National College of Ireland

Lung Cancer Classification from Histologic Images using Capsule Networks

MSc Research Project
Data Analytics

Bedanga Bikash Roy Medhi
Student ID: x18182127

School of Computing
National College of Ireland

Supervisor: Dr. Rashmi Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Bedanga Bikash Roy Medhi
Student ID:	x18182127
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Rashmi Gupta
Submission Due Date:	28/09/2020
Project Title:	Lung Cancer Classification from Histologic Images using Capsule Networks
Word Count:	5760
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	27th September 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Lung Cancer Classification from Histologic Images using Capsule Networks

Bedanga Bikash Roy Medhi
x18182127

Abstract

Cancer is one of the most dangerous and invasive disease in the human body. Lung cancer has the highest mortality rate amongst the other cancers. Therefore, detection and classification has become critical for the diagnosis of lung cancers. The manual determination of cancer from histology slides require expert supervision and takes time. Computer vision systems using Convolutional Neural Network (CNN) have shown remarkable performance in automated detection of cancers. But the current CNN based system have some limitations. In this study the state-of-the art Capsule Networks are used for lung cancers classification from histopathology images. Capsule Network does have the capability to preserve the orientation, pose and texture which is achieved by the vector transformation of the extracted features. For pre-processing the images were stain normalized. The classification performance of the model is evaluated using Matthew's correlation coefficient (MCC), specificity, sensitivity, false negative and false positive rate and accuracy. The Capsule Network implemented in the study achieved 0.99 MCC score and 99% accuracy for lung cancer classification.

1 Introduction

In the ongoing years there are a number of diseases that has overburdened the health system in different countries. Lung cancer is one of the disease that has the highest number of cases and deaths across the globe. Treatment of cancer requires proper examination and governance of expert professionals. Lung cancer is an unrestrained growth of lung tissues within the lung. It can spread to other parts of the human body (malignant) or may confine to a specific region in the lungs(benign tumors). The study considers the classification three different types of lung cancers defined below:

- Benign lung tissue is unusual growth of cells in a specified part of the lung. They do not spread to other paths of the organ like malignant cancers and rather confine at a specified location within the lungs (Henderson and Klebe; 2006).
- Squamous cell cancers[1] is another type of lung cancer that occurs in the squamous cells which are present in the respiratory tract. These cells are found in the airways of the lungs.

[1]https://www.cancer.gov/publications/dictionaries/cancer-terms/def/squamous-cell-carcinoma

- Lung adenocarcinoma[2] mostly occurs in the secretory cells of the lungs. It produces substances like mucus and are localized primarily in the alveoli or smaller airway regions.

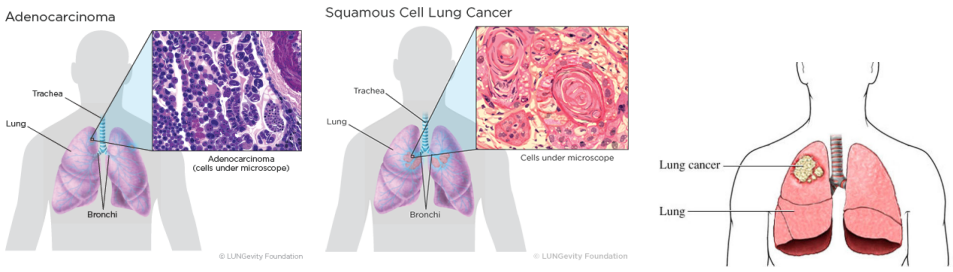


Figure 1: *Left to right:* Lung adenocarcinoma, Lung squamous cell carcinoma, Benign lung cancer[2]

1.1 Background

Globally, lung cancer cases are rising at an alarming rate. In most of the cases it is caused because of smoking. Rapid industrialization has led to degradation of air quality and rise of pollutants in the atmosphere which is another factor for the cause of lung cancer.

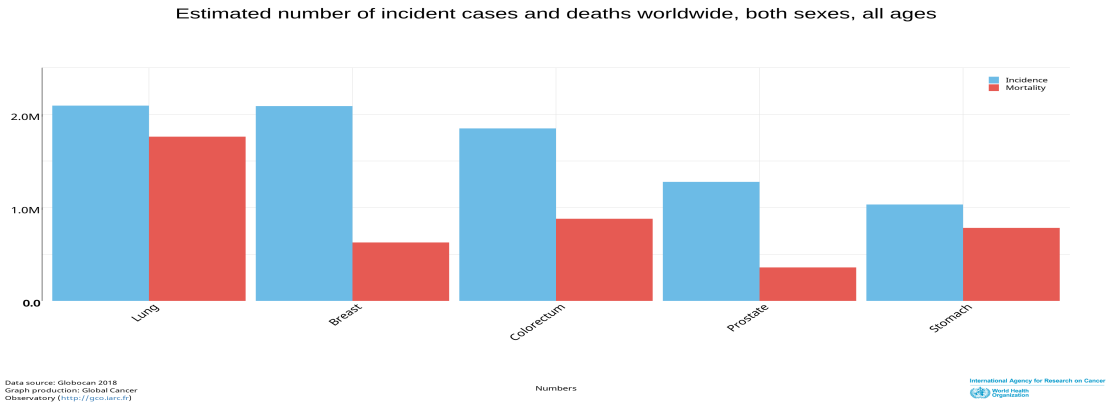


Figure 2: Cancer Statistics Incidence and Mortality Cancer Today (WHO)

The bar diagram in Figure 2 plotted using the data published by IARC (International Agency for Research on Cancer)[3] shows that there are more than 2 million active cases and nearly 2 million deaths because of lung cancer in 2018. After lung cancer, breast cancer takes the next spot with around 2 million cases but the mortality rate is less compared to lung cancers. From the plot it is clear that lung cancer has the highest rate in incidence and deaths among the top 5 cancers.

Histopathology is a salient diagnostic procedures for cancer detection. In histopathology, the doctors perform biopsy of the affected organ and analyse it under a microscope. To define a therapy for a cancer patient it is crucial for the doctors to identify the type

[2]https://www.cancer.gov/publications/dictionaries/cancer-terms/def/adenocarcinoma
[2][1],[2]: https://lungevity.org/;[3]: https://www.asthmafoundation.org.nz/your-health/lung-cancer
[3]https://gco.iarc.fr/today/online-analysis-multi-bars

of cancer. In the current scenario doctors have to examine tissue samples under a microscope to determine the cancer type which is a long time taking process, thus affecting the survival rate of the patient.

