

Introduction To Artificial Intelligence Project

Group Members

Addo Abdul-Razak Musah – 10677768

Martey Noah - 10678268

Prostate Cancer Analysis

School of Physical and Mathematical Sciences

Department of Computer Science

University of Ghana, Legon.

1. Introduction

Prostate Cancer is the most common cancer in American men. It is a slow-growing, potentially lethal disease usually found in men over the age of 50. Although cases of the disease have been reported in all age groups, more than 80 percent of all prostate cancers occur in men over the age of 65.

According to the National Cancer Institute, doctors diagnosed 198,100 new cases of prostate cancer in 2001, and about 31,500 men died from the disease. That means about 19 out of every 100 men born today will be diagnosed with prostate cancer, and four of every 100 men will die from the disease, or about one death every 16 minutes [1].

This is mostly so because most patients are not able to detect Prostate Cancer symptoms until it is too late. Recent studies in medicine have made efforts to identify factors that lead to Prostate Cancer. Researchers have also found that detecting these risk factors early can help prevent the occurrence of Prostate Cancer through lifestyle adjustments and medication. Screening is recommended if you are a man between 55–69 years old, an African–American and have a family history of prostate cancer. This research aims at developing an effective and efficient method for predicting prostate cancer using data mining techniques which is cheaper and effective and will help the early detection of the disease using the risk factors and in effect reduce the rate of deaths associated with prostate cancer [2].

1.1. Research Objective

The main objective of this research is to develop a model and integrate it into a web page to analyze prostate MRI scan to determine the probability that the prostate is cancerous or non-cancerous.

The project aims to employ the use of Machine Learning technology to detect cancer in an MRI of a prostate.

2. The Prostate

The prostate is a small, walnut-sized and shaped gland deeply embedded in the center of the pelvis where it produces a milky fluid that carries sperm during ejaculation [1]. The prostate surrounds

Prostate cancer prediction from MRIs 2

the urethra. The urethra is a tube that carries urine from the bladder out through the penis. The main job of the prostate and seminal vesicles is to make fluid to bathe semen. During ejaculation, sperm is made in the testicles, and then moves to the urethra. At the same time, fluid from the prostate and the seminal vesicles also moves into the urethra. This mixture of semen and fluid from the prostate and seminal vesicles forms the ejaculate that passes through the urethra and out of the penis. Prostate is known more for the problems it causes than the function it serves [2]. Below are some of the factors that have been shown to increase the risk of Prostate Cancer:

- ✓ Age
- ✓ Ethnicity
- ✓ Family History
- ✓ Smoking
- ✓ Diet
- ✓ Environmental factors

3. Materials and Methods

3.1. Data Source

The dataset used to train the machine learning model consists of a total of 867 prostate MRIs; 767 cancerous prostate MRIs and 100 non-cancerous prostate MRIs. The cancerous prostate MRIs were obtained from The Cancer Imaging Archive (TCIA) [3][4][5], while the non-cancerous prostate images were obtained from the Prostate MRI Database [6].

