Prostate Cancer Detection using Deep Learning

Major Project (EC448)

Under the guidance of

Dr. Amit V. Patel
Associate professor
CSRTC

Prepared by

Pankti Shah (17EC078)
8th Semester student
CSPIT EC

Content

- Definition and need of the project
- Project plan
- Functional and non-functional requirement
- Software and Hardware requirement
- Flowchart of project
- Brief description of prostate cancer
- Gleason grading system and histological features
- Information of the dataset
- Transfer learning
- MobileNet

Content

- Feature Extraction
- Data augmentation
- Working of code
- Output
- References
- Conclusion

Definition and Need of the project

- Bio-medical imaging is the emerging domain with the integration of artificial intelligence technology.

- Prostate cancer detection is the new research domain in the field of deep learning.

- Therefore, I wish to understand explore in this field.

Project plan

Date	Task
13/2/2021	Understanding prostate cancer. Exploring dataset Brief introduction to its implemented algorithms.
13/3/2021	Understanding various algorithms Transfer learning Data augmentation Image segmentation
16/4/2021	Code Implementation
30/4/2021	Project completion with final project report

Functional requirement

- Model must work accurately on the prostate cancer dataset and also on disparate datasets.

Software and hardware requirement

- S/W requirement:

  Anaconda software (64-bit)

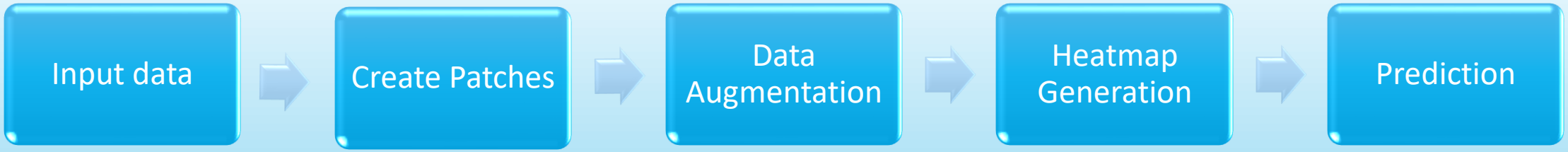
  With installation of TensorFlow, Keras, NumPy, Matplotlib, OpenCV and Os

- H/W requirement:

  8 GB RAM laptop

  i5/i7 processor

Flowchart of project



Brief description of prostate cancer

What is Prostate?

The prostate is a gland which is about the size of the walnut and located below the bladder. It is present in men. Prostate helps in protecting the urinary tract from infections. Prostate produces prostate specific antigen (PSA) which is an enzyme to help liquefy the ejaculation process.

What is Cancer?

Mainly, it is uncontrolled growth of cells in the body.

This abnormal growth is also sometimes known as tumour because there occur a swelling without any inflammation, due to abnormal growth of tissue.

What is Benign Prostate Hyperplasia (BPH)?

With respect to the elevation in the age of men, there is an increase in the size of the prostate which is termed as BPH or benign prostatic hypertrophy. This non – cancerous enlargement may lead to difficulty in urination process and can be easily treated with medication or surgery. BPH is not cancerous and it does not indicate or play role in developing cancer.

What is Prostate Cancer?

A small walnut-sized gland that produces seminal fluid which germinates cancerous cells in the man’s prostate. Cells in the human body are normally tight and bound to the other cell in the core. After generation of the destructive cells in the body the pattern of the cells starts deforming which helps pathologists and experts to indicate the presence of the cancer in the prostate.

What type of cancer is Prostate cancer?

Cancer is an Adenocarcinoma type of cancer.

Adenocarcinomas develop in an organ or a gland.

What is the main cause of Prostate Cancer?

Prostate cancer development mainly depends on the type of food or beverages being consumed. In the present era, most of the edible items have higher rate of fats in it which in turn leads to the improper synthesis and obesity. Simulation of testosterone and other types of hormones elevates due to fats present in the body. Subsequently, this testosterone indirectly starts boosting the cancerous cells to activate at a speedy rate in the prostate. Escalation in the quantity of testosterone leads to simulation of dormant prostate cancer cells into activity state.

What are the signs of Prostate Cancer?

Burning sensation or pain during the urination or ejaculation process. Frequent call of urination especially at the night time. Even, sometimes there occurs difficulty in starting and stopping urination process. Withal, sudden erectile dysfunction takes place. Unexpected flow of blood occurs via the medium of urine or semen. Consequently, insufferable pain in the back, hips, rips and other bones happens.

Which types of food terminates the growth of the cancerous cells in the Prostate?

Mainly, the types of cruciferous vegetables help in languishing the growth of such deadly cells and the risk of advanced stage of cancer. These vegetables include broccoli, cauliflower, cabbage, brussels sprouts, bok choy, spinach and kale.

What is the general age and reasons for the occurrence of the prostate cancer?

The chance of having or the development of prostate cancer in the men below the age of 40 is very rare however it jerkily inclines after the age of 50. On an average, 6 in 10 cases of this cancer are detected in the men who are elder than 65. Among 10000 men of age 40 or below, there is possibility that 1 man might get Prostate cancer. From 40 to 59 age, 1 in 38 have chances of getting it. Man with the age range of 60 to 69 have the highest chances of getting cancer and it is about 1 in 14. Coincidentally, it may transpire due to family history of prostate cancer such as a heredity (ethnicity).

Chances of getting prostate cancer is highest in African American men and its lowest in Asian men. Even, the genes such as HOXB13, BRCA1, BRCA2, MSH2 and MLH1 plays vital role in its generation. The consumption of red meat at a higher rate and low in vegetables are highly prone to this.

What type of genes are involved in the development of Prostate Cancer?

Mutations which are inherited in specified genes such as HOXB13, BRCA1 and BRCA2 play major role for the development of heredity prostate cancer. Even, this mutation seldomly leads to various other types of cancer in man. Men who develop metastatic cancer have a higher rate of mutations in BRCA1 and BRCA2 or in other genes which help in DNA repair. So, around 10% of men have germline mutation and another 10% have somatic mutations.