1.2 Motivation

The advancement of computer vision and machine learning, has made it possible to design systems that can automatically detect the type of disease a patient is suffering from. The key motivation that drives this research is to design an application that would automatically detect lung cancers from histopathology slides. For the ML research community, there is a high demand to replace the legacy systems with faster and better systems. This would directly benefit the medical departments across the world to replace the current process with a much faster system that would help in saving lives and reduce the mortality rate through out the world.
This study states the use of state-of-art Capsule Network for lung cancer classification(Iesmantas and Alzbutas; 2018). The advantages of Capsule Network over CNN could give an edge for its utilization in medical imaging. One of the main advantage is the ability to capture the spatial orientation of the features in the images by transforming them into vectors (Hinton et al.; 2011). Apart from this, the dynamic routing algorithm that is used within the network allows the capsules to associate the feature information vectors from the lower level to the higher level capsules in the network (Sabour et al.; 2017). It has also shown exceptional results in many standard datasets like MNIST digits, CIFAR-10, MNIST Fashion and SVHN(Rajasegaran et al.; 2019). The study is focused on answering **how well capsule networks perform lung cancer classification from histology scans?**

1.3 Research Objectives

1. Implement a capsule network for lung cancer classification.
2. Use pre-trained models with the same dataset for performance evaluation of the Capsule Network.
3. Use Matthew's correlation coefficient, specificity, sensitivity, false positive rate, false negative rate, accuracy and F1 score to analyse the model performance.

Section 2 contains the literature review of the related work. The Methodology is explained in the section 3. Section 4 and 5 contains the Design Specification and Implementations of the proposed approach. Section 6 discusses the Evaluation strategies that are considered for the study. Lastly, section 7 concludes of the findings of the research and the future work possible using this approach.

2 Related Work

2.1 Introduction

The cases of cancer across the world are rising in an incremental manner. There have been significant efforts from the machine learning community to build systems that could classify and detect cancers from histopathology slides. The availability of cutting-edge computing systems and computer vision techniques, it possible to design such systems.

This section gives a detail review of the studies that have been carried out in this research area.

2.2 Review of techniques applied on cancer detection using deep learning from histopathology images

The time taking and complexity of cancer detection has motivated researches to fast track the existing process of detection for different types cancer that exist today. Lu and Mandal (2015) designed multiple segmentation techniques for classification of skin melanoma. The automated approach consists of five modules: keratinocyte segmentation, epidermis segmentation, melanocytes detection, feature construction and classification. For pathologists the epidermis, melanocytes and keratinocytes are important region of interest that help them to identify melanoma. The model was evaluated using 10 fold cross validation. In many instances, deep learning application have demonstrated better classification and detection capability against pathologists. Here, Hekler et al. (2019) proposed a model for classification of benign nevus and malignant melanoma. The framework consist of a pretrained ResNet50 architecture trained with 595 cropped histopathologic images of different patients. The evaluations on the model showed that it achieves better classification when compared with multiple pathologists on the same image slides.

In another distinctive approach Graham et al. (2019) proposes HoVer Net model which is based on Convolutional neural network(CNN). It is an encoder-decoder framework that performs segmentation and classification from histopathology images. The model calculates the nuclei distances in the image which are used for segmenting of the areas having a group of nucleus in close proximity. The entire study was focused on a single cancer type and emphasises on future work to be extended to other cancer types. In similar context Vo et al. (2019) proposes a boosted convolutional neural network for breast cancer classification. The use of ensemble techniques for better classification was demonstrated in the study. In the study DCNN (Deep Convolutional Neural Network) was used for feature extraction which were then applied to Gradient Boosting models. It also stated that the use of augmentation techniques helps in increasing the dataset size considered in the research.

Pre-trained CNN models for feature extraction and transfer learning has been widely appreciated in the medical imaging studies. Here, Khan et al. (2019) used transfer learning for breast cancer classification from H&E stained images. Color normalization technique was applied to enhance the quality of the image. Pre-trained models like GoogleNet, ResNet and VGGNet were applied in the study. The models were tested by validating their respective accuracy on test samples. It proposes the use of CNN extracted features and annotated features combine together to improve the classification performance, as a future study. On the other hand Gupta and Chawla (2020) used pre-trained networks for a two phase classification model for detection of breast cancer. Feature extraction was carried out with the help of CNN models. The extracted features were used in Logistic Regression and Support Vector Machine for the classification. The pre-trained models like ResNet50, Xception, VGG16 and VGG19 were used for the research.

The ongoing studies in Artificial Intelligence have led to the development of systems that assist doctors in the treatment of cancer. It should not be limited to the detection of cancer. There are other areas in cancer treatment that are still unexplored and there lies a huge scope for deep learning to demonstrate its potential. Liang et al. (2020) performed a review study on the machine learning systems that assist in different cancer

treatment stages. As pointed out in the study there are areas like drug development, chemotherapy, radiology and immunotherapy where deep learning techniques are yet to be fully applied. These developments in AI systems for cancer treatment have shown promising results. Coccia (2020) reviewed the ongoing challenges that doctors face and how deep learning methods could be a game changer in detection of cancers. The onset of these new technologies in computer vision have made it possible for the development of biomedical systems that can assist in cancers treatment.

2.3 Lung cancer classification and detection using deep learning approaches

For the detection of lung cancer, CT scanned images are analysed for the growth of tumors. There have been researches that have presented with techniques for automated detection from CT scans. Liu and Kang (2017) designed a model using CNN that takes multiple views of the lung nodules. The multiple views of the nodules help in identifying the deep features in the images. The study also states that this approach can be used for object detection as well. In the same context, Sun et al. (2017) focused on designing a region based model using multiple deep networks for automatic feature extraction. The proposed feature extraction technique was able to match the performance of the current CADx systems. The study emphasises on the fact that deep learning algorithms have the capability for solving such problems. It also stated that the performance of these algorithms should be tested with large datasets and with deeper architectures.

Histopathology serves as key diagnostic procedure for cancer detection. It is also a tedious and time consuming process, as pathologists have to manually investigate the slides. In such scenario Khosravi et al. (2018) and Teramoto et al. (2019) implemented pre-trained CNN models for faster detection of lung cancer from whole slide images. These pre-trained models, trained with very large dataset have shown proven results in feature extraction. These approaches have shown potential for deep models to classify lung cancer from histopathology slides. Using ensemble approach ALzubi et al. (2019) designed a model for the classification lung cancer. The model is called Weight Optimized Neural Network with Maximum Likelihood Boosting (WONN-MLB). To minimize the classification time of the model Newton-Raphsons Maximum Likelihood was used which helped in selecting the important attributes, leading to higher accuracy. The study showed that the proposed architecture had good diagnostic accuracy and minimum error rate.

Staging and Grading of tumors allows doctors to determine the amount it has spread and the rate of spread of a tumor. In one of the study done by Moitra and Mandal (2020), 1-D Convolutional Neural Networks were applied to determine the staging and grading of non-small cell lung cancer which is one of the major type of lung cancer. The model showed good performance in terms of classification at different stages of the cancer. The study also focuses on further research which can be done using more number of images. In the similar area of interest Bonavita et al. (2020) uses a 3D convolutional neural network for the detection of malignant pulmonary nodules. The detection of malignant pulmonary nodules which helps in better prognosis of lung cancer. This new approach was integrated with the existing system of lung cancer classification thus enhancing the performance of the system. Multiple dataset were taken to test the new approach and results of the evaluation proved its efficacy.