Images were split into three separate folders namely: train, test and validation with each folder having a positive and negative folder containing MRIs of cancerous prostates and non-cancerous prostates respectively. The train folder contains 460 positive images and 60 negative images, test folder contains 154 positive images and 20 negative images, and validation folder contains 153 positive images and 20 negative images. The images were then used to train the model, which is a 10-layer neural network, in 30 epochs, with 50 iterations per epoch. The response attribute had a value of 1 for patients with prostate cancer and 0 for patients without prostate cancer. Figure 1.0 shows the text representation of accuracy (y-axis) and loss (y-axis) for each epoch (both x-axes) and Figure 1.1 is a graphical representation of accuracy (y-axis) and loss(y-axis) for each epoch (both x-axes). Figure 2.0 is tabular representation of model summary. Overall accuracy was 99%, loss was 1%.

```

Epoch 1/30
50/50 [=====] - 211s 4s/step - loss: 0.2566 - acc: 0.9681 - val_loss: 2.1744e-04 - val_acc: 1.0000
Epoch 2/30
50/50 [=====] - 192s 4s/step - loss: 0.0257 - acc: 0.9932 - val_loss: 1.3473e-05 - val_acc: 1.0000
Epoch 3/30
50/50 [=====] - 186s 4s/step - loss: 3.3865e-04 - acc: 1.0000 - val_loss: 0.8820 - val_acc: 0.6304
Epoch 4/30
50/50 [=====] - 183s 4s/step - loss: 0.0210 - acc: 0.9931 - val_loss: 2.0068e-06 - val_acc: 1.0000
Epoch 5/30
50/50 [=====] - 191s 4s/step - loss: 2.1735e-05 - acc: 1.0000 - val_loss: 5.1717e-07 - val_acc: 1.0000
Epoch 6/30
50/50 [=====] - 193s 4s/step - loss: 0.0071 - acc: 0.9965 - val_loss: 4.4230e-08 - val_acc: 1.0000
Epoch 7/30
50/50 [=====] - 246s 5s/step - loss: 1.5084e-06 - acc: 1.0000 - val_loss: 3.3031e-08 - val_acc: 1.0000
Epoch 8/30
50/50 [=====] - 291s 6s/step - loss: 7.3241e-07 - acc: 1.0000 - val_loss: 6.0729e-08 - val_acc: 1.0000
Epoch 9/30
50/50 [=====] - 284s 6s/step - loss: 0.0558 - acc: 0.9972 - val_loss: 1.3515e-07 - val_acc: 1.0000
Epoch 10/30
50/50 [=====] - 242s 5s/step - loss: 4.9217e-05 - acc: 1.0000 - val_loss: 5.9263e-08 - val_acc: 1.0000
Epoch 11/30
50/50 [=====] - 247s 5s/step - loss: 0.0048 - acc: 0.9986 - val_loss: 1.8620e-07 - val_acc: 1.0000
Epoch 12/30
50/50 [=====] - 205s 4s/step - loss: 6.2748e-05 - acc: 1.0000 - val_loss: 2.9856e-09 - val_acc: 1.0000
Epoch 13/30
50/50 [=====] - 187s 4s/step - loss: 1.4482e-07 - acc: 1.0000 - val_loss: 2.3945e-09 - val_acc: 1.0000
Epoch 14/30
50/50 [=====] - 197s 4s/step - loss: 1.0536e-07 - acc: 1.0000 - val_loss: 2.1743e-09 - val_acc: 1.0000
Epoch 15/30
50/50 [=====] - 198s 4s/step - loss: 1.1749e-06 - acc: 1.0000 - val_loss: 1.7623 - val_acc: 0.8774
Epoch 16/30
50/50 [=====] - 175s 3s/step - loss: 0.0434 - acc: 0.9966 - val_loss: 1.7280e-09 - val_acc: 1.0000
Epoch 17/30
50/50 [=====] - 185s 4s/step - loss: 6.2863e-06 - acc: 1.0000 - val_loss: 8.6563e-10 - val_acc: 1.0000
Epoch 18/30
50/50 [=====] - 262s 5s/step - loss: 9.2047e-07 - acc: 1.0000 - val_loss: 3.0473e-10 - val_acc: 1.0000
Epoch 19/30
50/50 [=====] - 281s 6s/step - loss: 1.1638e-07 - acc: 1.0000 - val_loss: 5.6092e-10 - val_acc: 1.0000
Epoch 20/30
50/50 [=====] - 268s 5s/step - loss: 0.0068 - acc: 0.9979 - val_loss: 2.7504e-07 - val_acc: 1.0000
Epoch 21/30
50/50 [=====] - 247s 5s/step - loss: 5.6400e-06 - acc: 1.0000 - val_loss: 6.8943e-09 - val_acc: 1.0000
Epoch 22/30
50/50 [=====] - 253s 5s/step - loss: 3.3129e-07 - acc: 1.0000 - val_loss: 9.5654e-10 - val_acc: 1.0000
Epoch 23/30
50/50 [=====] - 232s 5s/step - loss: 3.0365e-08 - acc: 1.0000 - val_loss: 1.7916e-10 - val_acc: 1.0000
Epoch 24/30
50/50 [=====] - 188s 4s/step - loss: 0.0042 - acc: 0.9979 - val_loss: 1.8295e-07 - val_acc: 1.0000
Epoch 25/30
50/50 [=====] - 178s 4s/step - loss: 8.6940e-08 - acc: 1.0000 - val_loss: 1.7458e-09 - val_acc: 1.0000
Epoch 26/30
50/50 [=====] - 175s 4s/step - loss: 3.4376e-08 - acc: 1.0000 - val_loss: 1.2279e-09 - val_acc: 1.0000
Epoch 27/30
50/50 [=====] - 187s 4s/step - loss: 3.9153e-08 - acc: 1.0000 - val_loss: 1.0244e-08 - val_acc: 1.0000
Epoch 28/30
50/50 [=====] - 182s 4s/step - loss: 4.8309e-04 - acc: 1.0000 - val_loss: 7.6333e-09 - val_acc: 1.0000
Epoch 29/30
50/50 [=====] - 196s 4s/step - loss: 4.5961e-08 - acc: 1.0000 - val_loss: 2.6936e-10 - val_acc: 1.0000
Epoch 30/30
50/50 [=====] - 181s 4s/step - loss: 1.1970e-08 - acc: 1.0000 - val_loss: 1.4025e-10 - val_acc: 1.0000

