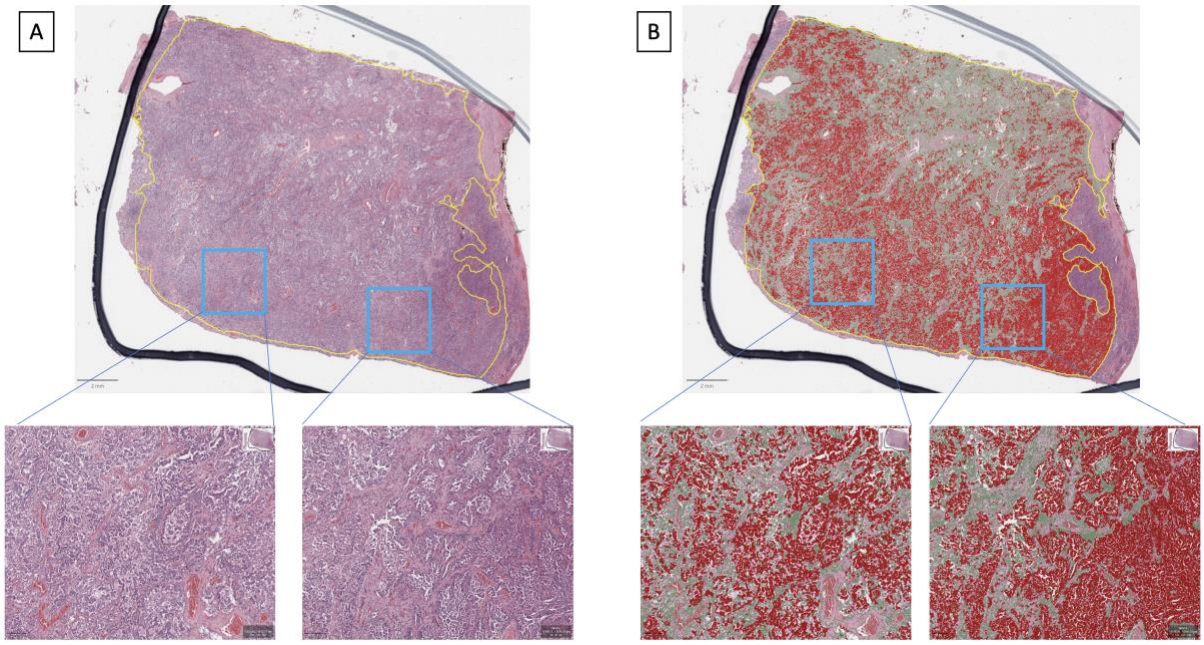


Figure 7: Random Forest Cell Classifier. A) The original H&E whole slide image, and B) the H&E image overlaid with the ML classifier. The tumor regions are indicated by the color red, while the stroma is represented by green.

Digital tumor purity estimation

The tumor purity estimation done by the digital and the manual approach is illustrated in **Figures 8-10**. The differences in the tumor purity estimation done by the ML-classifier and the pathologists are shown when comparing the two bar plots (Figure 9 and 10), which shows some differences in terms of standard deviation of the mean (22 vs 23 for digital vs mean manual scores).

We wanted to determent how consistent the ML tumor purity estimate is compared to the mean tumor purity estimate done by the two pathologists and used ICC for that purpose. The computed ICC coefficient was 0.838, which was statistically significant with a P-value of <0.001, and the 95% confidence interval was between 0.743 and 0.899. These results indicate good reliability, as ICC values falling within the range of 0.75 and 0.9 typically denote this level of reliability (49). Therefore, these findings suggest that the ML model and the pathologists exhibit a good degree of consistency with each other in their estimations of tumor purity.

Then we aimed to assess the inter-rater reliability between the manual tumor purity estimation made by the pathologists and the digital estimation made by the ML model. Our analysis yielded a Cohen's Kappa value of 0.563, which signifies moderate inter-rater reliability (50). The findings were statistically significant with a P-value of < 0.001. Based on our results, it appears that there is moderate agreement between the manual method employed by the pathologists and the digital method implemented through the ML model in estimating tumor purity.

