Utilizing Image Classification Techniques t
Distinguish Cancerous Cell Tissue From
Healthy Cell Tissue

Sammie Haskin

Kennesaw State University

Introduction

The implementation of a successful cancer treatment hinges on:

- Correct diagnoses
- Proper treatment protocol
- Timely treatment

Artificial Intelligence (AI) can thus play a pivotal role in the accurate identification of cancer.

Introduction

Within this analysis, we sought to create an algorithm that coul
reliably classify cell tissue as benign or cancerous.

To do so, two key learning methods were implemented:

- Machine Learning - the process of utilizing statistical algorithms to di
  patterns within the data to perform:
  - Classification
  - Clustering
  - Prediction

- Deep Learning - Augments the capabilities of machine learning throu
  utilization neural networks that simulate the capacities of the humar
  learn.

Data

- From the works of Borkowski et. al (2019), images of both ben cancerous lung and colon cell tissue were utilized.
- The resulting data set contained two features:
  - The image
  - The associated classification of each image into one of five categorie
    - Benign Lung Tissue
    - Benign Colon Tissue
    - Lung Squamous Cell Carci
    - Colon Adenocarcinoma
    - Lung Adenocarcinoma
- The images of histopathological cell tissue are similarly measu 768x768 .jpeg files that exhibited a high degree of quality for

Data

- Derived from various sources made openly available, the original data set contained 250 images within each of the 5 classes.
- Utilizing image augmentation techniques, Borkowski et al. (2019) increased that total to exactly 25,000 images with 5000 images belonging to each class.
- Using these data, an algorithm could be created to improve the diagnostic capabilities of the medical system.

Figure 1: Pie Chart of Sample Size for each Classification of Lung and Colon Cells

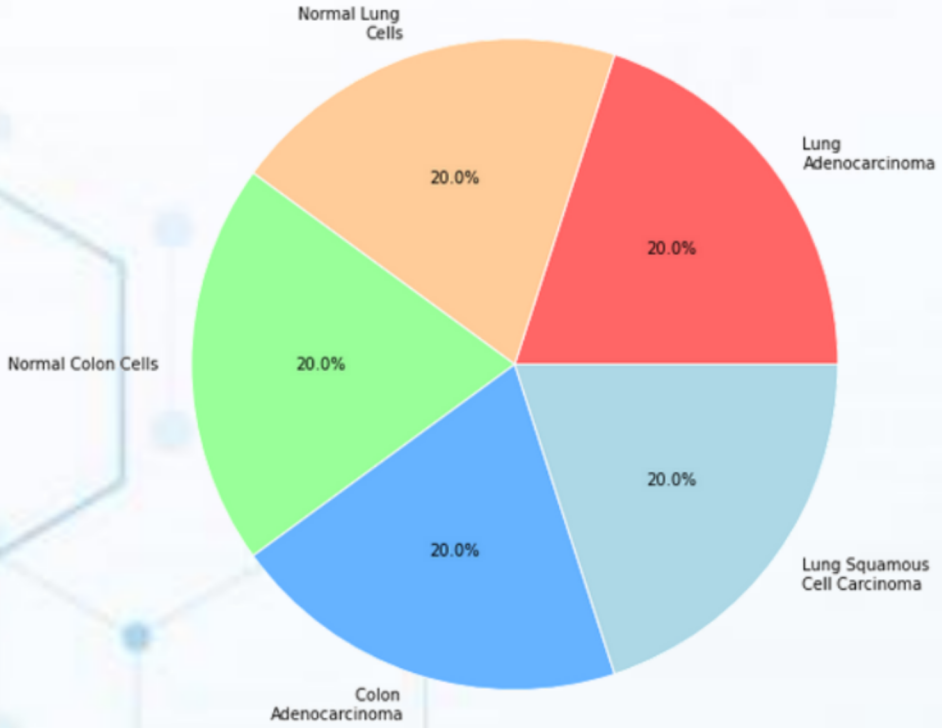
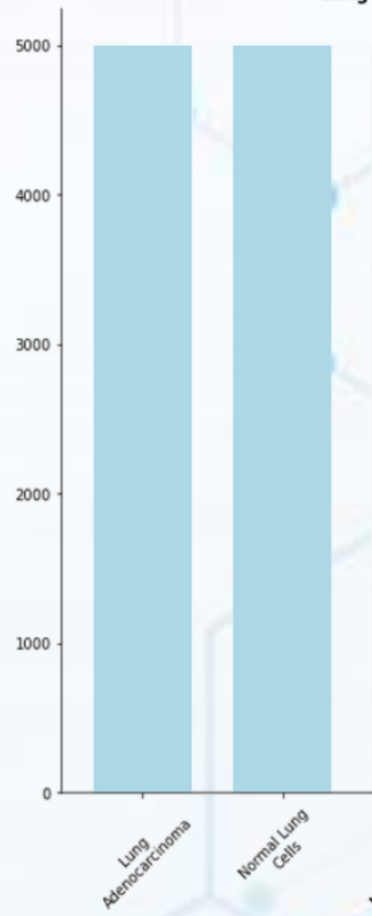


Figure 2: Bar Chart of Sa
Lung



Methods

Within this analysis, the python programming language was util

In the Extraction-Transform-Loading process (ETL) images were:

- retrieved from the repository of Burkowski et al. (2019)
- read in batches into memory as 150x150, 3 color channel images
- represented as a normalized numerical array

As images were read in batches, the process of modeling could
performed without loading all images into memory simultaneou

With the Keras image generator, the two utilized models were subsequently created, trained, and evaluated.