2.4 Application of capsule network in medical imaging

In the latest implementations of medical imaging, Capsule Networks (CapsNet) are being considered as a potential replacement of CNN. Recent studies have proved its capability to stand against the state-of-art approaches. For lung cancers classification Mobiny and Van Nguyen (2018) used CapsNet model with a modified dynamic routing algorithm. The new consistent dynamic routing algorithm was faster and reduced computational cost against high dimensional data. The study showed that CapsNet requires less amount and performs better than the CNNs. A novel CNN based decoder was also implemented to demonstrate better reconstruction. The CapsNet model compared against CNNs showed good results in classification of cancerous and non-cancerous nodules in the lungs from CT scans.

Iesmantas and Alzbutas (2018) demonstrates the use of CapsNet as a classifier for cancer classification. The research uses color normalization technique to resolve the uneven staining problem and for up-sampling of training data. The study showed that CapsNet can be used in medical imaging domain and can be further improved for better performance. A similar study on breast cancers was done by Anupama et al. (2019) using CapsNet. The study demonstrated a comparative analysis with state-of-art models like CNN, InceptionResnetv2 and DCNN against CapsNet; here the CNN models under performed against CapsNet model in classifying breast cancer. This comparative study showed the scope for CapsNet's application for critical disease diagnostic systems.
In another study Baydilli and Ümit Atila (2020) uses CapsNet as a classifier for the classification of white blood cells. The motive of the study is to find the count of these white blood cells which serve as an indication of certain diseases. The study showed the use of the network as a classifier utilizing the encoder section. The evaluations were carried with state of art CNN models like DenseNet, InceptionNet, VGG etc.

In the current scenario of the pandemic, CapsNet model was used by Toraman et al. (2020) for detection of corona-virus infections in lungs. The study focused on binary as well as multi-class classification of patients using X-ray images. It showed that the classification process was faster and accurate. The model had 5 convolutional layers with input size of 128 x 128. The CapsNet model showed good training performance without any issue of over-fitting.

In another unique implementation Lee et al. (2020) used CapsNet for hand gesture recognition. The study proposes the use of this system in the surgical room to prevent the contamination of CAD systems during a surgery. The hand gestures would help in operating the machines without touching it. The study evaluated CapsNet model with a CNN model and a VGG network. The CapsNet model showed the superior performance against the other two models.

2.5 Conclusion

From the review of the different studies, it is evident that there is huge potential for deep learning in cancer diagnostics. The complexity of the disease has brought in, the scope for advance researches that would help in detection and classification. Capsule Network being very new to the medical imaging domain do possess the capability to be utilized for CAD systems that could further enhance detection capability of the existing systems. In some recent applications it has shown remarkable performance thus proving its capability to be extended to larger problems statements. The below Table 1 gives the summary of the entire literature.

Table 1: Literature Review Summary

Author	Methodology	Findings
Lu and Mandal (2015)	Used multiple segmentation techniques to extract region of interest	Improved the classification algorithm designed for skin cancer.
Sun et al. (2017)	Used ROI based approach for feature extraction	Multiple models were used in the study which showed better capability than the existing CAD system.
Liu and Kang (2017)	Used convolutional neural network for early detection from CT images	Detection of cancer by extracting multiple views of the nodules in lungs.
Khosravi et al. (2018)	Used pre-trained Google Inception network	Demonstrated the use of pre-trained CNN for lung cancer classification.
Mobiny and Van Nguyen (2018)	Capsule Network	Used Capsule network for classification of cancer nodules in lungs. Implemented a modified dynamic routing algorithm to improve performance of CapsNet.
Iesmantas and Alzbutas (2018)	Capsule Network	Showed good performance for breast cancer classification. Used stain normalization.
Vo et al. (2019)	Application of CNN for breast cancer classification	Designed an ensemble model using DCNN and gradient boosting which gave improved performance against traditional approaches.
Khan et al. (2019)	Application of transfer learning for breast cancer classification	Showed the used Stain Normalization and Transfer Learning using pre-trained Neural Network for breast cancer classification.
Graham et al. (2019)	Used segmentation and CNN for classification of cancer from histopathology images	Segmentation helps to segment the image into different regions which helps in identifying different features thus boosting classification performance
Hekler et al. (2019)	Used pre-trained model for classification of malignant melanoma	evaluation of the model was better than 11 pathologists.
Teramoto et al. (2019)	Applied VGG16 network for classification of benign and malignant lung cancer	Showed promising results in classification but there is scope for further study with large dataset.
ALzubi et al. (2019)	Ensembled Classifier using Neural Network	Improved the classification performance using ensemble approaches.
Anupama et al. (2019)	Capsule network, DCNN, InceptionResnetv2, CNN	Comparative analysis showed better results in CapsNet.
Liang et al. (2020)	Review study of Artificial Intelligence on medical imaging and drug development	The study showed that there is a huge scope for Artificial Intelligence on Medical Imaging.
Coccia (2020)	Review study of Deep Learning on cancer imaging	The study showed the use of Deep learning on Cancer Imaging and its present day challenges.
Gupta and Chawla (2020)	Application of pre-trained CNN for classification	Demonstrated the use of pre-trained models for breast cancer classification.
Moitra and Mandal (2020)	Used CNN for feature extraction	The approach showed promising results thus showing the potential for deep learning techniques to estimate the staging and grading of a cancer.
Bonavita et al. (2020)	Applied 3D CNN model	Focused on detecting malignant pulmonary nodules for early detection of lung cancer.
Baydilli and Ümit Atila (2020)	Capsule Network	Designed a classifier for different blood cells helps in the detection of blood related diseases.
Toraman et al. (2020)	Hybrid CNN and Capsule Network	Classifier for detection of COVID-19, pneumonia and uninfected patients.
Lee et al. (2020)	Capsule Network framework	Hand gesture classifier for the operation of CAD systems in surgery rooms.

3 Methodology

In the current study of lung cancer classification the CRISP-DM (Cross-industry standard process for data mining) methodology has been chosen.

3.1 Business Understanding

Cancer being one of the most dangerous diseases in the world that has many different types. To identify the type of cancer in a patient requires several medical cross examinations. It is a time consuming process and requires expert professionals with extensive domain knowledge. CAD systems have greatly improved the diagnostic process of many diseases. But such sophisticated system are only available to the hospitals that have huge capital. There has been constant efforts to build a system that is cost-effective and is highly available. The study aims to achieve this by build a image classification system, that would classify lung cancers in less time and better availability.

3.2 Data Understanding

The histopathology image dataset is collected from Electrical Engineering and Systems Science archive[4] of Cornell University which is open for free download (Borkowski et al.; 2019). The images of the following cancer types are present in the dataset; lung adenocarcinoma, benign lung tissue, lung squamous cell carcinoma, benign colon tissue and colon adenocarcinoma. There are total 25,000 images corresponding to the five types of cancers stated in the documentation. In the current research the lung cancer images will be used which comprises of 15,000 images corresponding to the three types of lung cancer. The dataset has been pre-augmented with geometric transformations. As these are medical images there is no information regarding patients, validated with supervision and are HIPAA compliant.