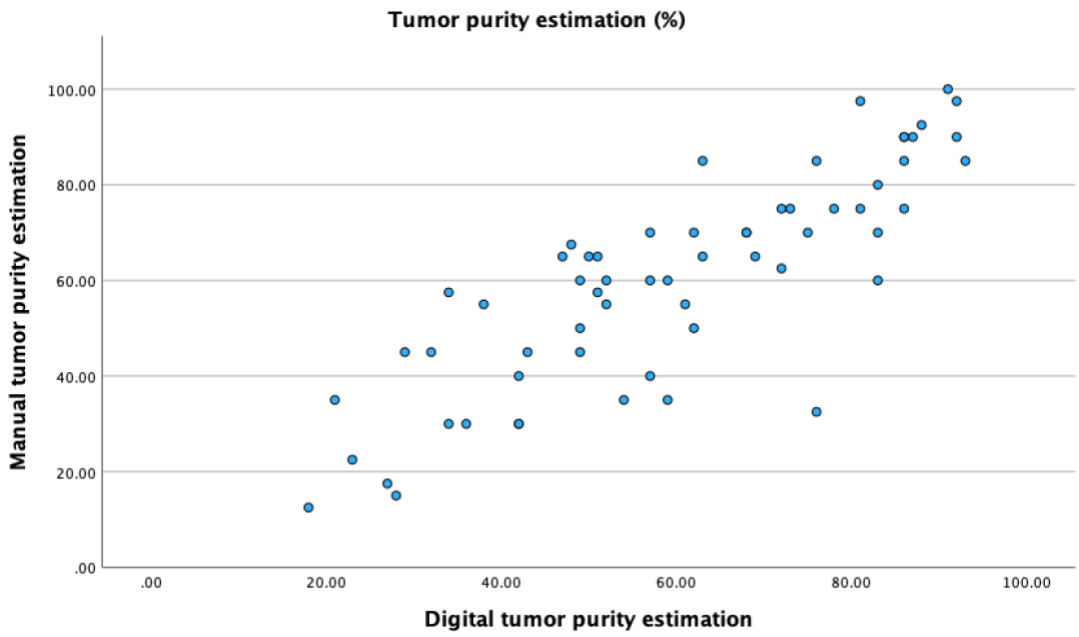


Figure 8: Scatter plot showing the tumor purity estimation done digitally and manually. The manual tumor purity estimation is represented by the mean tumor purity estimation done by Pathologist 1 and Pathologist 2.

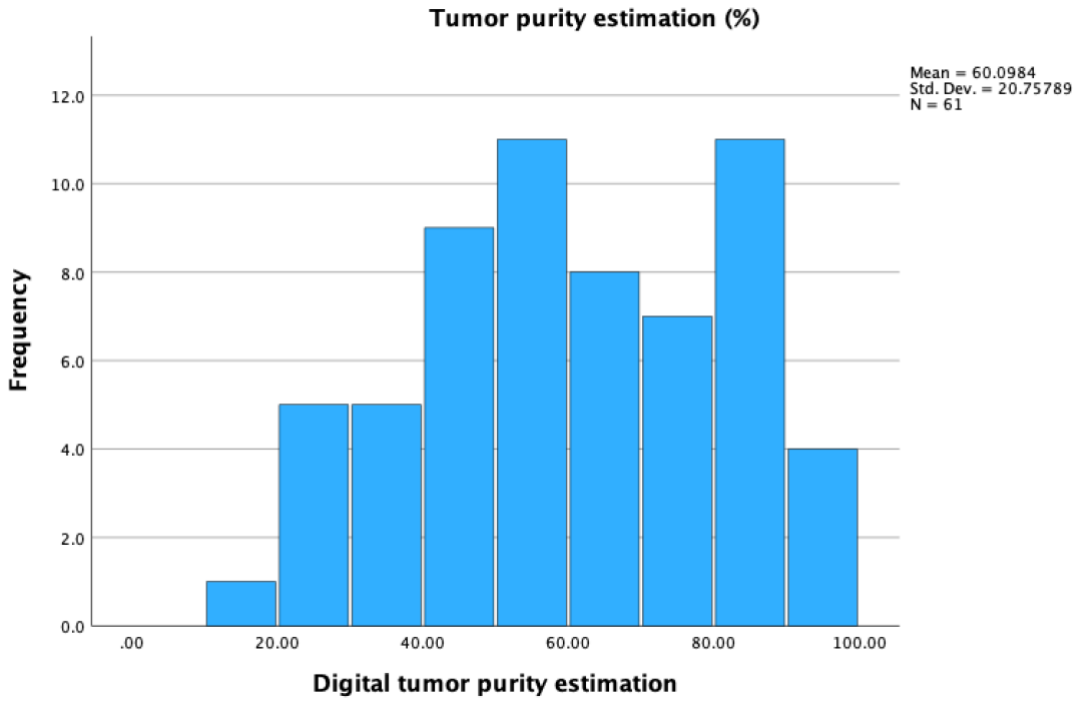


Figure 9: Bar plot showing the tumor purity estimated by the ML classifier.

