**Project Proposal: Group 7**

**Project Title:**

**Prostate Cancer Grade Assessment**

**Team members:**

Ssuna Mulindwa, Tynan Hollyer, Brandon Swanberg, Raj Vekeria, Yueying Liu

**Abstract:**

Prostate cancer is one of the most common cancers among males. Diagnosis of prostate cancer can be achieved through a visual classification of tissue called a Gleason pattern (integer from 3-5). Machine learning visual classification can be applied to a dataset of 20,000 images of tissue and Gleason patterns.

**Motivation and Significance:**

Prostate cancer is the second common cancer among males around the world. It results in over 350,000 deaths annually, and over 1 million new diagnoses reported every year, but it can be treated successfully if identified earlier in the stages. Diagnosis of prostate cancer is generally through grading the prostate tissue samples by a pathologist with the Gleason score system they use. It plays an important role on how a patient should be treated. If the grading process went wrong, the patient is going to take a risk of missing a severe diagnosis or unnecessary treatment. However, sometimes, even pathologists have disagreements on the correct Gleason score. This makes the problem of assigning Gleason scores valuable since it has the potential to create better health outcomes than manual analysis. The key to avoiding solving this problem is to develop an automated deep learning system to assist pathologists.

**Related Work and Background:**

One challenge that must be considered is the problem of incorrect labels. As previously stated, the diagnosis of prostate cancer severity is difficult for pathologists.The field of machine learning mitigates this issue by treating the diagnosis as an image classification problem. The AlexNet and ResNet18 networks are a few examples of visual classification machine learning models that have been deployed to identify prostate cancer severity. The most successful model, Resnet34, achieved an accuracy of 97.6% (Tătaru et al). The network was created out of a need to add more layers to convolutional neural networks (CNNs) which is achieved through skip connections. Our team hopes to match and exceed this accuracy by exploring different classification machine learning models such as LeNet-5 or VGG-16. We may also experiment with other factors such as number of epochs and or image transformations.

**Proposed Work and Methods:**

1. Problem to solve:

* We want to build a visual classification machine learning model to classify different levels of cancer tissue into levels from 3 to 5 based on growth patterns of the tumor. According to this, we can determine Gleason score, then convert it to ISUP grade.

(Gleason score and ISUP grade follow the guidelines of the international society of Urological Pathology)

* Our goal is to reach the highest accuracy we can in determining Gleason patterns.
* The more accurate the Gleason pattern is, the better it is at determining Gleason scores, which is mapped to an ISUP grade.
* ISUP grades determine the level of treatment. If ISUP grades are determined effectively, this method of machine learning can be more useful in real settings.
* The best ML algorithms reached >.9 accuracy. We would like to reach similar levels of accuracy.

1. Dataset:

Dataset source:<https://www.kaggle.com/competitions/prostate-cancer-grade-assessment/data>

The dataset is provided by the MICCAI 2020 challenge, it contains more than 20,000 images, train.csv and test.csv file. train.csv contains image\_id, data\_provider,isup\_grade, gleason\_score, as shown in Figure 1.

| Figure.1Dataset   |  | | --- | |
| --- | --- |
|  |

Image\_id: image code

Data\_provider: institution name of data provider

Isup\_grade: the severity rate for cancer

Gleason\_score: an alternate cancer severity rating system

1. Proposed approach and methods

step :

1. Clean dataset

2. Pre-process images

3. Generate data patches

4. Data augmentation (if necessary)

5. Build a visual classification ML model

6. Train our model

7. Based on the performance of our model, if possible we are going to make some improvements to our model.

Potential Approach Method:

CNN model(LeNet, ResNet,etc)

Or Other potential model (custom CNN model)

**Conclusions:**

In conclusion, at 350,000 deaths annually, prostate cancer poses a huge threat, however, with some of the methods we looked at above (Resnet34 for example), we should be able to classify the severity of some prostate cancers from scans with a fairly high accuracy, and using something like this earlier on in a diagnosis could potentially help save lives.

Works Cited

Choi, Hyungeun, et al. “Fig. 5. the Structure of Resnet-12.” ResearchGate, 10 Dec. 2020, <https://www.researchgate.net/figure/The-structure-of-ResNet-12_fig1_329954455>.

*Prostate cANcer graDe Assessment (PANDA)* *Challenge, MICCAI 2020 Challenge.* (2020, April 21). Retrieved April 7, 2022. <https://www.kaggle.com/competitions/prostate-cancer-grade-assessment/overview>

*Residual networks (resnet) - deep learning*. GeeksforGeeks. (2022, January 27). Retrieved April 7, 2022, from <https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/>

Tătaru, O. S., Vartolomei, M. D., Rassweiler, J. J., Virgil, O., Lucarelli, G., Porpiglia, F., Amparore, D., Manfredi, M., Carrieri, G., Falagario, U., Terracciano, D., de Cobelli, O., Busetto, G. M., Del Giudice, F., & Ferro, M. (2021, February 20). *Artificial Intelligence and machine learning in prostate cancer patient management-current trends and future perspectives*. Diagnostics (Basel, Switzerland). Retrieved April 7, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7924061/>