3.3 Data Preparation

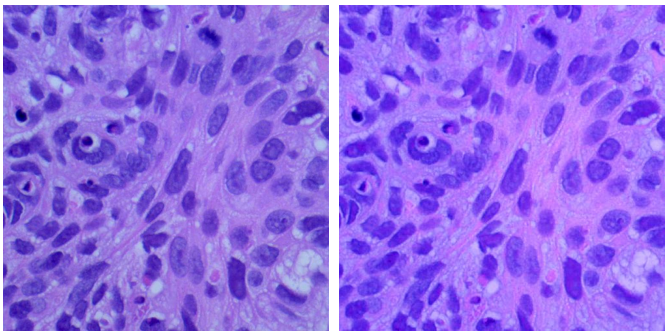
For training the model the raw data has to be transformed and cleaned such that the training of the model is fruitful. Another factor that is important for model training is to have adequate number of data samples. So, it is pivotal to pre-process the data before the training phase. The below mentioned techniques have been applied in the data pre-processing phase.

3.3.1 Stain Normalization

For the stain normalization of the images, the methodology proposed by Vahadane et al. (2016) has been used. The approach focuses on preserving the biological structures in the the tissues. The proposed method has three key steps in achieving the desired color normalization. Firstly it considers the pixels in the image with optical density greater than equal to zero. Secondly, it calculates the sparsity of the stains in the image. In the last step it tries to perform a soft classification for smoothening the borders between the biological features. The images in the dataset are H&E stained. Due to different staining standards, the color intensities of the stained images varies greatly. This color variation

[4]https://arxiv.org/abs/1912.12142

can be resolved by stain normalization which normalizes the color intensities in the histopathology images through color transfer; thus improving the quality of image (Iesmantas and Alzbutas; 2018; Khan et al.; 2019; Anupama et al.; 2019).



(a) (b)

Figure 3: (a): Image without stain normalization, (b): Stain normalized image

3.3.2 Image Augmentation

The total number of images for the three lung cancers counts to 15,000. After splitting there are 9,600 images for training, 3,000 for testing and 2,400 for validation. Keras provides a real time data augmenter, this allows the image loader to augment the image in real time and provide it for training. For the augmentation we will be using rotation, flip and zoom transformations on the images.

3.4 Modelling

In this section the modelling of the Capsule Network and the pre-trained models are described. The number labels in the workflow diagram (Figure 4) shows the phases in

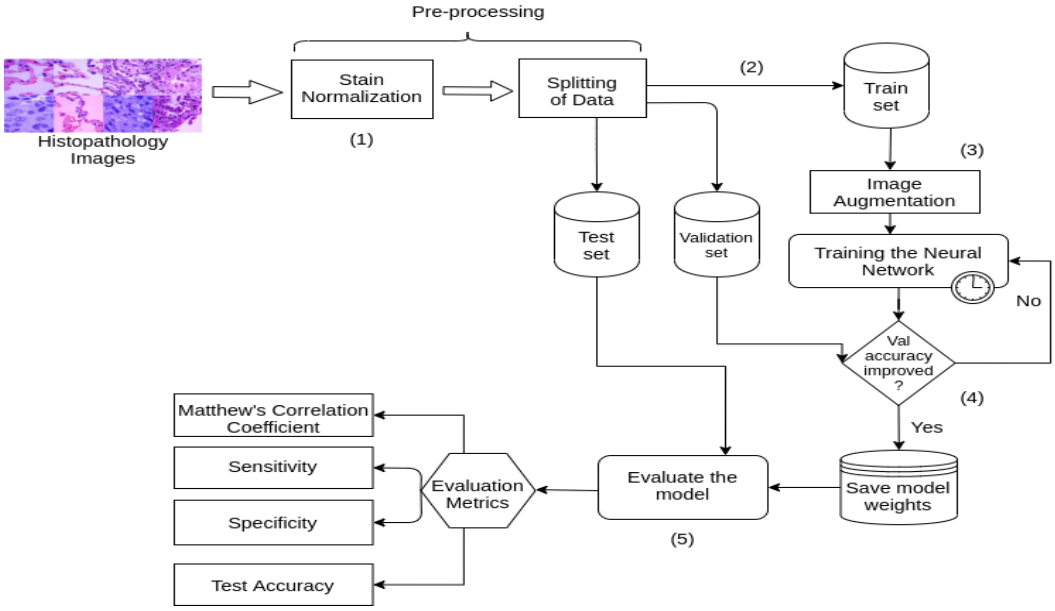


Figure 4: Modelling workflow diagram

the modelling of the neural network:

1. Firstly, the images from the data source are unzipped and stain normalization is applied.
2. In step two, the data set is split into training, testing and validation set.
3. After splitting, the images in the training set are made to go through real time augmentations.
4. In this step the training set and the validation set is used in training of the model where the model is evaluated using the validation image and the validation accuracy is assessed. If the validation accuracy of the model improves in the training epochs, the weights are saved.
5. After the completion of the training process the model is validated using the testing set images. Here, various evaluation metrics are applied to test the model performance.

4 Design Specification

4.1 Capsule Network

Capsule Networks are deep networks that extracts the features from the data and convert them into vectors. It is one of the promising concepts that has demonstrated several advantages in computer vision applications. For the first time Hinton et al. (2011) introduced Capsule Networks, these networks have the capability to determine pose, orientation and texture of the feature. The presence of a feature in a capsule can be determined by the output vector's length. These vectorized form of the features allows the Capsule Networks to determine the its presence even if the angle of view is altered. The feature information from the lower layers is passed on to the high layers if there exists any hierarchical dependency between the features, which is governed by the dynamic routing algorithm (Sabour et al.; 2017). The output vectors are applied to the squash function that shrinks the value between 0 and 1. A vector output value near to 1 is indicative for strong presence of a feature and vice-versa.
For the implementation of Capsule Network Keras/Tensorflow does not provide any dedicated API or library like CNN models. Being new to the field of research, Capsule Network has been applied to only a limited applications and have been evaluated with some of the benchmarking datasets. For the current study the architecture is based on (Rajasegaran et al.; 2019). The network is made up of Convolutional Capsule blocks and a Fully-connected Capsule layer. It also uses a novel dynamic routing algorithm based on 3D convolution that is less expensive in terms of computation cost.

The Figure 5 is the pictorial depiction of the implemented architecture where 5 Convolutional Capsule blocks are used with an image input size of 128 x 128 pixels.

- **Input Layer** : The input layer specified in the architecture accepts images with dimensions of 128 x 128.
- **Convolutional Capsule Layer** : In each block there are contains 3 Convolutional Capsule layer and a skip layer . The outputs of the Convolutional Capsule Layer are 4 dimensional vectors. The skip layer is used to tackle the vanishing gradient problem.