```

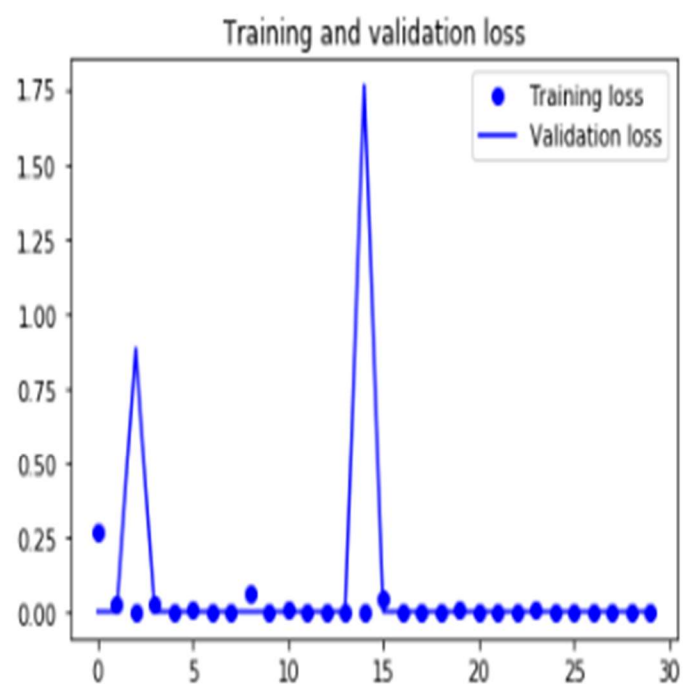
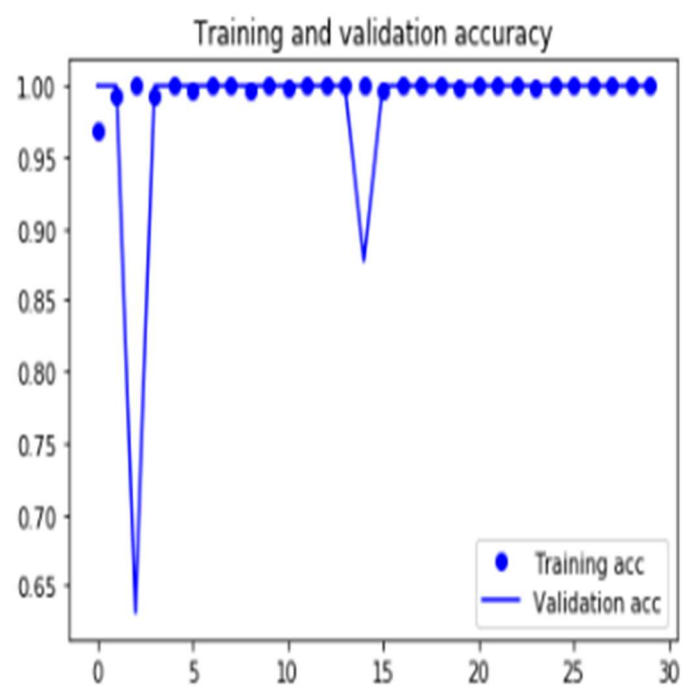


Figure 2.0 Tabular representation of model summary

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 254, 254, 32)	896
max_pooling2d (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_1 (Conv2D)	(None, 125, 125, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 64)	0
conv2d_2 (Conv2D)	(None, 60, 60, 128)	73856
max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 128)	0
flatten (Flatten)	(None, 115200)	0
dropout (Dropout)	(None, 115200)	0
dense (Dense)	(None, 512)	58982912
dense_1 (Dense)	(None, 1)	513
Total params: 59,076,673		
Trainable params: 59,076,673		
Non-trainable params: 0		

4. Conclusion

The overall objective of this research is to predict more accurately the presence of prostate cancer among patients. From our studies conducted at the moment, we have been able to come up with a model that makes a prediction more accurately.

This system may be further expanded to use other data mining techniques for the prediction, such as Clustering, Time series and Neural Networks Association rules.

References

1. <https://www.genome.gov/10005585/prostate-cancer-background-information>
2. <https://www.urologyhealth.org/urologic-conditions/prostate-cancer>
3. <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=3277260>
4. Choyke P, Turkbey B, Pinto P, Merino M, Wood B. (2016). Data From PROSTATE-MRI. The Cancer Imaging Archive. <http://doi.org/10.7937/K9/TCIA.2016.6046GUDv>.
5. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057.
6. <https://prostatemrimagedatabase.com/Database/index.html>

Appendix: How to set up and run the project

1. The model was not included in the zip archive because the file size surpassed the limit specified on Sakai, so I uploaded it to Google Drive instead. Download it here: <https://drive.google.com/file/d/16okdOrfepodrPIjxwKzR6aaePM4Sqef2/view?usp=sharing> and put it in the prostateAnalysis folder, next to the manage.py file.
2. Open the project folder(prostateAnalysis) in Command Prompt
3. Run conda activate
4. Run pip install django djangorestframework django-cors-headers tensorflow
5. Run python manage.py runserver
6. Open <http://127.0.0.1:8000> in the browser.