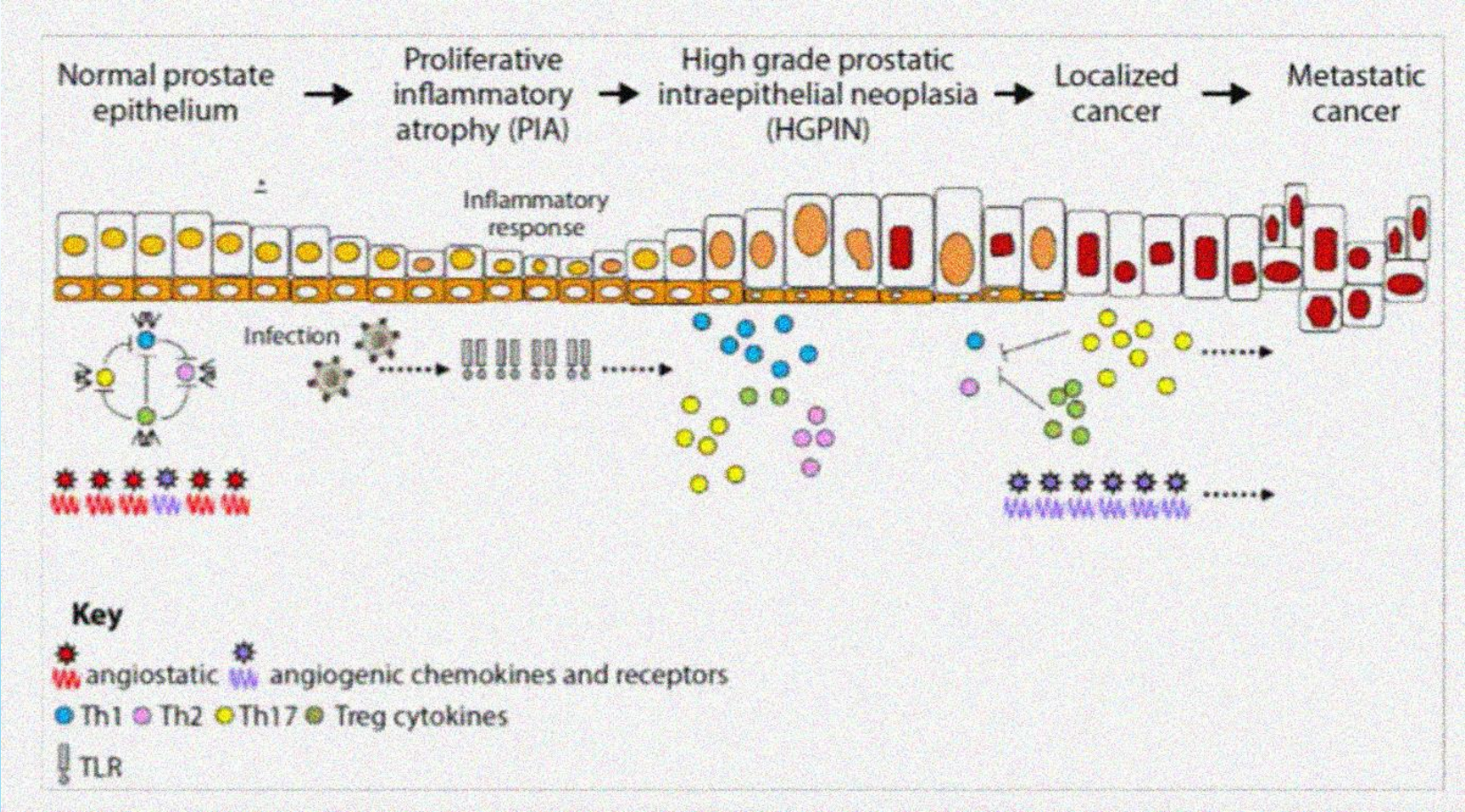
Which races have the chances of its development?

African and American men have maximum chances of developing Prostate cancer in comparison with Caucasian men. These North American and African men are nearly 2.5 times likely to die from this deadly disease. Asian men are at the lowest risk.

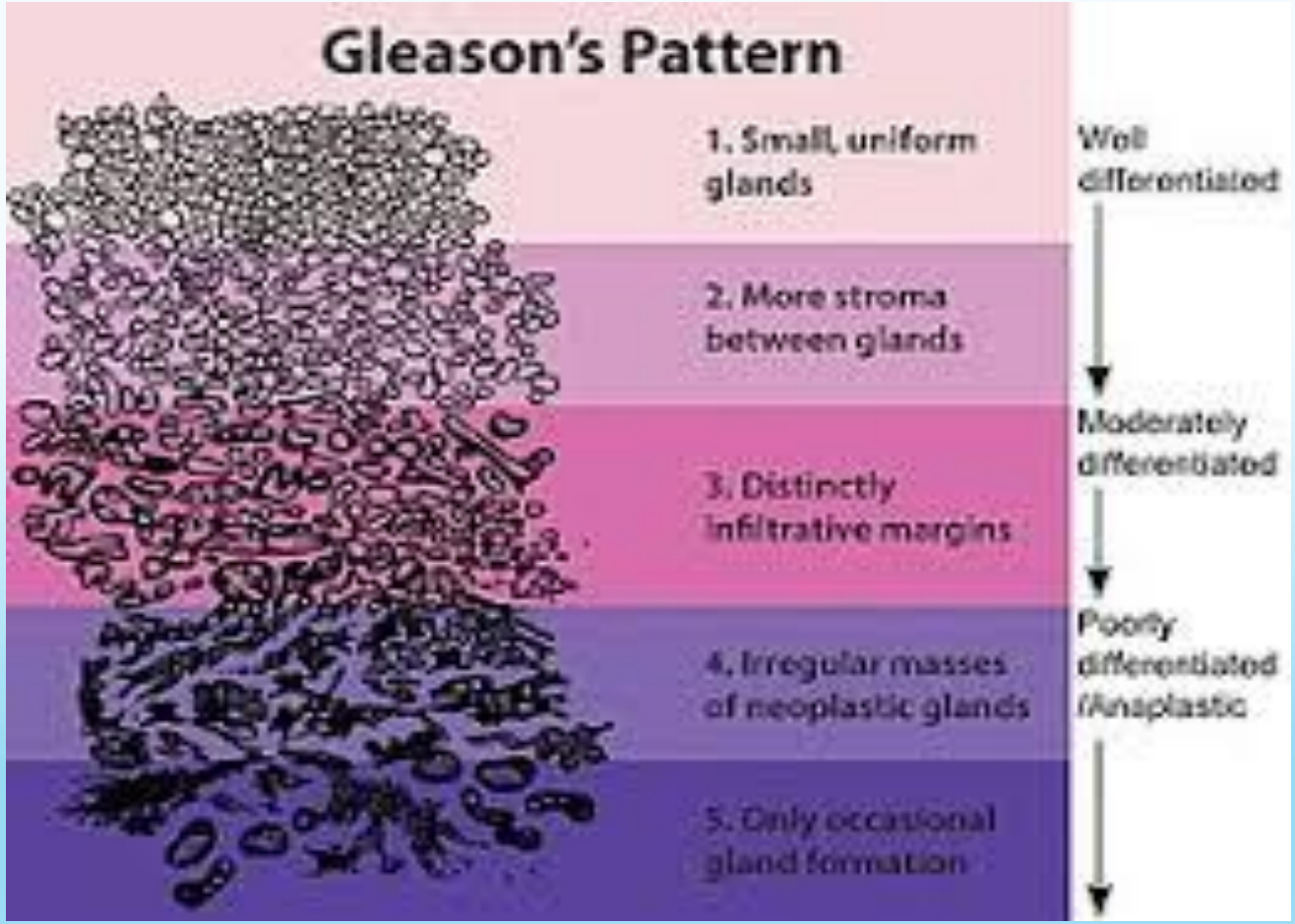
What are the methods for the diagnosis of this cancer?