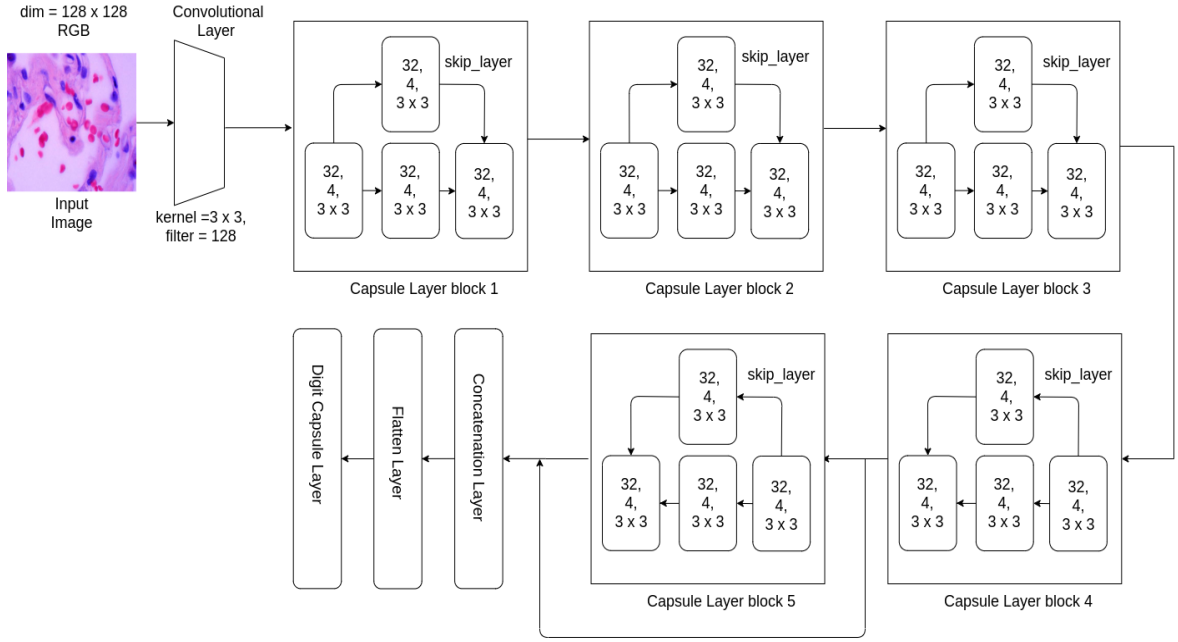


Figure 5: Capsule Network architecture

- **Flatten Capsule** : These layers are used to reshape the outputs of the Convolutional Capsule Layers before dynamic routing.
- **Digit Capsule Layer** : After the flatten layer the information from the capsules in the lower levels are routed to the class vectors using the Digit Capsule Layer thus encoding the image to a final capsule vector which is the output class.

4.2 Pre-trained Convolutional Models

For the study the following models trained on imagenet are used: VGG19, DenseNet-121 and Resnet-50. Keras provides a dedicated libraries for the models. Every pre-trained model have their own uniqueness in their architecture. These pre-trained models have been applied in lung cancer classification in previous research as discussed in Related Work.

- **VGG 19:** Simonyan and Zisserman (2015) designed the VGG19 model which comprises of multiple blocks of convolutional layers. In the entire network there are 19 weight layers i.e. 16 convolutional layers and 3 dense layers.
- **ResNet 50:** Unlike other sequential models the ResNet model introduces a new residual block (He et al.; 2016). The residual block in the ResNet has convolutional layers stacked together along with skip connection which is for identity mapping. This helps in the training performance, also avoid the vanishing or exploding gradient problem.
- **DenseNet 121:** Huang et al. (2017) proposed the DenseNet where the layers are grouped together into a block termed as dense block. In the architecture of the dense block each convolutional layers connected to the preceding layers allowing high amount of feature information traversal into the successive layers.

5 Implementation

The implementation task for the project is sub divided into multiple parts under which the detailed workflow is explained. The different subsections here would describe the execution of the phased explained in Modelling 3.4 .

5.1 Data pre-processing and splitting

The first phase of the implementation is data processing and preparation. The following steps were carried for the pre-processing:

5.1.1 Stain Normalization

1. In stain normalization a target image is selected form where the color information is extracted and applied to the other images. The target image is selected from the dataset by calculating the BRISQUE score of all the images. This is a no-reference image quality assessment technique (Mittal et al.; 2012). The image with the lowest score is the best quality[5]. Thus, the target images for the three type of lung cancer are chosen by evaluating the BRISQUE score of all the images for the respective classes and the one with the best score is considered as target for the respective class.
2. For the stain normalization process the python library staintools is required.
3. The proposed methodology for the stain normalization can be initiated by creating an object using staintools.StainNormalizer(method='vahadane').
4. The object of the StainNormalizer() calls the fit() function with the target image as argument for extracting the stain information of the target images which is then applied to all the other images. This step is repeated for the three lung cancer types.
5. In total there are 15,000 images considering all the classes of lung cancer. So the use of a loop was computationally expensive here. Hence, parallel execution was used for processing the images. It creates a pool of threads that select an images from the directory and calls the method stain_Normalize() which is a user defined function and executes the normalization process.
6. The stain normalized images are stored into a new directory named in accordance to their original class names.

5.1.2 Data splitting

In the second phase the data is split into train, test and validation. These sets will be later used in training and evaluation of the models. The detailed steps are mentioned below:

1. Using a loop the images names and the directory name i.e. the class label is extracted and then stored into a dataframe variable having two columns. The image names are stored in the **Image** column and the labels are stored in **CancerType**.
2. Once the dataframe is created, the rows in the dataframe are shuffled row-wise thus eliminating biases.

[5]https://in.mathworks.com/help/images/ref/brisque.html

3. The labels in the data frame are store as object . These values are transformed to binary labels using the sklearn's **LabelBinarizer**.
4. The **Dataset** dataframe is now split using the **train_test_split** function. The training and the test set are split in the ratio of 80:20 and stored in new **train** and **test** dataframes. For the validation set the training set is further split, where 20% of the training data is sliced off and stored in new dataframe **val**.
5. The train, test and validation sets are exported to .csv files to preserve the labels and the file names.
6. The image names are then fetch from the three dataframes created in step 4 and using a loop the image file are transferred to newly created testing, training and validation folders.
7. These labels in the .csv file and the images in the test, train and validation folder will be used for model training and evaluation.

5.2 Training

For the training steps of the capsule network and the pre-trained CNN models are described as following:

1. For Capsule Network, the images and the labels are loaded into numpy arrays. The images are reshaped to dimensions of 128 x 128 x 3 which is the input layer shape for the Capsule Network.
2. The pre-trained models are initialised using the keras.applications library which gives the provision for including weights of "imagenet". The three pre-trained models VGG19, ResNet50 and DenseNet121 are used for comparative study against Capsule Network.
3. The ImageDataGenerator from is used to generate batches of images which will be used for training. The batch size of 32 and 30 epochs have been considered for the training of the models.
4. The augmentation feature of the ImageDataGenerator, is used on the images dedicated for training of the model.
5. Callbacks like CSVLogger, ModelCheckpoint and LearningRateScheduler are used for logging the epoch output, save the weights if the loss/accuracy improves and decrease the learning rate which helps in fine tuning of the model respectively(Konar et al.; 2020).