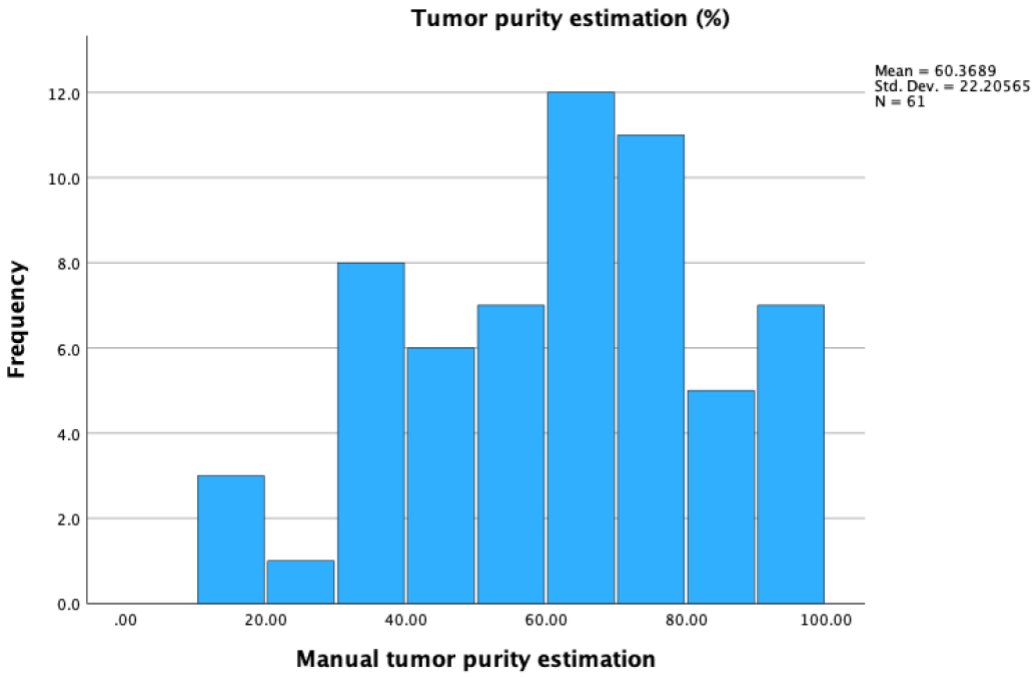


Figure 10: Bar plot showing the manual tumor purity estimation, represented by the mean tumor purity estimation done by Pathologist 1 and Pathologist 2.

Discussion

In this sub-study conducted as part of the TNM-I trial, we have developed a ML classifier to estimate tumor purity from histological slides. Utilizing the ICC to measure the consistency of our digital tumor purity estimation against manual methods, we found that our ML model demonstrated good reliability. This suggests a favorable level of consistency between the ML model and the pathologists in estimating tumor purity. Furthermore, we analyzed the inter-rater reliability using Cohen's Kappa. The result value indicated a moderate level of agreement between the digital and manual methods of estimating tumor purity. Despite the relative novelty of digital tumor purity estimation, the moderate agreement in this early stage underscores the potential of our ML classifier as a reliable tool in tumor purity estimation.

As mentioned earlier in the thesis, a problem with manual tumor purity estimation high inter-observer variability between pathologists' estimates. A study concluded that manual estimates of tumor purity on H&E-stained slides are not accurate due to significant difference between the nine pathologists who were making these estimations (51). Similarly, another study involving nine pathologists evaluating 18 tumors revealed that pathologists generally tend to overestimate tumor purity (52). These findings on inter-observer variability are supported by the Cohen's Kappa value calculated in our study, as it indicates only moderate agreement between the two pathologists included in our study when it comes to estimating tumor purity.

In addition, studies have found high inter-observer variability between pathologists when evaluating other histopathological features. An article by Kos Et al. looked at 3 studies who evaluated pathologists scoring of stromal tumor-infiltrating lymphocytes (sTILs), which are an important prognostic factor in triple-negative and human epidermal growth factor receptor 2 (HER2)-positive breast cancer (44). In total, results from 220 slides were evaluated and the average ICC values were 0.7 for study 1, 0.89 for study 2 and 0.76 for study 3. The ICC values indicate moderate to good reliability for the sTILs scoring done by the 66 pathologists in the 3 studies included (44). A large-scale international multicenter study involving 39 pathologists who evaluated 149 ductal carcinomas in situ slides concluded with inter-observer variability being considerable, but variable between the different histopathological features evaluated and being acceptable at most (53). Notably, lobular cancerization, nuclear atypia, and stromal architecture exhibited the highest inter-observer variability, with Krippendorff's Alpha (KA) values of 0.39, 0.42, and 0.45, respectively. Conversely, solid ductal carcinoma

in situ architecture and calcifications displayed the lowest inter-observer variability, with KA values of 0.60 and 0.67, respectively (53). These findings underline the potential of ML as a viable approach for evaluating histopathological features beyond just tumor purity and to tackle these substantial diagnostic challenges created by inter-observer variability between pathologists.

Manual estimation of tumor purity is tedious, time consuming, and prone to inter-observer variability (38). These problems can be solved by using ML to estimate tumor purity. By developing classifiers like in this study, one can secure a standardized method for estimating tumor purity. This can ensure a more reproducible estimation of tumor purity, by avoiding the possibility for human errors and subjectivity. In addition, the ML approach can identify features that may not be visible to humans, which means we may get a more accurate estimation of tumor purity than the traditional, manual way (40). At the same time using ML to estimate tumor purity can reduce pathologists' workload (54).