Among multiple diagnosis and screening methodology, there are four most common methods. Firstly, there is Prostate Specific Antigen (PSA). Secondly, there is Digital Rectal Examination (DRE). Next is Biopsy based examination. Lastly, Multi parametric - MRI (MP-MRI).

Prostate Cancer develops over Time and is Associated with Precursor Lesions



Stages of Cancer



Gleason grading system and histological features

Group 1	Gleason Score 6	5%
Group 2	Gleason Score 3+4=7	17%
Group 3	Gleason Score 4+3=7	35%
Group 4	Gleason Score 4+4=8	37%
Group 5	Gleason Score 9-10	76%

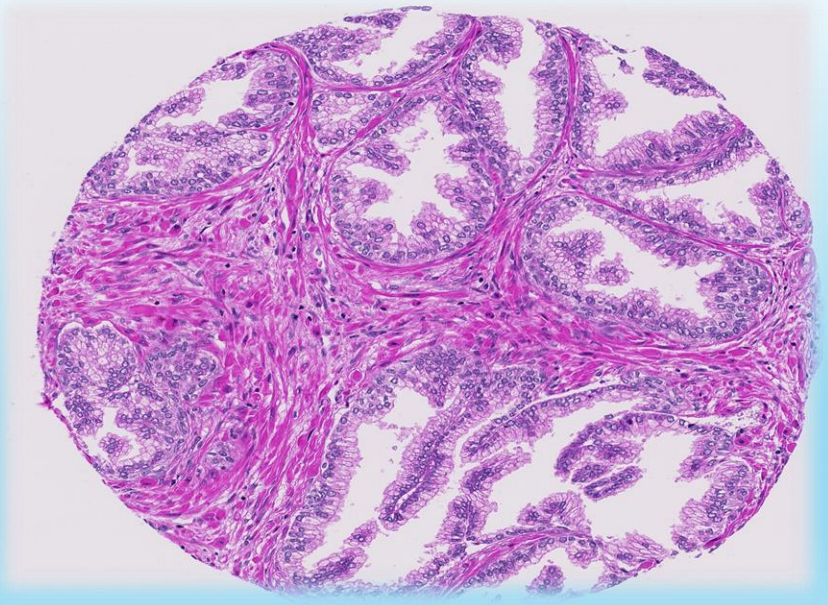
Risk Group*	Grade Group	Gleason Score
Low/Very Low	Grade Group 1	Gleason Score ≤ 6
Intermediate (Favorable/Unfavorable)	Grade Group 2	Gleason Score 7 (3 + 4)
	Grade Group 3	Gleason Score 7 (4 + 3)
High/Very High	Grade Group 4	Gleason Score 8
	Grade Group 5	Gleason Score 9-10

Gleason Grading System

The pathologist looking at the biopsy sample will assign one Gleason grade to the most predominant pattern in your biopsy and a second Gleason grade to the second most predominant pattern. For example: 3 + 4. The two grades will then be added together to determine your Gleason score. Theoretically, Gleason scores range from 2-10. However, since Dr. Gleason's original classification, pathologists almost never assign scores 2-5, and Gleason scores assigned will range from 6 to 10, with 6 being the lowest grade cancer.

What is histological features?

It shows rudimentary necessary features which detects the specified disease by the biological features.



Information of the dataset

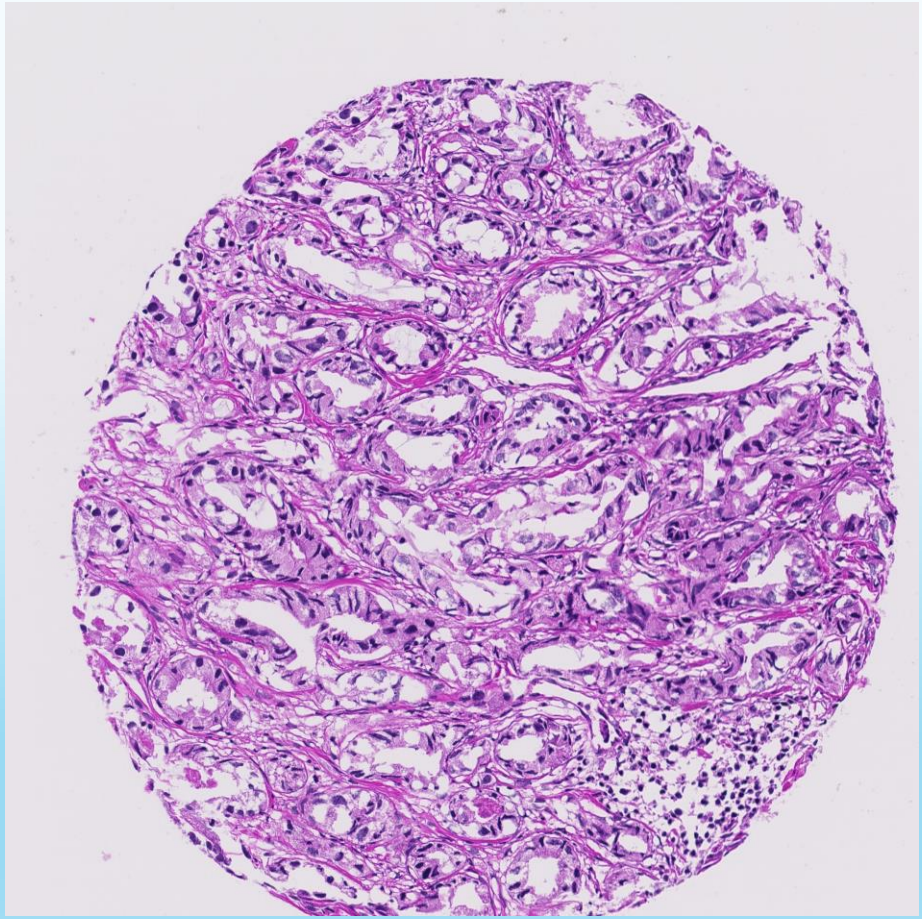
- This dataset has been taken from the HARVARD DATAVERSE
- Two pathologists have graded the test set.
- ZTMAs 76, 80, 111, 199, 204 is the overall data
- ZTMA 80 is graded by pathologists and also used as the test set.
- ZTMAs 76, 111, 199, 204 is used for training data.