Methods

- The machine learning procedure of Multiple Logistic Regressio then implemented using the Keras API.
- This machine learning algorithm is widely established as a mea classifying observations into one of multiple classes.
- This algorithm allows for the classification of observations give information known about them.

$$\Pr(Y_i = 1) = \frac{e^{\beta_1' \cdot \mathbf{X}_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k' \cdot \mathbf{X}_i}}$$

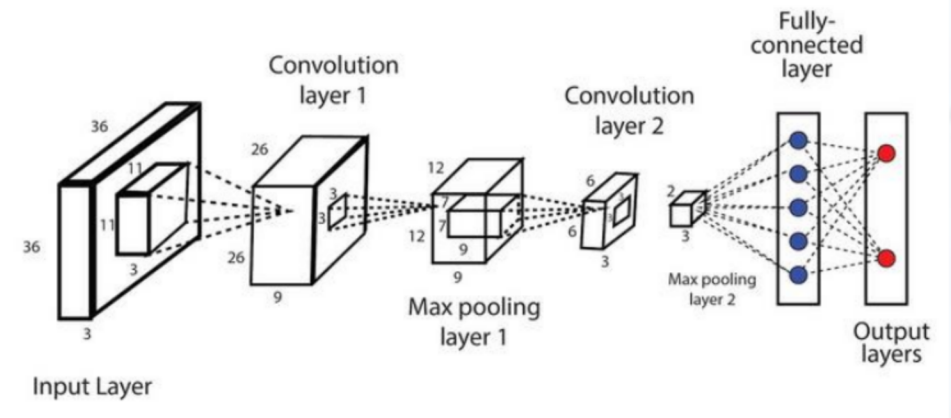
$$\cdots\cdots$$

$$\Pr(Y_i = K-1) = \frac{e^{\beta_{K-1}' \cdot \mathbf{X}_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k' \cdot \mathbf{X}_i}}$$

$$\Pr(Y_i = K) = \frac{1}{1 + \sum_{k=1}^{K-1} e^{\beta_k' \cdot \mathbf{X}_i}}$$

Methods

- For the purposes of comparison, a deep learning procedure kn
  a convolutional neural network was created using Keras.
- Convolutional neural networks are generally accepted as an ex
  effective technique for the classification of images.
- These algorithms function similarly to the inner workings of th
  human brain in its classification of visual stimuli.



Methods

- For the evaluation of each model, 20,000 images were utilized each model while the remaining 5,000 images were utilized to each model.
- To compare these two models, the accuracy and F1 score of ea model was evaluated and compared.

$$\text{Precision} = \text{TP} / (\text{TP}+\text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP}+\text{FN})$$

$$F = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Results

In the evaluation of the efficacy of the logistic regression model

- The logistic regression model resulted in accurate predictions of each 40.65% of the time on the training data.
- Accurate predictions resulted 29.75% of the time when testing on da had not been previously exposed to.
- Similarly, this model resulted in an F1 score of .2975 on the validation set.

- These results seemed to indicate that the logistic regression m predicted slightly better than chance.
- The accuracy of this model was nowhere near optimal for use clinical settings, however.

Results

In the evaluation of the Convolutional Neural Network:

- A high degree of accuracy on the training data was indicated with a s
  94.45%.
- When exposed to the remaining 5,000 images, the algorithm predict
  correct class of each image 92.0% of the time and resulted in an F1 s
  .9210.

These metrics seemed to indicate that:

- this model generalizes well to new data points
- the algorithm could be relatively reliable for use in clinical settings

Discussion

In summary:

- The deep learning model was significantly accurate in its classificatio of each tissue as benign or cancerous.
- The logistic regression procedure, alone, could not produce the same of accuracy as produced by the deep learning model.
- The convolutional neural network appeared to generalize well to new points.

Such indicates the possibility that the model's utilization in clinic settings in the classification of histopathological images of lung cell tissue could greatly aid in future diagnoses.

Conclusion

- Through the use of convolutional neural networks, the efficacy of the classification of images in various fields can be greatly increased.
- The effectiveness of these procedures was illustrated in this project through the classification of colon and lung cell tissue as benign or cancerous.
- With more data, the effectiveness of the model can be further optimized for its utilization for the diagnoses of cancers.

References

Borkowski AA, Bui MM, Thomas LB, Wilson CP, DeLand LA, Mast SM. Lung and Colon Cancer Histopathological Image Dataset (LC arXiv:1912.12142v1 [eess.IV], 2019. Retrieved from https://www.kaggle.com/andrewmvd/lung-and-colon-cancer-histopathological-images.