As mentioned, we created the ML classifier using QuPath. This is an advantage since QuPath is an open source bioimage analysis platform, which means it is available for anyone to use and therefore also cost-effective. It is a both comprehensive and user-friendly program for analyzing whole slide images (55). The program was designed to be user-friendly to such a degree that users without computer programming skills can use it easily, at the same time QuPath allows scientists with software development skills to add their own extensions (56). QuPath is optimal for the purpose of creating a ML classifier, since the software primarily was designed to count cells, classify objects and pixels in large whole slide brightfield, fluorescent or H&E images without the need for cropping or down sampling to lower file sizes (56). The pros of using QuPath as an open-source program lie in its ability to foster collaboration and adoption within the scientific community. Its open nature allows the tumor purity of other developed algorithms to be widely shared and implemented by Clinical Laboratory Improve Amendments (CLIA) laboratories without significant financial burdens. On the other hand, there are cons associated with QuPath’s open-source nature and lack of FDA approval. The absence of regulatory agency approval may raise doubts about its reliability for certain users. Additionally, the use of ML in QuPath may introduce concerns and discomfort due to its novel and evolving nature. Furthermore, CLIA laboratories need to possess digital infrastructure and capabilities to effectively utilize QuPath for assay validation and analysis (57).

For our classifier, we opted for a method based on ML, a branch of artificial intelligence (AI). Another approach that could have been employed is deep learning (DL), a subset of ML, which is being used in other studies to build similar classifiers. DL is a subset of ML and refers to a specific class of ML algorithms called neurol networks. It can be said that neural networks try to mimic the learning process of the human brain by layering algorithms and computing units into artificial neural networks. ML uses experience to learn and adapt automatically without being explicitly programmed. This means that ML requires more human intervention to correct and learn from past mistakes than DL which learns and corrects on its own (58, 59). The ML model requires a scientist which can feed large amounts of data into the system and thereby "train" the machine, the more date the machine analyses, the better it gets in performing the task. In contrast, DL models can improve their outcomes through repletion and without human intervention. ML models require shorter training than DL models, but in contrast it delivers lower accuracy. Another important difference is that ML models can be trained on smaller data sets, while DL models requires large amounts of data (58, 59).

According to ICC, our ML model estimates tumor purity relatively consistent to the two pathologists which is the traditionally used method to estimate tumor purity. A study which developed a deep multiple instance learning model predicting tumor purity from H&E-stained digital histopathology slides concluded that their model predicted tumor purity highly consistent with tumor purity estimates using genomic tools (38). Another study used various ML models and concluded that the models predicted tumor purity accurately and showed a high correlation with well-established gold standard methods for estimating tumor purity (60). A study using whole slide digital scanned images of colorectal, lung and breast cancer specimens showed that tumor purity estimation done by digital image analysis were highly concordant with manual estimation in 86% of the samples (61). Consistently, a study on colon cancer showed that their developed computer algorithm had a slight median deviation from the criterion standard, which was set by pathologists manually estimating tumor purity, of 5.4% on the training set and 6.2% on the validation set (62). Furthermore, another study conducted on H&E-stained lung cancer scans revealed that digital tumor purity estimations significantly correlated with benchmark tumor cell counts created by multiple pathologists (63). A study employing transfer learning to estimate tumor purity in H&E-stained breast cancer images found a strong correlation, with a ICC value of 0.94, compared to estimations made by pathologists (64). In addition, a study focusing on breast cancer using DL

approaches found a Cohen's Kappa coefficient of 0.69 and an ICC coefficient of 0.89 when comparing their estimations to pathologist scoring of tumor purity (65). In summary, all these studies on tumor purity using ML/DL methods support our finding that ML methods correlate well with traditional, well-established methods for detecting tumor purity.

When it comes to the limitations of our study, some points should be mentioned. One point is that we should have included more slides in the study. As mentioned above a ML model requires a large amount of data to be "trained" ideally, since the principle is that the more data the machine analyses, the better it gets in performing the task. Based on this statement one can argue that 61 whole slide images are not sufficient in training a ML model ideally. Another fact worth mentioning is that our study lacks a validation dataset. A validation dataset is used to test the ML models performance. It is a sample of data held back from training the ML model, and it is different from the test dataset because it is used to give an unbiased estimate of the skill of the final tuned model (66). It can also be argued that we should have included more pathologists than two in our study to get more manual tumor purity estimations. This would have contributed to a better inter-observer variability estimation and a better mean value to compare the digital tumor purity estimation to.

Our study compared the ML derived tumor purity with manually estimated tumor purity done by pathologist. A different approach is to compare digital tumor purity estimation to genomic tumor purity, as Oner et al. did in the pan-cancer study on ML and tumor purity estimation (38). Genomic tumor purity is tumor purity inferred from different types of genomic data, such as somatic copy number and mutations, gene expression data and DNA methylation data. Genomic purity values are used in correlational studies to investigate the associations between tumor purity and clinical variables and in genomics analysis to mitigate confounding effects of normal cell contamination, but genomic tumor purity have also now been accepted as an accurate value for downstream analysis. Another fact about genomic tumor purity is that it does not correlate well with pathologists' tumor purity estimations (38). An idea for future studies on digital tumor purity estimations is to compare the values to genomic tumor purity instead of pathologists' tumor purity estimations and making a ML/DL model such that the predictions are consistent with the genomic tumor purity values.