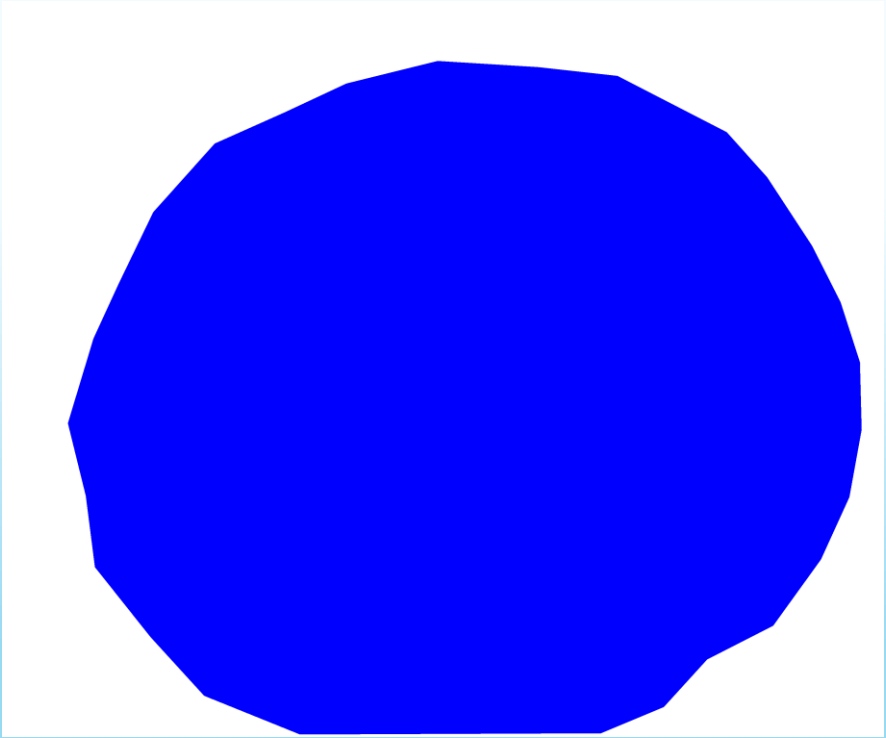
	Benign	6 (3+3)	7 (3+4, 4+3)	8 (4+4, 5+3, 3+5)	9 (4+5, 5+4)	10 (5+5)	Total
TMA 76	42	35	25	15	2	14	133
TMA 80	12	88	38	91	3	13	245
TMA 111	0	95	39	69	16	8	227
TMA 199	61	69	17	26	2	1	176
TMA 204	0	1	2	25	8	69	105

Colour code for masking

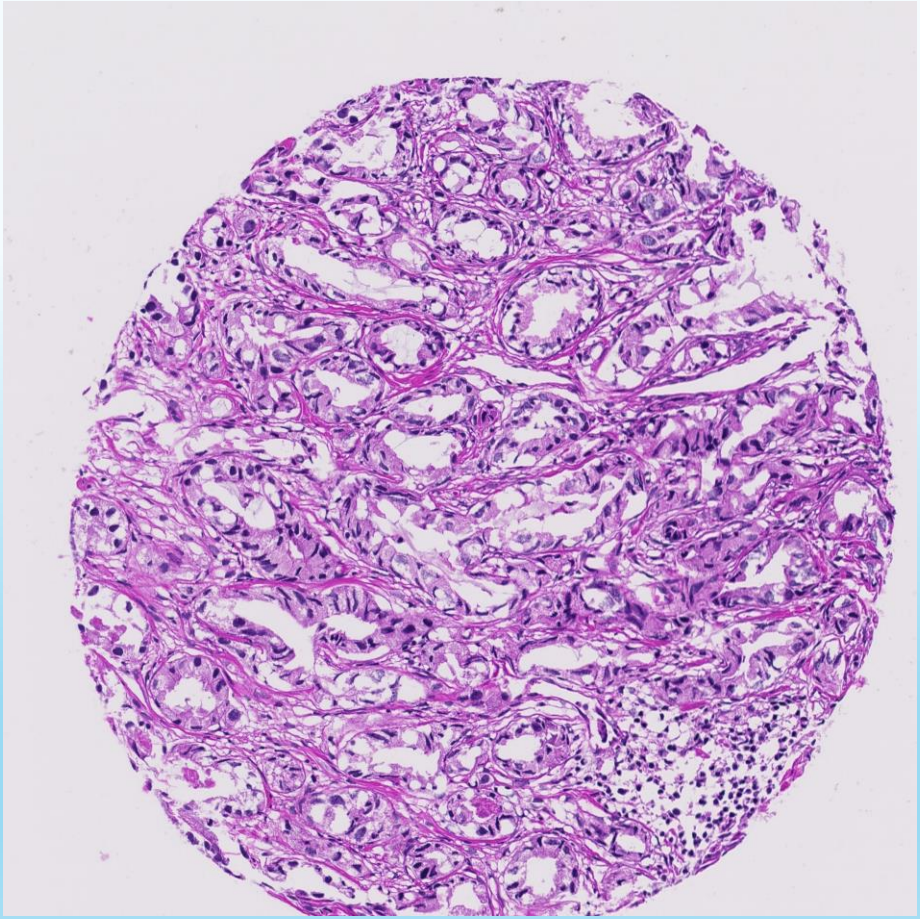
- benign is green
- Gleason 3 is blue
- Gleason 4 is yellow
- Gleason 5 is red
- ignore class is white

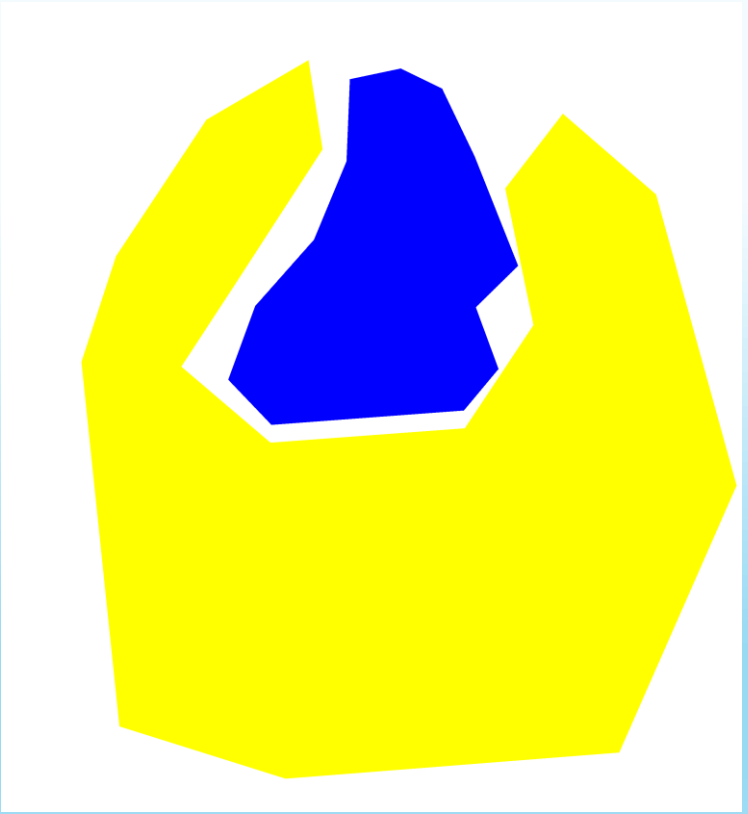
Masked Data by Pathologist1





Masked Data by Pathologist2





Transfer learning

- It is the method of machine learning where a developed model is reused as the starting point for the model of the second task.