6 Evaluation

To determine the model performance, it has to be validated on different evaluation metrics. In the current study the Capsule Network is evaluated with the performance of other pre-trained models. The models will be evaluated based on certain metrics to analyse their performance. The evaluations on the models would give an comparative analysis of Capsule Network against the state of art CNN architectures. The evaluations are based on the following metrics: Matthew's correlation coefficient (MCC), false negative rate, false positive rate, specificity, sensitivity,F1 score and accuracy which are deemed important in classification problems (Chicco and Jurman; 2020; Tharwat; 2018). For the study our focus is on the MCC score of the models as it is considered to be a better evaluation metric compared to accuracy, as the later only gives the overall performance of the

model which does not completely justify the classification potential of a model (Chicco and Jurman; 2020).

6.1 Evaluation 1 : Evaluating the performance of Capsule Network

In deep models to determine the case of over-fitting or under-fitting the learning curves must be analysed. The below Figure. 6 shows the training_validation accuracy and loss curves. By analysing the graphs we can infer that there is no evidence of over-fitting or

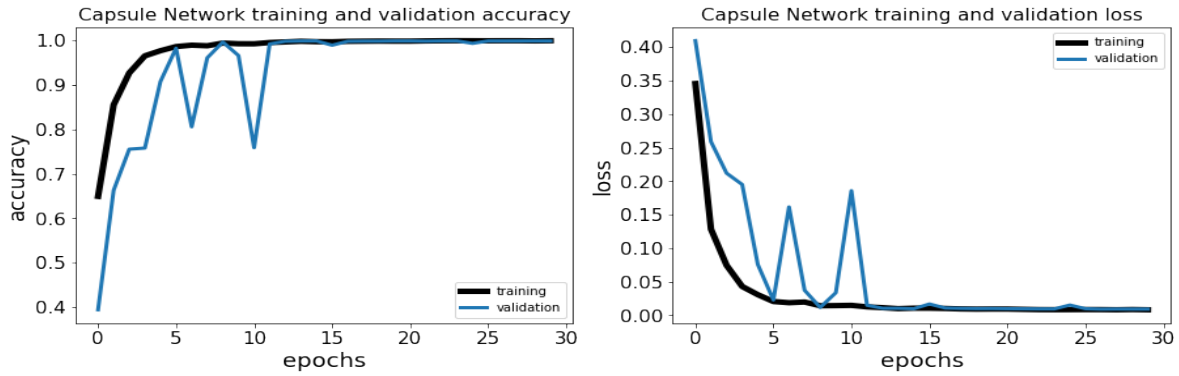


Figure 6: Learning curves of the Capsule Network

under-fitting. It is evident from the graphs that after epoch 10 both the training and validation accuracy and losses converge together and remains constant till the last epoch. The Capsule Network when evaluated gave validation accuracy of **99.9** and test accuracy of **99.8** with a MCC score of **0.99** which means there is high correlation between the actual and predicted values.

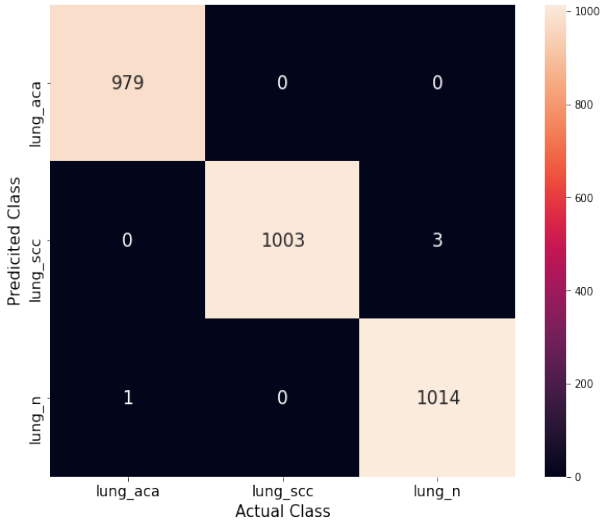


Figure 7: Confusion Matrix

The confusion matrix generated after evaluation in Figure. 7 shows the total number of correctly classified lung cancers and their corresponding incorrect predictions. There are total 4 miss-classifications out of 3000 sample of test data. Form the confusion matrix we

Table 2: Sensitivity and Specificity of the three classes of lung cancer

Lung Cancers	**Sensitivity**	**Specificity**	**False Positive Rate**	**False Negative Rate**
lung adenocarcinoma	1	0.99	0.00	0.00
lung squamous cell carcinoma	0.99	1	0.00	0.00
benign lung tissue	0.99	0.99	0.00	0.00

can calculate the Specificity and Sensitivity which is shown in Table 2. The calculated values of Specificity and Sensitivity for the lung cancers are close to 1 which indicate that the model has a very high true positive rate and true negative rate. Also, the low values of false positive rate and negative rate is indicative of the fact that there would be very minimal chance of a miss classification done by the Capsule Network. Therefore, the model was able to classify the lung cancers accurately and with very minimum error. This is important for any system in medical domain to give accurate results and generate less error which justifies the credibility of the system.

6.2 Evaluation 2: Estimating the Capsule Network performance with state of art CNN architectures

To evaluate the accuracy and validation of the predictions made by the Capsule Network the previously mentioned scoring metrics will be compared with CNN architectures.

The Test accuracy and MCC values have been computed for the Capsule Network and the pre-trained models and as shown in Table 3.

Table 3: MCC, Test and Validation Accuracy

Model	**MCC values**	**Test Accuracy**	**Validation Accuracy**
Capsule Network	0.99	0.99	0.99
DenseNet-121	0.99	0.99	0.99
ResNet50	0.93	0.95	0.93
VGG19	0.99	0.99	0.99

From Table 3 we can observe that Capsule Network was able to achieve good test and validation accuracy which is above **99%** and MCC score of **0.99** when compared with the pre-trained networks. The pre-trained networks are trained with the imagenet weights. Capsule Network on the other has used no pre-trained weights and was able to give comparable results against the pre-trained models which were leveraging the advantage of transfer learning. So, even not being trained with million images like the pre-trained models Capsule Network was able to give matching performance against the CNN models and in certain cases even better.

Table 4 shows the evaluation results of the models. Observing the F1 score, Specificity and Sensitivity values of the Capsule Network to the other models gives a clear picture that the Capsule Network was able to match the performance on par with the CNN models. In fact it was able to perform better than the ResNet model; but equivalent in the case of VGG19 and DenseNet. This shows the Capsule Network was able to match the performance against the state of art CNN architectures thus proving its competence.

Table 4: Evaluation of the Capsule Network with CNN architectures

Models	Sensitivity (lung adenocarcinoma)	Sensitivity (lung squamous cell carcinoma)	Sensitivity (benign lung tissue)	Specificity (lung adenocarcinoma)	Specificity (lung squamous cell carcinoma)	Specificity (benign lung tissue)	Weighted F1 score
Capsule Network	1	0.99	0.99	0.99	1	0.99	1
DenseNet-121	0.99	1	1	1	0.99	1	1
ResNet50	0.96	0.95	0.96	0.98	0.98	0.97	0.95
VGG19	1	0.99	0.99	0.99	1	1	1

6.3 Evaluating the performance with extended dataset

The LC25000 data used for the study contains histopathology images for colon cancer as well. To further test the performance of the Capsule Network two more classes of cancers were added. So, in total there are 5 types of cancers; 3 types for lung cancer and 2 for colon cancer. After pre-processing the colon cancer images the models were trained with the new data.