Conclusion

In this master thesis we have successfully developed a ML based classifier for estimating tumor purity in lung cancer tissue. When comparing manual tumor purity estimation to those obtained through the ML based classification they correlate well in terms of ICC and Cohen's Kappa, which proposes accuracy and reproducibility of the model. All the results were statistically significant and indicates good agreement between manually and digitally tumor purity estimations. The results in this study supports findings in previously done studies on the subject, and suggests that further research should be done in terms of developing a ML based method which can replace or complement the traditionally manual method of detecting tumor purity.

Bibliography

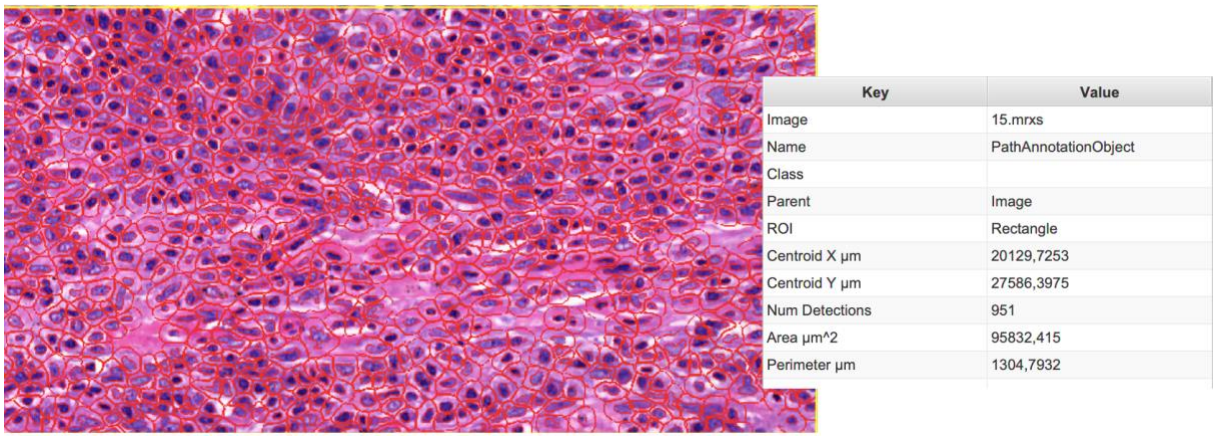
1. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: An overview. Int J Cancer. 2021.
2. Helsedirektoratet. Lungekreft, mesoteliom og thymom - handlingsprogram helsedirektoratet.no: Helsedirektoratet; 2021 [updated 23.12.2021; cited 2021 21.12]. Available from: https://www.helsedirektoratet.no/retningslinjer/lungekreft-mesoteliom-og-thymom-handlingsprogram/epidemiologi/forekomst-og-dodelighet.
3. Kreftregisteret. Kreft i Norge kreftregisteret.no2022 [updated 08.06.2022; cited 2022 21.12]. Available from: https://www.kreftregisteret.no/Temasider/om-kreft/.
4. Kreftregisteret. Lungekreft kreftregisteret.no2022 [updated 02.11.2022; cited 2022 21.12]. Available from: https://www.kreftregisteret.no/Temasider/kreftformer/Lungekreft/.
5. Brustugun OT, Møller B, Helland A. Years of life lost as a measure of cancer burden on a national level. Br J Cancer. 2014;111(5):1014-20.
6. Malhotra J, Malvezzi M, Negri E, La Vecchia C, Boffetta P. Risk factors for lung cancer worldwide. European Respiratory Journal. 2016;48(3):889-902.
7. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. Bmj. 2004;328(7455):1519.
8. Smoke T, Smoking I. IARC monographs on the evaluation of carcinogenic risks to humans. IARC, Lyon. 2004;1:1-1452.
9. Hackshaw AK, Law MR, Wald NJ. The accumulated evidence on lung cancer and environmental tobacco smoke. Bmj. 1997;315(7114):980-8.
10. Boffetta P. Involuntary smoking and lung cancer. Scand J Work Environ Health. 2002;28 Suppl 2:30-40.
11. Lam TK, Gallicchio L, Lindsley K, Shiels M, Hammond E, Tao XG, et al. Cruciferous vegetable consumption and lung cancer risk: a systematic review. Cancer Epidemiol Biomarkers Prev. 2009;18(1):184-95.
12. Marmot M, Atinmo T, Byers T, Chen J, Hirohata T, Jackson A, et al. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. 2007.
13. Sinha R, Kulldorff M, Swanson CA, Curtin J, Brownson RC, Alavanja MC. Dietary heterocyclic amines and the risk of lung cancer among Missouri women. Cancer Res. 2000;60(14):3753-6.
14. Mayne ST, Buenconsejo J, Janerich DT. Previous lung disease and risk of lung cancer among men and women nonsmokers. Am J Epidemiol. 1999;149(1):13-20.
15. Wu AH, Fontham ET, Reynolds P, Greenberg RS, Buffler P, Liff J, et al. Previous lung disease and risk of lung cancer among lifetime nonsmoking women in the United States. Am J Epidemiol. 1995;141(11):1023-32.
16. Gao YT, Blot WJ, Zheng W, Ersnow AG, Hsu CW, Levin LI, et al. Lung cancer among Chinese women. International journal of cancer. 1987;40(5):604-9.
17. Aoki K. Excess incidence of lung cancer among pulmonary tuberculosis patients. Jpn J Clin Oncol. 1993;23(4):205-20.
18. Lissowska J, Bardin-Mikolajczak A, Fletcher T, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, et al. Lung cancer and indoor pollution from heating and cooking with solid fuels: the IARC international multicentre case-control study in Eastern/Central Europe and the United Kingdom. Am J Epidemiol. 2005;162(4):326-33.
19. Afaj A. IARC Monograph on the Evaluation of Carcinogenic Risk to Humans, Volume 95: Household Use of Solid Fuels and High‐temperature Frying. Taylor & Francis; 2011.
20. Ionizing radiation, part 1: X- and gamma-radiation, and neutrons. Overall introduction. IARC Monogr Eval Carcinog Risks Hum. 2000;75 Pt 1(Pt 1):35-115.