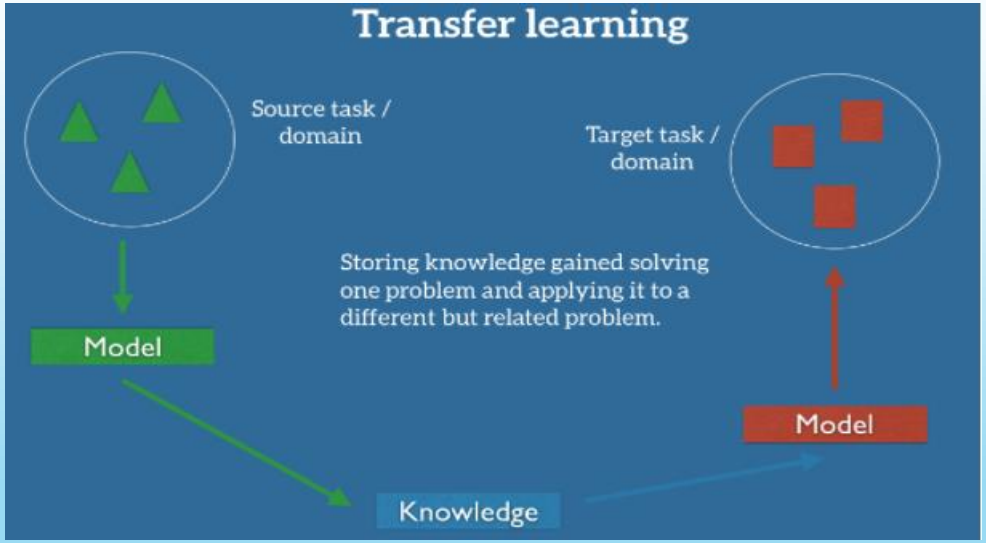


Fig 1. Transfer learning flow

https://ruder.io/transfer-learning/

Transfer learning - Develop Model Approach

- **Select Source Task**. You must select a related predictive modeling problem with an abundance of data where there is some relationship in the input data, output data, and/or concepts learned during the mapping from input to output data.

- **Develop Source Model**. Next, you must develop a skillful model for this first task. The model must be better than a naive model to ensure that some feature learning has been performed.

- **Reuse Model**. The model fit on the source task can then be used as the starting point for a model on the second task of interest. This may involve using all or parts of the model, depending on the modeling technique used.

- **Tune Model**. Optionally, the model may need to be adapted or refined on the input-output pair data available for the task of interest.

Transfer learning - Pre-trained Model Approach

- **Select Source Task**. A pre-trained source model is chosen from available models. Many research institutions release models on large and challenging datasets that may be included in the pool of candidate models from which to choose from.

- **Reuse Model**. The model pre-trained model can then be used as the starting point for a model on the second task of interest. This may involve using all or parts of the model, depending on the modeling technique used.

- **Tune Model**. Optionally, the model may need to be adapted or refined on the input-output pair data available for the task of interest.

When to use transfer learning?

- **Higher start**. The initial skill (before refining the model) on the source model is higher than it otherwise would be.

- **Higher slope**. The rate of improvement of skill during training of the source model is steeper than it otherwise would be.

- **Higher asymptote**. The converged skill of the trained model is better than it otherwise would be.

Description of deep learning architectures

- Architectures with variety of design are deployed as per the requirements.

- In the transfer learning approach, most commonly used architecture is ImageNet.

ImageNet

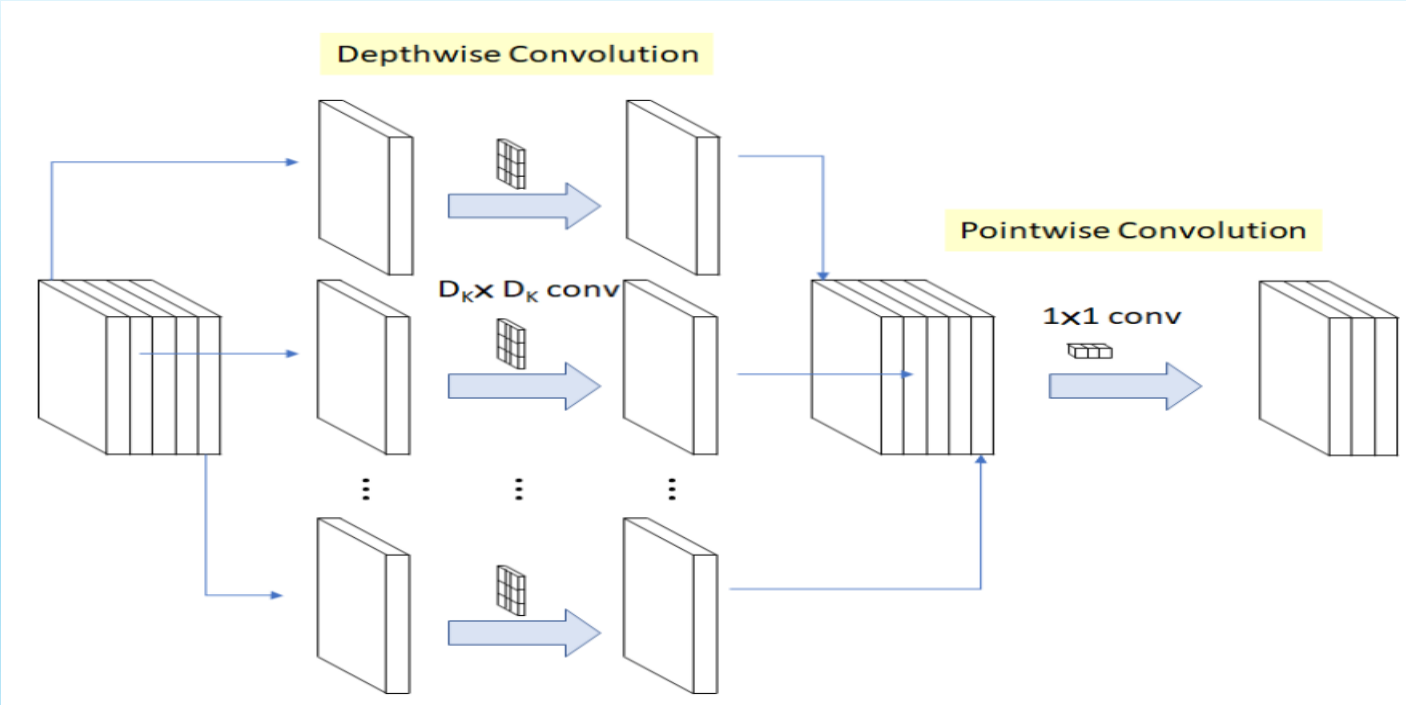
- The Image Net Challenge is an annual competition helped between 2010 and 2017 in which challenge tasks use subsets of the ImageNet dataset.

