**Histopathology Cancer Detection**

**Sasidhar Guthi**

**IJ43577**

**Data Science Department, University of Maryland Baltimore County**

**Data 690-Modern Practical Deep Learning**

**Prof. Andre Nguyen**

**Spring 2023**

**Abstract**

Pathologists face a substantial increase in workload and complexity of histopathologic cancer diagnosis due to the advent of personalized medicine. Traditionally, pathologists analyze tissue samples under a microscope to identify cancer cells. However, this process is time-consuming, subjective, and prone to errors. The emergence of deep learning, specifically Convolutional Neural Networks (CNNs), has revolutionized cancer detection in histopathology images. This report focuses on how CNN find evidence of cancer in images and also by using autoencoders an unsupervised deep learning model, which would improve the accuracy.

**Introduction**

For the last hundred years, cancer diagnosis and grading have been primarily based on the microscopic examination of haematoxylin and eosin (H&E) stained sections. Detailed protocols, which include microscopic analysis, are available for many prevalent cancer types such as lung, breast, and prostate, and their use has resulted in effective prognostic and grading strategies such as the Gleason grading system. CNNs are neural networks specifically designed for image analysis and can learn to identify relevant features in the image. In histopathology, CNNs can be trained on large datasets of labelled images to accurately detect cancer cells. Deep learning models can also be trained to identify cancer subtypes, providing valuable information for personalized treatment planning. Autoencoders, on the other hand, are unsupervised deep learning models that can learn a compressed representation of the input data while preserving the essential features.In this report, we detect histopathologic cancer by using a CNN and a combination of CNNs and autoencoders. We train a CNN on a large dataset of Histopathologic images and classify them into normal images or images with evidence of cancer. The same process we are going to perform with autoencoders and observe how the models are going to perform.This approach demonstrates the potential of combining CNNs and autoencoders in histopathologic cancer detection. If unsupervised learning would again take over supervised learning.

**Literature Review**

Over the past five years, there have been some initial studies discussing the use of "deep learning" methods for analysing microscopic and histopathologic images. Ciresan et al. were the first to apply convolutional neural networks to count mitosis in primary breast cancer grading. (Ciresan, 2013). In another publication, they also demonstrated the suitability of patch-driven convolutional neural networks for segmentation tasks. (Ciresan, 2012). Wang et al. later extended the work on mitosis detection by combining handcrafted features and convolutional neural networks. (Wang, 2014). Other applications of convolutional networks include primary breast cancer detection, glioma grading, and epithelium and stroma segmentation. Su et al. used another deep learning method called stacked denoising auto-encoders to perform cell detection and segmentation in lung cancer and brain tumours. (Su, 2015).

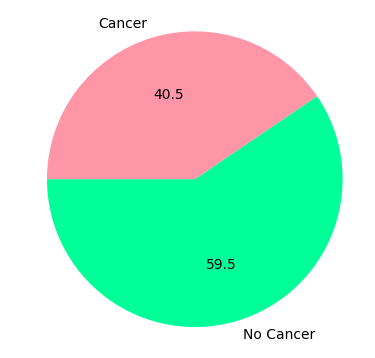
Unlike CNN-based feature representation which uses convolutional and subsampling operations to learn a set of locally connected neurons for feature extraction using local receptive fields. Instead, we employ an autoencoder architecture. In this architecture, the "encoder" network models pixel intensities using lower dimensional attributes, while the "decoder" network reconstructs the original pixel intensities using these low-dimensional features. (J.Xu, 2016). Then it is an uncontrolled learning model that aids in data reconstruction by processing features derived from compressed data. It is essentially a neural network model that can learn the internal structure of data to regenerate the input data. In other words, it captures data characteristics to rebuild the input data. The autoencoder model is trained to rebuild data using a neural network. (Toğaçar, 2020).

**Modelling**

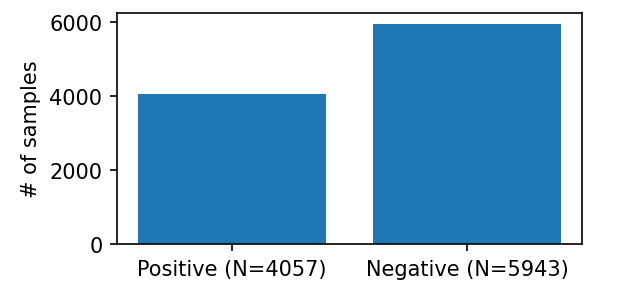
**Data and Data Preparation**

Collected the data from a Kaggle Competition dataset. This data consists of test and train data which we have used to train our model. The dataset size is 6.3 GB, used google drive to warehouse the dataset and then by using the colab environment extracted the data to train models.

The training dataset consists of 57,458 values, whereas the test dataset consists of 220,025 values. In the training dataset, the value counts of images having cancer and images with no cancer are as follows.



We have loaded 10,000 images to analyse and use for modelling. We have got 4057 positive samples and 5943 negative samples.



**Data Pre-Processing**

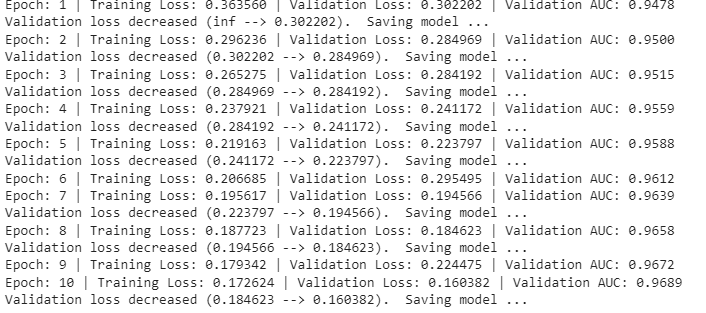
Performed transformations to the data by using techniques like flipping, rotating and normalization. This is to ensure the diversity of training data is increased and to prevent overfitting the model to the training data.