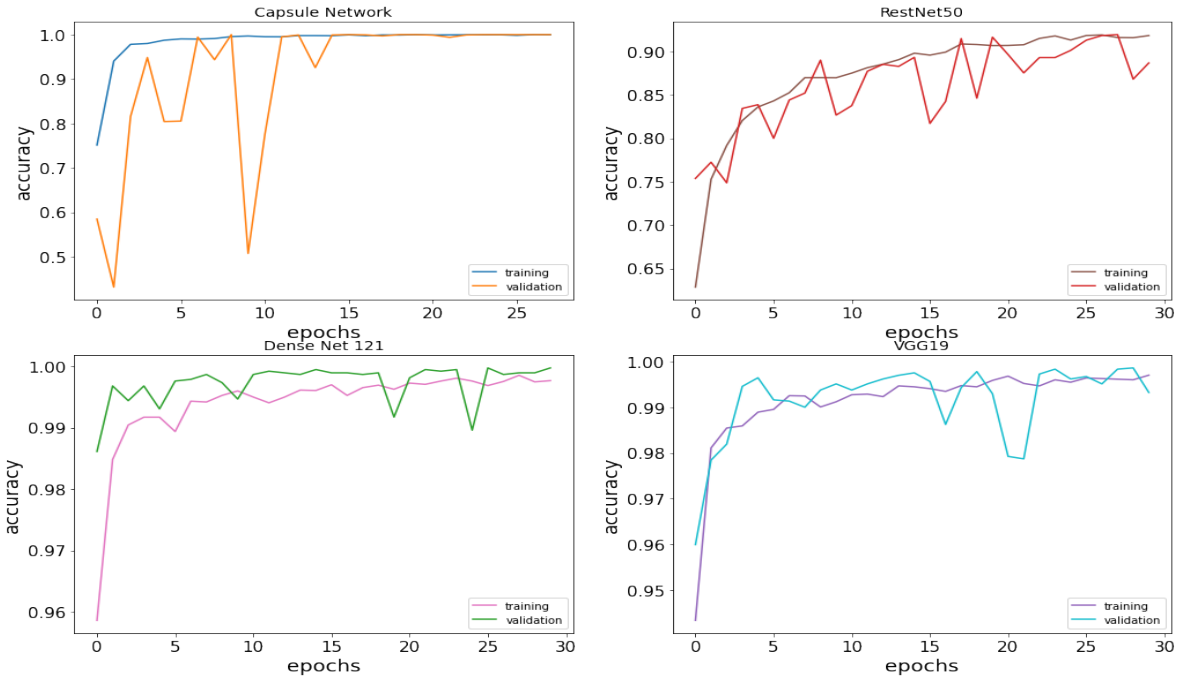


Figure 8: Learning Curves of the models

Figure 8 shows the learning curves of the Capsule Network, ResNet50, DenseNet121 and VGG19. The plots indicate that there is no case of over or under fitting in any of the models. The MCC score and testing accuracy were estimated shown in Table 5. The MCC score, test and validation accuracy of Capsule Network was better than the ResNet50 and VGG19 and equivalent against the DenseNet-121. Considering the MCC score it can be deduced that the prediction made by the Capsule Network have very high correlation as compared to VGG19 and ResNet50. This shows that even after introducing a new class cancer the performance of the Capsule Network remains unaffected, thus proving it capability in terms of data scalability.

Table 5: MCC and Test Accuracy with expanded dataset

Model	MCC values	Test Accuracy	Validation Accuracy
Capsule Network	1.0	1	1
DenseNet-121	1.0	1	0.99
ResNet50	0.89	0.88	0.88
VGG19	0.99	0.99	0.99

6.4 Discussion

From the above evaluations it is evident that Capsule Networks does possess the capability to perform equally against the pre-trained models and even better in certain cases. The evaluations the Capsule Network has shown remarkable results with very less error rate and good MCC score.

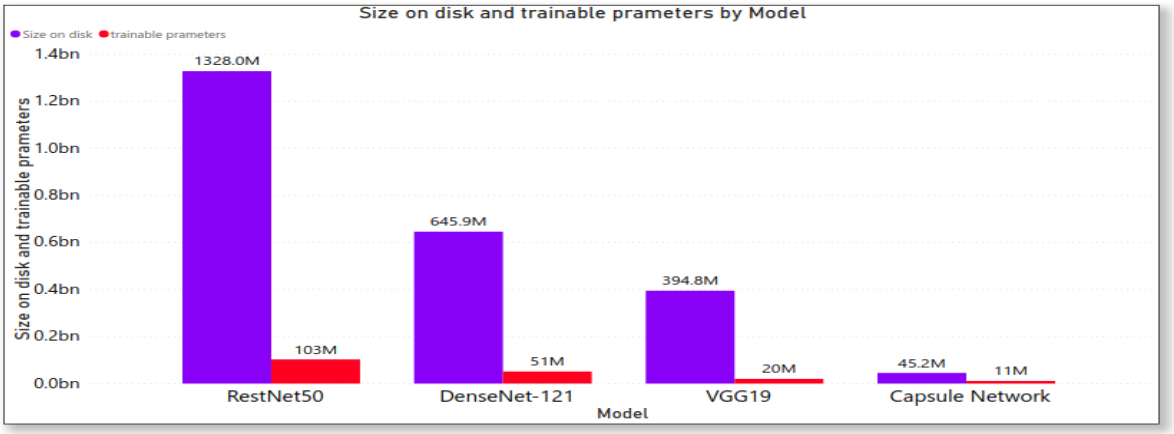


Figure 9: Trainable Parameter and size on disk

The Figure 9 shows comparison on the size of the model on disk after training and the total trainable parameters. After training the ResNet50 model was around 1.3 GB .h5 file, about 64 5MB for DenseNet-121 and 394 MB for VGG, whereas the Capsule Network was just 45 MB. Hence, CNN architectures consume more disk space as compared to Capsule network. Similarly, considering the trainable parameter Capsule Network has less number of trainable parameters i.e 11 million, against the CNN models which were 103 million, 51 million and 20 million for ResNet50, DenseNet-121 and VGG respectively. Thus, Capsule Network is light-weight in comparison to the CNN models and gives comparable performance against the CNN architectures in lung cancer classification.