- The goal of the challenge was to both promote the development of better computer vision techniques and to benchmark the state of the art.

- So, in that architectures were proposed such as DenseNet, ResNet, GoogleNet, Inception Model and many more.

MobileNet

- MobileNet uses depthwise separable convolutions. It significantly reduces the number of parameters when compared to the network with regular convolutions with the same depth in the nets. This results in lightweight deep neural networks.



Feature Extraction

- Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process.

Data augmentation

- Data augmentation in data analysis are techniques used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data. It acts as a regularize and helps reduce overfitting when training a machine learning model.

- It helps us to increase the size of the dataset and introduce variability in the dataset, without actually collecting new data. The neural network treats these images as distinct images anyway. Also Data Augmentation helps reduce over-fitting.

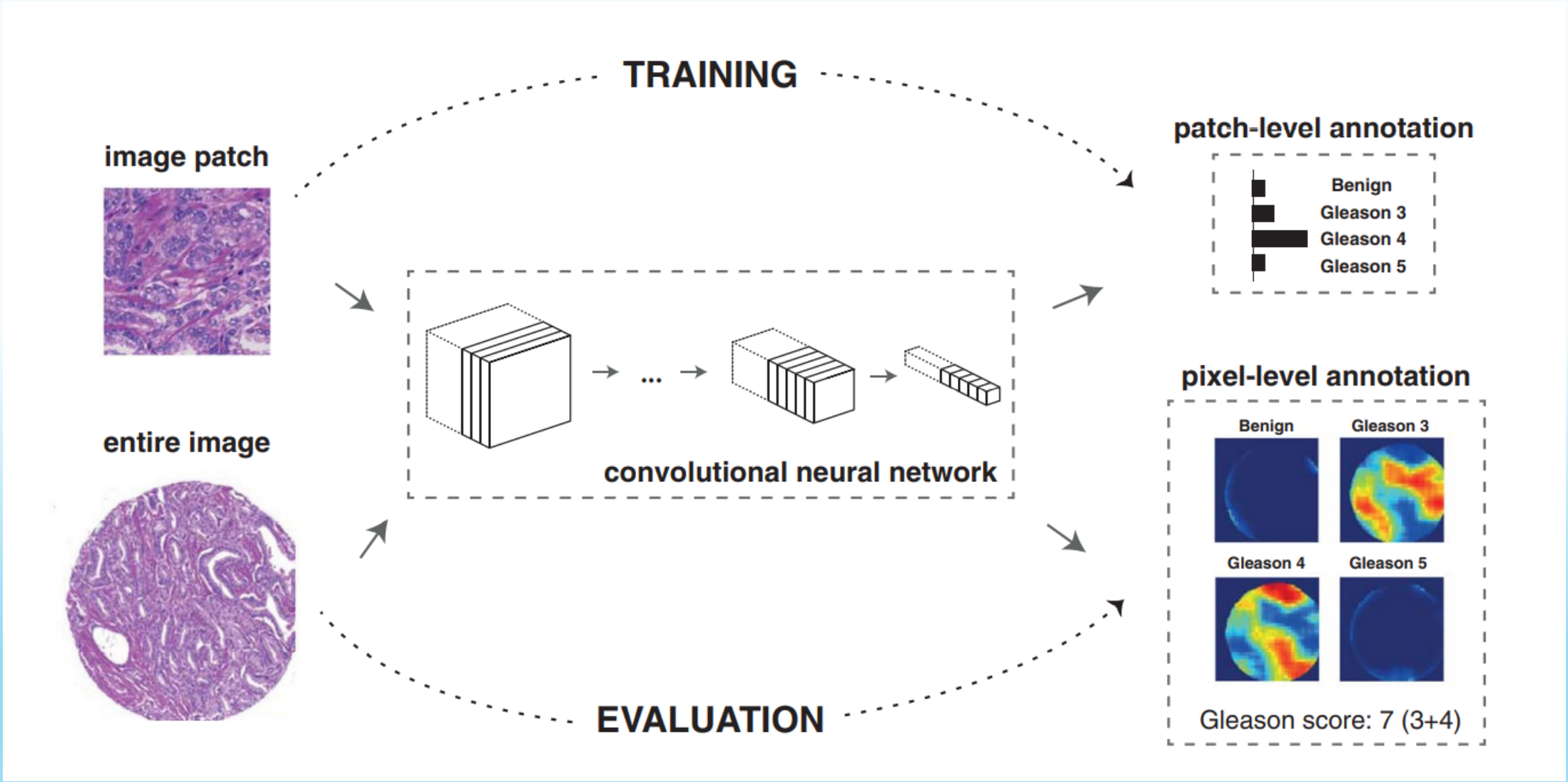
- Trying to implement the code with datagen keras and scipy.ndimage

Image masking

- Masking is a way to tell sequence-processing layers that certain timesteps in an input are missing, and thus should be skipped when processing the data. Padding is a special form of masking where the masked steps are at the start or the end of a sequence.