They created a batch size of 128 and 10 per cent of training data to use for validation. Prepared data loaders for training and validation.

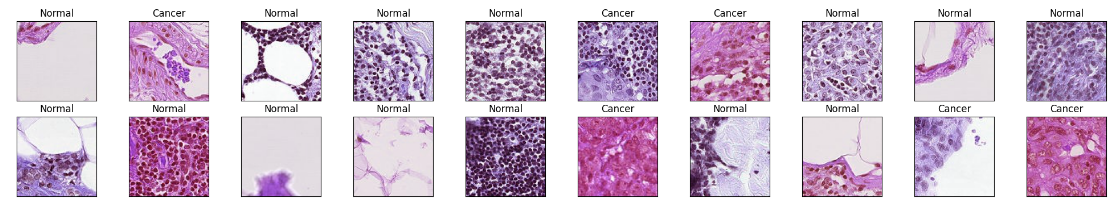
**Modelling**

**Convolutional Neural Network**

Built a Convolutional Neural network (CNN) with 5 convolutional layers with an increasing number of filters 32 to 512. Each convolutional layer is followed by batch normalization, ReLU activation function, and max pooling operation for downsampling. The output of the convolutional layers is flattened and fed into a fully connected neural network with three linear layers and ReLU activation functions. The last layer uses a sigmoid activation function to output a binary classification result. This CNN model has given the following evaluation report.



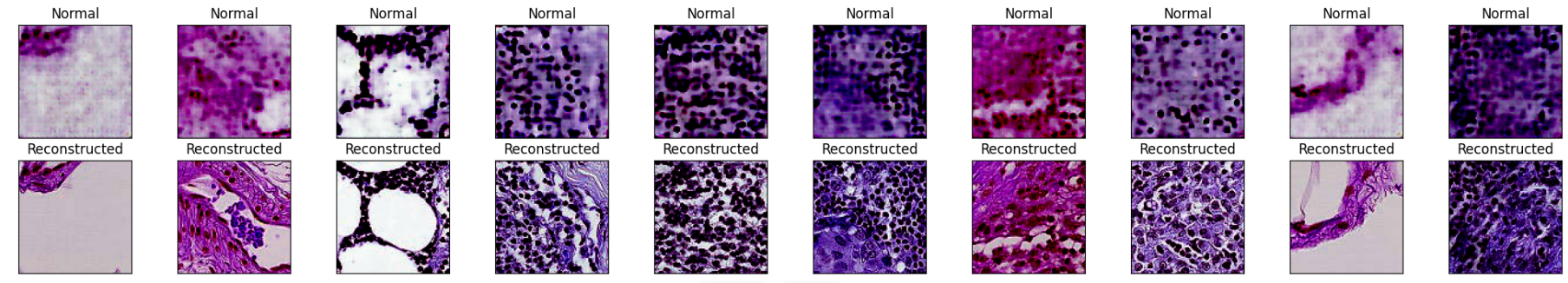
A validation area under the curve of 96.89% is achieved by the CNN model. Where these are the image predictions this CNN performed on the Histopathology test cancer dataset.



**Autoencoder Model**

An Autoencoder model with 3 colour channels is used, it is passed through a series of convolutional layers with an increasing number of filters (32,64,128,256, and 512) and max pooling layers, resulting in a compressed representation of the input. The decoder then takes this compression representation and passes it through a series of transposed convolutional layers with decreasing numbers of filters (512, 256, 128, 64, and 32) and no pooling, resulting in a reconstructed output image with the same dimensions as the input. The Sigmoid function is applied to the output of the final layer of the decoder, and the forward () method defines the flow of the data through the model.

Evaluated the test dataset using this autoencoder model.



With Unsupervised learning, the autoencoder model has helped in image processing by denoising and reconstructing the images in the test dataset.

***Challenges Faced and Future Improvements***

Some challenges experienced:

* The RAM of the colab processor was not enough to run over 10 epochs. The results would have been better if the model is trained for 20 epochs.
* The Reconstructed images were stored on the drive, these reconstructed images will be more useful in detecting the evidence of cancer.
* This was not possible because of an issue with the path name access.

**Conclusion**

This project discusses creating a CNN model to identify the evidence of Cancer in the Histopathology images. The CNN has performed well resulting in a 97% Area Under the Curve. It also goes on to show how the Unsupervised learning model Autoencoder has helped to enhance the images. From the results, we can observe that there was a lot of noise in the dataset images. The Autoencoder has denoised the images and reconstructed them so that the pathologists would have a better visual of the evidence of cancer-affected cells.

**References**

Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., ... & Van Der Laak, J. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports*, *6*(1), 1-11.

Ciresan, D. C., Giusti, A., Gambardella, L. M. & Schmidhuber, J. Mitosis detection in breast cancer histology images with deep neural networks. Med Image Comput Comput Assist Interv. 8150, 411–418 (2013)

Ciresan, D., Giusti, A., Gambardella, L. M. & Schmidhuber, J. Deep neural networks segment neuronal membranes in electron microscopy images. Advances in Neural Information Processing Systems. 25, 2843–2851 (2012)

Wang, H. et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Journal of Medical Imaging 1, 034003–034003 (2014).

Su, H. et al. Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders. Med Image Comput Comput Assist Interv, 9351, 383–390 (2015)

J. Xu et al., "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images," in IEEE Transactions on Medical Imaging, vol. 35, no. 1, pp. 119-130, Jan. 2016, doi: 10.1109/TMI.2015.2458702.

Toğaçar, M., Ergen, B., & Cömert, Z. (2020). Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders. *Medical hypotheses*, *135*, 109503.