21. Organization WH. Weight control and physical activity. Weight control and physical activity. 2002.
22. Hassfjell CS, Grimsrud TK, Standring WJ, Tretli S. Lungekreftforekomst knyttet til radoneksponering i norske boliger. Tidsskrift for Den norske legeforening. 2017.
23. Helsedirektoratet. Inngang til pakkeforløp for lungekreft helsedirektoratet.no2016 [updated 01.08.2016; cited 2023 08.01]. Available from: https://www.helsedirektoratet.no/nasjonale-forlop/lungekreft/inngang-til-pakkeforlop-for-lungekreft.
24. Helsedirektoratet. Utredning av lungekreft helsedirektoratet.no2022 [updated 29.04.2022; cited 2023 08.01]. Available from: https://www.helsedirektoratet.no/nasjonale-forlop/lungekreft/utredning-av-lungekreft#utredning.
25. Baugstø VS. Derfor er MDT-møtet så viktig dagensmedisin.no2017 [updated 01.11.2017; cited 2023 08.01]. Available from: https://www.dagensmedisin.no/artikler/2017/01/11/-derfor-er-mdt-motet-sa-viktig/.
26. Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. Nat Rev Dis Primers. 2021;7(1):3.
27. Helsedirektoratet. Behandling av lungekreft helsedirektoratet.no2022 [updated 29.04.2022; cited 2023 08.01]. Available from: https://www.helsedirektoratet.no/nasjonale-forlop/lungekreft/behandling-av-lungekreft#hovedgrupper-av-behandlingsforlop.
28. Mamdani H, Matosevic S, Khalid AB, Durm G, Jalal SI. Immunotherapy in Lung Cancer: Current Landscape and Future Directions. Front Immunol. 2022;13:823618.
29. Vanneman M, Dranoff G. Combining immunotherapy and targeted therapies in cancer treatment. Nat Rev Cancer. 2012;12(4):237-51.
30. Adib E, Nassar AH, Abou Alaiwi S, Groha S, Akl EW, Sholl LM, et al. Variation in targetable genomic alterations in non-small cell lung cancer by genetic ancestry, sex, smoking history, and histology. Genome Med. 2022;14(1):39.
31. Prabhu S, Prasad K, Robels-Kelly A, Lu X. AI-based carcinoma detection and classification using histopathological images: A systematic review. [Amsterdam] :2022. p. 105209.
32. Arneth B. Tumor Microenvironment. Kaunas :2019.
33. Chakravarty D, Solit DB. Clinical cancer genomic profiling. England :2021. p. 483-501.
34. Wang F, Zhang N, Wang J, Wu H, Zheng X. Tumor purity and differential methylation in cancer epigenomics. Oxford :2016. p. 408-19.
35. Deng Y, Song Z, Huang L, Guo Z, Tong B, Sun M, et al. Tumor purity as a prognosis and immunotherapy relevant feature in cervical cancer. Albany, NY :2021. p. 24768-85.
36. Mao Y, Feng Q, Zheng P, Yang L, Liu T, Xu Y, et al. Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. [Auckland, N.Z.] :2018. p. 3569-77.
37. Hiley CT, Le Quesne J, Santis G, Sharpe R, de Castro DG, Middleton G, et al. Challenges in molecular testing in non-small-cell lung cancer patients with advanced disease. London2016. p. 1002-11.
38. Oner MU, Chen J, Revkov E, James A, Heng SY, Kaya AN, et al. Obtaining spatially resolved tumor purity maps using deep multiple instance learning in a pan-cancer study. [New York] :2022. p. 100399.
39. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. London :2019. p. 703-15.
40. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. New York, N.Y. :2018. p. 1559-67.

41. Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. New York, NY :2018. p. 30.
42. 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. Ghatkopar, Mumbai :2016. p. 38.
43. Sobhani F, Robinson R, Hamidinekoo A, Roxanis I, Somaiah N, Yuan Y. Artificial intelligence and digital pathology: Opportunities and implications for immuno-oncology. [Amsterdam] :2021. p. 188520.
44. Rakaee M, Adib E, Ricciuti B, Sholl LM, Shi W, Alessi JV, et al. Association of Machine Learning-Based Assessment of Tumor-Infiltrating Lymphocytes on Standard Histologic Images With Outcomes of Immunotherapy in Patients With NSCLC. JAMA Oncol. 2023;9(1):51-60.
45. Rakaee M, Adib E, Ricciuti B, Sholl LM, Shi W, Alessi JVM, et al. Artificial intelligence in digital pathology approach identifies the predictive impact of tertiary lymphoid structures with immune-checkpoints therapy in NSCLC. Journal of Clinical Oncology. 2022;40(16_suppl):9065-.
46. Parwani AV. Whole Slide Image Analysis. Cham : . 203-21 p.
47. Väyrynen JP, Lau MC, Haruki K, Väyrynen SA, Dias Costa A, Borowsky J, et al. Prognostic Significance of Immune Cell Populations Identified by Machine Learning in Colorectal Cancer Using Routine Hematoxylin and Eosin-Stained Sections. Clin Cancer Res. 2020;26(16):4326-38.
48. Rakaee M, Andersen S, Giannikou K, Paulsen EE, Kilvaer TK, Busund LR, et al. Machine learning-based immune phenotypes correlate with STK11/KEAP1 co-mutations and prognosis in resectable NSCLC: a sub-study of the TNM-I trial. Ann Oncol. 2023.
49. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.
50. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.
51. Smits AJJ, Kummer JA, de Bruin PC, Bol M, van den Tweel JG, Seldenrijk KA, et al. The estimation of tumor cell percentage for molecular testing by pathologists is not accurate. New York, NY :2014. p. 168-74.
52. Mikubo M, Seto K, Kitamura A, Nakaguro M, Hattori Y, Maeda N, et al. Calculating the Tumor Nuclei Content for Comprehensive Cancer Panel Testing. [Philadelphia, Pa.] :2020. p. 130-7.
53. Dano Hln, Altinay S, Arnould L, Bletard N, Colpaert C, Dedeurwaerdere F, et al. Interobserver variability in upfront dichotomous histopathological assessment of ductal carcinoma in situ of the breast: the DCISion study. New York, NY :2020. p. 354-66.
54. Metter DM, Colgan TJ, Leung ST, Timmons CF, Park JY. Trends in the US and Canadian Pathologist Workforces From 2007 to 2017. JAMA Netw Open. 2019;2(5):e194337.
55. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. London :2017. p. 16878.
56. Humphries MP, Maxwell P, Salto-Tellez M. QuPath: The global impact of an open source digital pathology system. Gothenburg, Sweden :2021. p. 852-9.
57. Rimm DL. An Algorithm as a Biomarker for Response to Immune Checkpoint Inhibitor Therapy. JAMA Oncology. 2023;9(1):60-1.

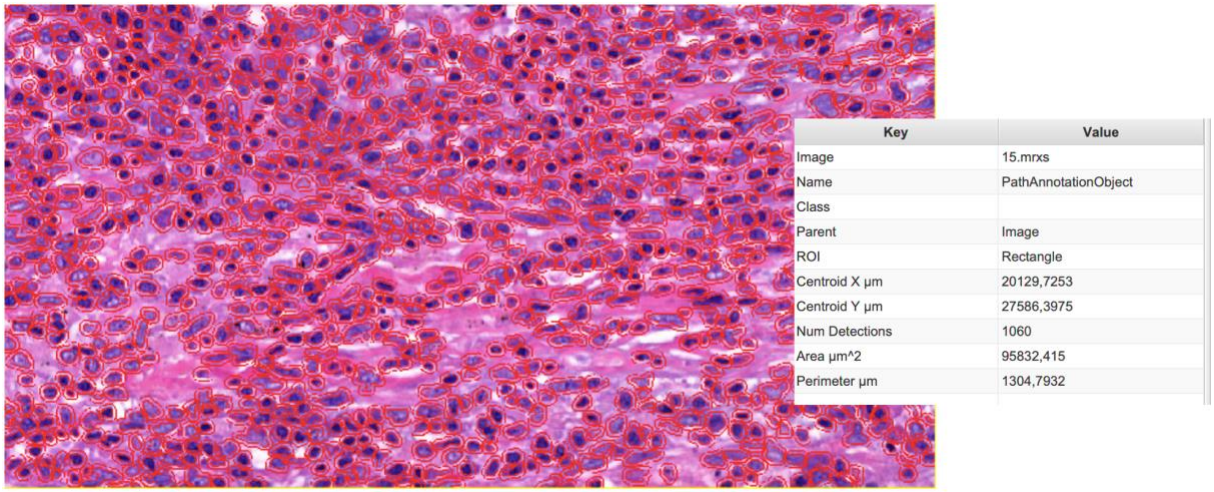
58. Coursera. Deep Learning vs. Machine Learning: Beginner's Guide www.coursera.org: Coursera; 2023 [updated 22.03.23; cited 2023 11.05]. Available from: https://www.coursera.org/articles/ai-vs-deep-learning-vs-machine-learning-beginners-guide.
59. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. [New York, N.Y.] :2019. p. 477-87.
60. Koo B, Rhee J-K. Prediction of tumor purity from gene expression data using machine learning. [London] :2021.
61. Greene C, O'Doherty E, Abdullahi Sidi F, Bingham V, Fisher NC, Humphries MP, et al. The Potential of Digital Image Analysis to Determine Tumor Cell Content in Biobanked Formalin-Fixed, Paraffin-Embedded Tissue Samples. New Rochelle, NY :2021. p. 324-31.
62. Viray H, Coulter M, Li K, Lane K, Madan A, Mitchell K, et al. Automated objective determination of percentage of malignant nuclei for mutation testing. [Philadelphia, PA] :2014. p. 363-71.
63. Hamilton PW, Wang Y, Boyd C, James JA, Loughrey MB, Hougton JP, et al. Automated tumor analysis for molecular profiling in lung cancer. Albany, N.Y. :2015. p. 27938-52.
64. Pei Z, Cao S, Lu L, Chen W. Direct Cellularity Estimation on Breast Cancer Histopathology Images Using Transfer Learning. [London] :2019. p. 3041250.
65. Rakhlin A, Tiulpin A, Shvets AA, Kalinin AA, Iglovikov VI, Nikolenko S, editors. Breast tumor cellularity assessment using deep neural networks. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops; 2019.
66. Brownlee J. What is the Difference Between Test and Validation Datasets? machinelearningmastery.com2017 [updated 14.08.2020; cited 2023 13.05]. Available from: https://machinelearningmastery.com/difference-test-validation-datasets/.