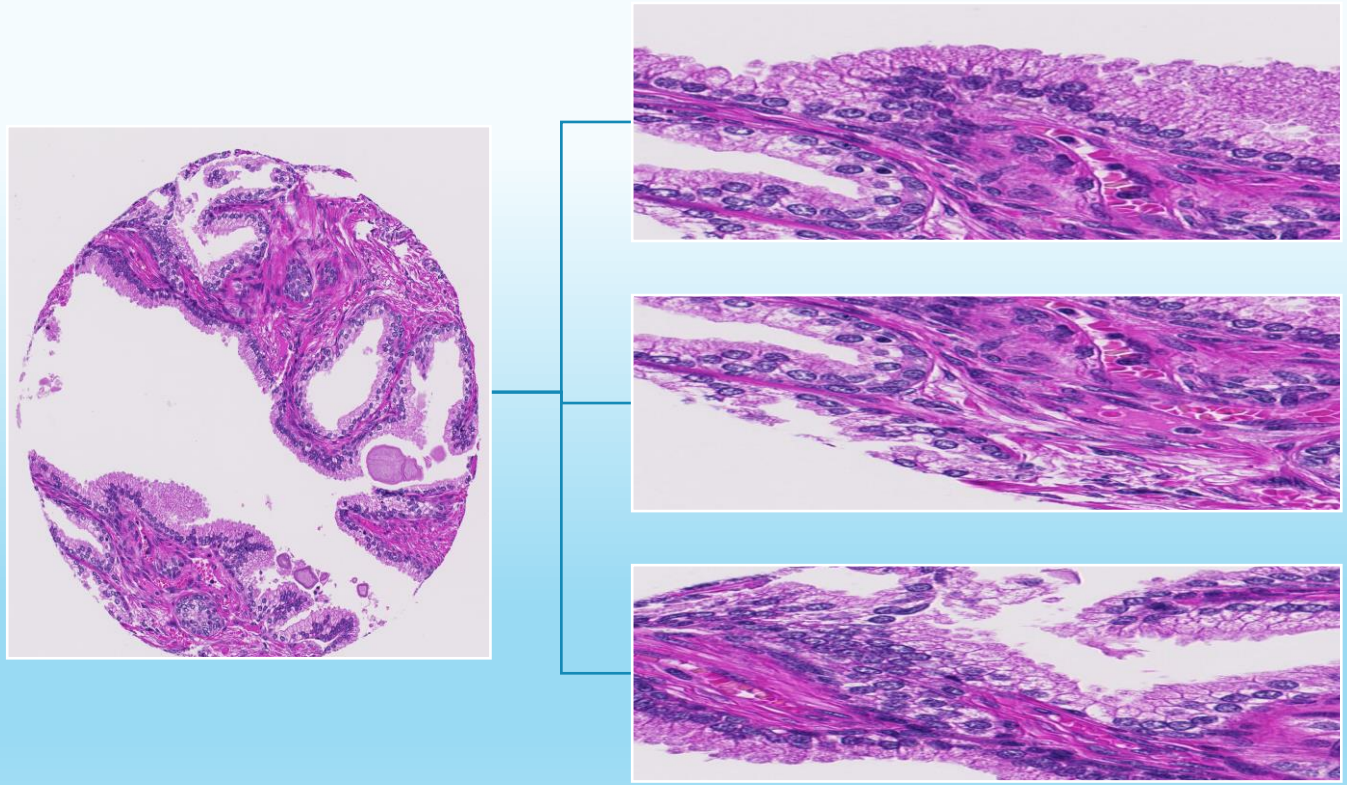
Working of code

- The first step is to create image patches for training and validation.
- We then train a convolutional neural network, e.g. using the MobileNet architecture, to classify image patches
- as benign, Gleason pattern 3, 4 or 5. Given the small size of our dataset, fine-tuning models pre-trained on
- ImageNet achieved better results.
- Once having a good model (the model used in the manuscript is
- provided: (MobileNet_gleason_weights.h5``), we can use it to produce pixel-level probability maps
- for entire TMA spots and visualise class-discriminative regions via class activation mappings
- We only want to make predictions on tissue regions,
- so we first need to create tissue region masks for the TMA images.
- Finally, we compare the model's and pathologists' Gleason annotations on the test cohort.

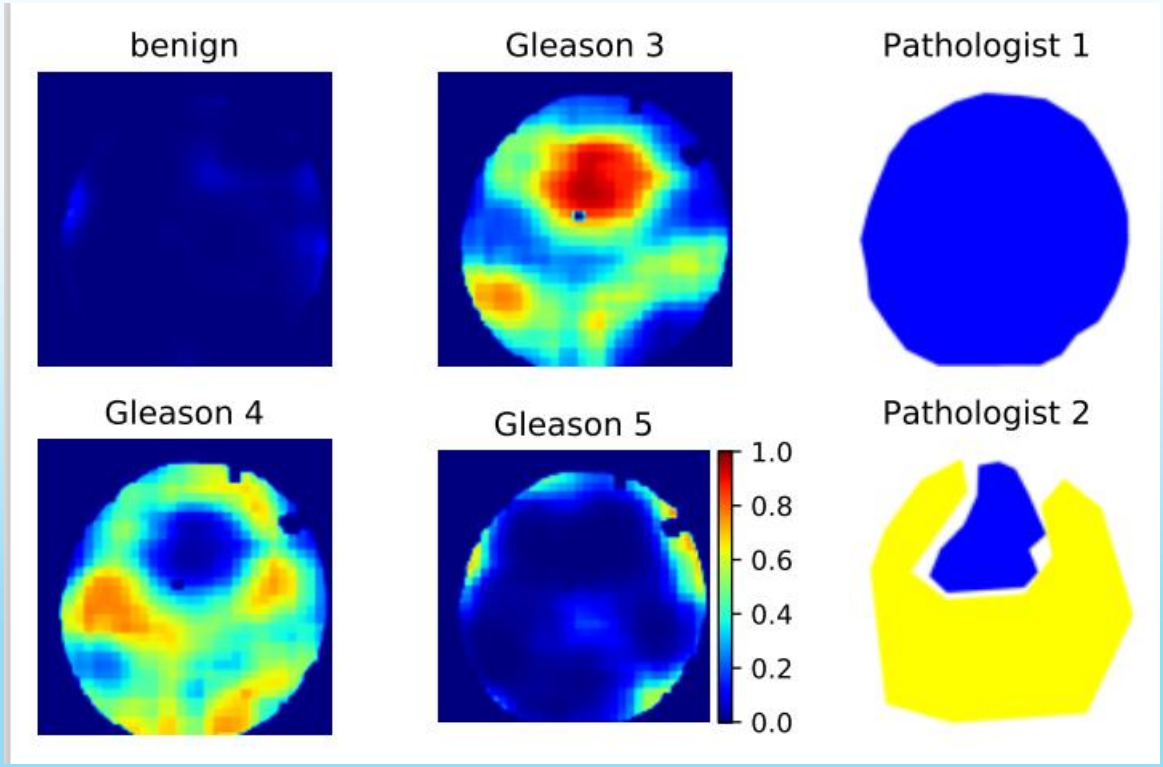

TRAINING
image patch
entire image
convolutional neural network
patch-level annotation
Benign
Gleason 3
Gleason 4
Gleason 5
pixel-level annotation
Benign
Gleason 3
Gleason 4
Gleason 5
Gleason score: 7 (3+4)
EVALUATION