7 Conclusion and Future Work

For cancer diagnosis automated detection and classification systems are safety-critical systems that have to demonstrate good performance in the detection and classification of the cancer types. It becomes easy for a doctor to identify the disease and provide proper medication and treatment. So, it is pivotal for the automated detection and classification system to have very low false positive and false negative rate which would help in building confidence on the system. The research focused on lung cancers classification from stained images using a Capsule Network. The model that was designed for the task, showed good performance with respect to the evaluations carried out. To prove

the Capsule Network's competence, three different pre-trained CNN architectures were trained with the same data using transfer learning. The comparative evaluation of the models showed that Capsule Networks was able to match the performance of state of art CNN architectures which are widely applied in medical imaging. The number of cancer type was also increased to test the model in case of diverse dataset, where it showed improved performance against the CNN models. Thus, concluding that Capsule Networks are able to classify lung cancers efficiently. In future, the application and the study can be extended to other kinds of aliments or larger set of cancers classes. Capsule Networks are at the initial stage of their application and its actual capability is yet to realised. So, there lies a bundle of scope in medical imaging to prove its potential. The study can be extended to CT scan or X-ray imaging problems, typically used in neurology and orthopedics.

References

ALzubi, J. A., Bharathikannan, B., Tanwar, S., Manikandan, R., Khanna, A. and Thaventhiran, C. (2019). Boosted neural network ensemble classification for lung cancer disease diagnosis, *Applied Soft Computing* **80**: 579 – 591.

Anupama, M. A., Sowmya, V. and Soman, K. P. (2019). Breast cancer classification using capsule network with preprocessed histology images, *2019 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0143–0147.

Baydilli, Y. Y. and Ümit Atila (2020). Classification of white blood cells using capsule networks, *Computerized Medical Imaging and Graphics* **80**: 101699.

Bonavita, I., Rafael-Palou, X., Ceresa, M., Piella, G., Ribas, V. and Ballester, M. A. G. (2020). Integration of convolutional neural networks for pulmonary nodule malignancy assessment in a lung cancer classification pipeline, *Computer Methods and Programs in Biomedicine* **185**: 105172.

Borkowski, A. A., Bui, M. M., Thomas, L. B., Wilson, C. P., DeLand, L. A. and Mastorides, S. M. (2019). Lung and colon cancer histopathological image dataset (lc25000).

Chicco, D. and Jurman, G. (2020). The advantages of the matthews correlation coefficient (MCC) over f1 score and accuracy in binary classification evaluation, *BMC Genomics* **21**(1).

Coccia, M. (2020). Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence, *Technology in Society* **60**: 101198.

Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y. W., Kwak, J. T. and Rajpoot, N. (2019). Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images, *Medical Image Analysis* **58**: 101563.

Gupta, K. and Chawla, N. (2020). Analysis of histopathological images for prediction of breast cancer using traditional classifiers with pre-trained cnn, *Procedia Computer Science* **167**: 878 – 889.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* .

Hekler, A., Utikal, J. S., Enk, A. H., Solass, W., Schmitt, M., Klode, J., Schadendorf, D., Sondermann, W., Franklin, C., Bestvater, F., Flaig, M. J., Krahl, D., von Kalle, C., Fröhling, S. and Brinker, T. J. (2019). Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images, *European Journal of Cancer* **118**: 91 – 96.

Henderson, D. and Klebe, S. (2006). TUMORS, BENIGN, *Encyclopedia of Respiratory Medicine*, Elsevier, pp. 312–320.

Hinton, G. E., Krizhevsky, A. and Wang, S. D. (2011). Transforming auto-encoders, *in* T. Honkela, W. Duch, M. Girolami and S. Kaski (eds), *Artificial Neural Networks and Machine Learning – ICANN 2011*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 44–51.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. (2017). Densely connected convolutional networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* .

Iesmantas, T. and Alzbutas, R. (2018). Convolutional capsule network for classification of breast cancer histology images, *Image Analysis and Recognition* p. 853–860.

Khan, S., Islam, N., Jan, Z., Din, I. U. and Rodrigues, J. J. P. C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, *Pattern Recognition Letters* **125**: 1 – 6.

Khosravi, P., Kazemi, E., Imielinski, M., Elemento, O. and Hajirasouliha, I. (2018). Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images, *EBioMedicine* **27**: 317 – 328.

Konar, J., Khandelwal, P. and Tripathi, R. (2020). Comparison of various learning rate scheduling techniques on convolutional neural network, *2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS)*, pp. 1–5.

Lee, A., Cho, Y., Jin, S. and Kim, N. (2020). Enhancement of surgical hand gesture recognition using a capsule network for a contactless interface in the operating room, *Computer Methods and Programs in Biomedicine* **190**: 105385.

Liang, G., Fan, W., Luo, H. and Zhu, X. (2020). The emerging roles of artificial intelligence in cancer drug development and precision therapy, *Biomedicine & Pharmacotherapy* **128**: 110255.

Liu, K. and Kang, G. (2017). Multiview convolutional neural networks for lung nodule classification, *International Journal Imaging System Technology* **27**(1): 12–22.

Lu, C. and Mandal, M. (2015). Automated analysis and diagnosis of skin melanoma on whole slide histopathological images, *Pattern Recognition* **48**(8): 2738 – 2750.

Mittal, A., Moorthy, A. K. and Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain, *IEEE Transactions on Image Processing* **21**(12): 4695–4708.

Mobiny, A. and Van Nguyen, H. (2018). Fast capsnet for lung cancer screening, *in* A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López and G. Fichtinger (eds), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Springer International Publishing, Cham, pp. 741–749.

Moitra, D. and Mandal, R. K. (2020). Classification of non-small cell lung cancer using one-dimensional convolutional neural network, *Expert Systems with Applications* **159**: 113564.

Rajasegaran, J., Jayasundara, V., Jayasekara, S., Jayasekara, H., Seneviratne, S. and Rodrigo, R. (2019). Deepcaps: Going deeper with capsule networks, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10717–10725.

Sabour, S., Frosst, N. and Hinton, G. E. (2017). Dynamic routing between capsules, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, p. 3859–3869.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition, *International Conference on Learning Representations*.

Sun, W., Zheng, B. and Qian, W. (2017). Automatic feature learning using multichannel roi based on deep structured algorithms for computerized lung cancer diagnosis, *Computers in Biology and Medicine* **89**: 530 – 539.

Teramoto, A., Yamada, A., Kiriyama, Y., Tsukamoto, T., Yan, K., Zhang, L., Imaizumi, K., Saito, K. and Fujita, H. (2019). Automated classification of benign and malignant cells from lung cytological images using deep convolutional neural network, *Informatics in Medicine Unlocked* **16**: 100205.

Tharwat, A. (2018). Classification assessment methods, *Applied Computing and Informatics* .

Toraman, S., Alakus, T. B. and Turkoglu, I. (2020). Convolutional capsnet: A novel artificial neural network approach to detect covid-19 disease from x-ray images using capsule networks, *Chaos, Solitons & Fractals* **140**: 110122.

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A. M., Esposito, I. and Navab, N. (2016). Structure-preserving color normalization and sparse stain separation for histological images, *IEEE Transactions on Medical Imaging* **35**(8): 1962–1971.

Vo, D. M., Nguyen, N.-Q. and Lee, S.-W. (2019). Classification of breast cancer histology images using incremental boosting convolution networks, *Information Sciences* **482**: 123 – 138.