Supplementary



Key	Value
Image	15.mrxs
Name	PathAnnotationObject
Class	
Parent	Image
ROI	Rectangle
Centroid X µm	20129,7253
Centroid Y µm	27586,3975
Num Detections	951
Area µm^2	95832,415
Perimeter µm	1304,7932

Figure S1: QuPath detects 951 cells in the annotation when using the QuPath suggested settings for cell detection.



Key	Value
Image	15.mrxs
Name	PathAnnotationObject
Class	
Parent	Image
ROI	Rectangle
Centroid X µm	20129,7253
Centroid Y µm	27586,3975
Num Detections	1060
Area µm^2	95832,415
Perimeter µm	1304,7932

Figure S2: When using the settings which was used in the prior classifier QuPath detects 1060 cells in the annotation. This is an increase of 109 cells compared to using the QuPath suggested settings on cell detection in the same annotation.

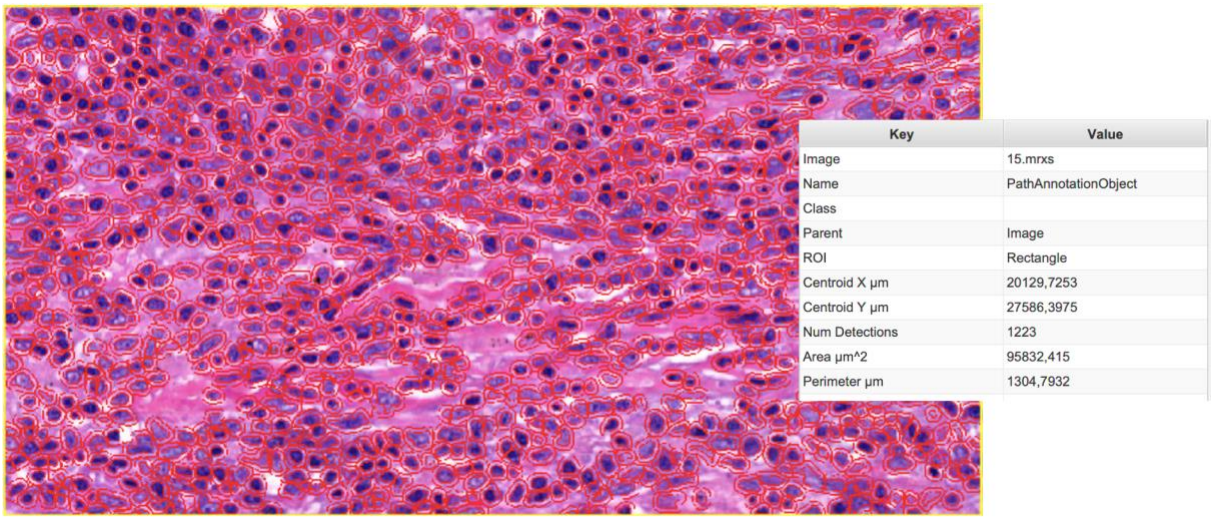


Figure S3: When changing the pixel size to 0.25 µm, in addition to using the settings which was used in the prior classifier, QuPath increases the detection to 1223 cells detected in the annotation. This is an increase of 163 cells compared to using only settings from the prior classifier and an increase of 275 cells compared to using the QuPath suggested settings.