Creating Patches

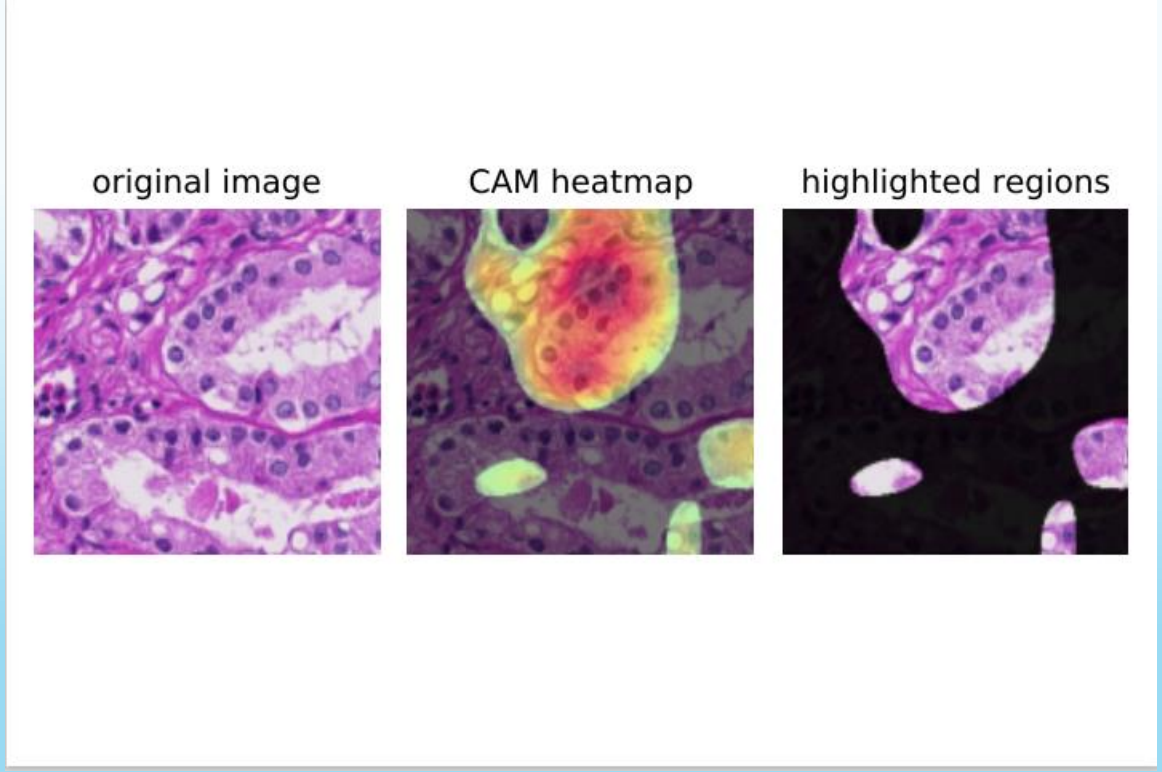
- ZT76_39_A_1_1



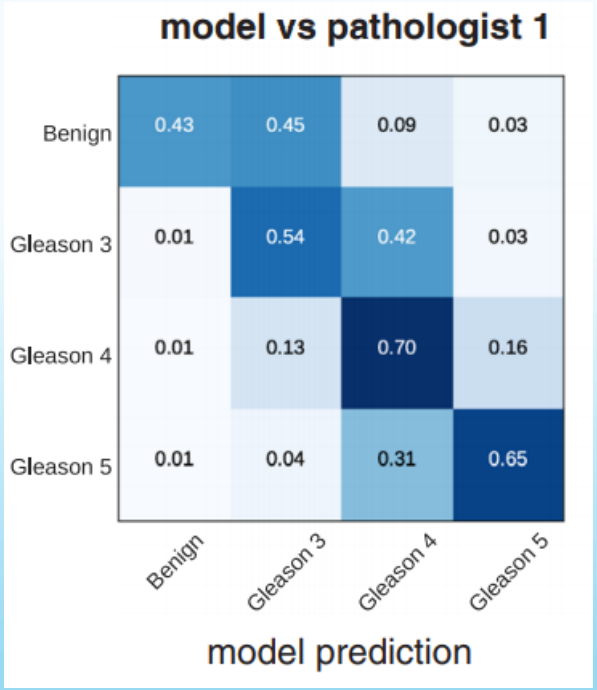
Generating heatmap

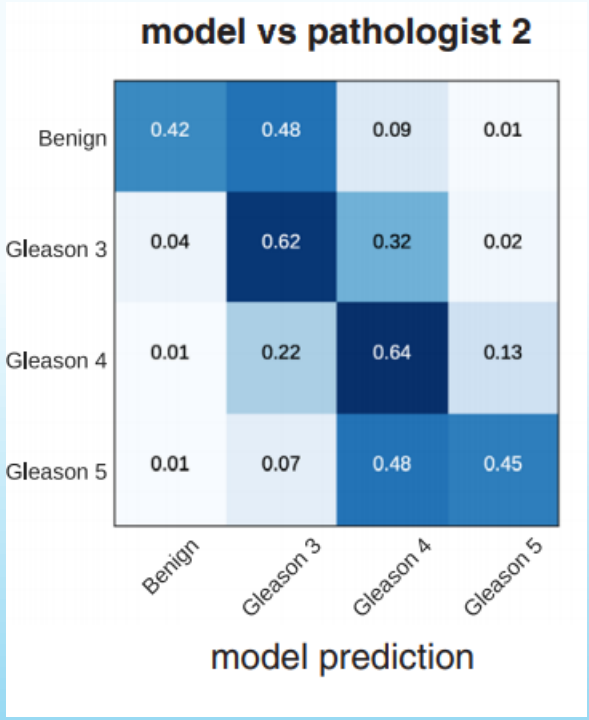


Class Activation Mapping



Result


model vs pathologist 1
Benign
Gleason 3
Gleason 4
Gleason 5
0.43
0.45
0.09
0.03
0.01
0.54
0.42
0.03
0.01
0.13
0.70
0.16
0.01
0.04
0.31
0.65
Benign
Gleason 3
Gleason 4
Gleason 5
model prediction


model vs pathologist 2
Benign
Gleason 3
Gleason 4
Gleason 5
0.42
0.48
0.09
0.01
0.04
0.62
0.32
0.02
0.01
0.22
0.64
0.13
0.01
0.07
0.48
0.45
Benign
Gleason 3
Gleason 4
Gleason 5
model prediction

References

[1] Automated Gleason grading of prostate cancer tissue microarrays via deep learning

[2]https://www.youtube.com/user/ThePCRI

[3] Understanding Prostate cancer course from Coursera

[4]AI for medical diagnosis course from Coursera

[5] Automated Gleason Grading of Prostate Biopsies using Deep Learning

[6] C- Net : A reliable CNN for Biomedical image classification