Early Detection of Laryngeal Cancer using Multiple Instance Learning Based Neural Network

MSc Research Project
Data Analytics

Harshal Milind Tayade
Student ID: x18182763

School of Computing
National College of Ireland

Supervisor: Dr. Rashmi Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Harshal Milind Tayade
Student ID:	x18182763
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Rashmi Gupta
Submission Due Date:	28/09/2020
Project Title:	Early Detection of Laryngeal Cancer using Multiple Instance Learning Based Neural Network
Word Count:	6575
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:	Harshal Milind Tayade
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):	

Early Detection of Laryngeal Cancer using Multiple Instance Learning Based Neural Network

Harshal Milind Tayade
x18182763

Abstract

Laryngeal cancer is the most common type of head and neck cancer which affects the soft tissues of the larynx. Early stage detection of laryngeal cancer is crucial to avoid further medical complications and better patient care. The primary aim of this research is to provide computer-aided cancer diagnosis powered by deep learning mechanism. To achieve this we develop a novel Multiple Instance Learning (MIL) technique which classifies healthy and unhealthy/cancerous tissues. Further, we also incorporate the traditional Convolutional Neural Network (CNN) and transfer learning DenseNet121 model on our dataset for better comparison and evaluation of our research. The models are evaluated using standard metrics which are specific to biomedical domain. Our proposed MIL architecture produced outstanding results when compared to other models. It also outperformed other state-of-the-art MIL models used in solving medical domain problems. The results imply that the proposed technique is highly effective in detecting early signs of laryngeal cancer. The results prove that this proposed approach can assist medical professionals in early and accurate diagnostic of laryngeal cancer.

1 Introduction

The laryngeal tissues form the inner lining of the larynx which is made up of soft tissues. These tissues are formed by squamous cells. The soft tissue cells are thin and flat which help in breathing, voice generation and swallowing of food. As shown in the Figure 1 below the larynx is situated in the anterior compartment of the neck, precisely, in front of the food-pipe (pharynx) and above the windpipe (trachea). It is commonly known as the "voice-box". It virtually separates the digestive and breathing systems in the human body. Few of its vital functions are producing voice, allowing safe air flow through lungs and protect the windpipe during swallowing.

Laryngeal Cancer (LCA) is a type of head and neck cancer characterized by uncontrollable growth of malignant (cancerous) cells in the larynx. Most of the LCAs occur in the squamous cells and are termed as Squamous Cell Carcinoma (SCC). The Figure 1 below depicts the possible areas where the laryngeal cancer may develop and spread. Major proportion of LCA's develop in glottis (60%) which is the middle part of the larynx near the vocal cords as shown in the enlarged section of Figure 1. This is followed by supraglottis (35%) which is the area above the vocal cords including the epiglottis. Finally, subglottis (5%) which is lower part of the larynx is less affected. Although, high intake of alcohol and tobacco are the key factors for the increased risk of LCA, other contributing factors are age, gender, previous medical history and occupation.

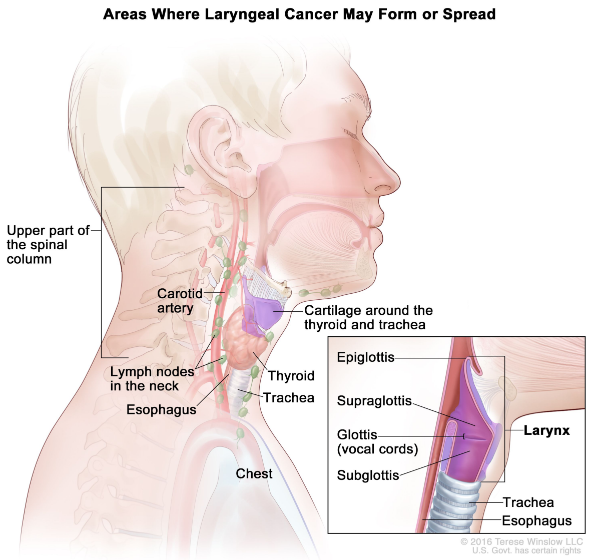

Areas Where Laryngeal Cancer May Form or Spread
Upper part of the spinal column
Carotid artery
Cartilage around the thyroid and trachea
Lymph nodes in the neck
Thyroid
Trachea
Esophagus
Chest
Epiglottis
Supraglottis
Glottis (vocal cords)
Subglottis
Larynx
Trachea
Esophagus
© 2016 Terese Winslow LLC
U.S. Govt. has certain rights

Figure 1: Body Parts Where Laryngeal Cancer May Form or Spread

According to NCBI (National Center for Biotechnology Information) more than 210,000 new cases of LCA have been diagnosed in 2017 [1]. The American Cancer Society has estimated that in the year 2020, around 12,000 people in the USA will be diagnosed with laryngeal cancer and nearly 3,700 will die [2]. Also, according to HSE, in Ireland, every year about 145 new cases of laryngeal cancer arise with higher probability in men [3]. Further, based on recent SEER statistics 12000 new LCA cases with 4000 deaths were reported in 2019 [4].

About 95% to 98% of the laryngeal cancers take the form of Squamous Cell Carcinoma. Based on the clinical literature study, early-stage detection of laryngeal cancer highly improves the survival rate of patients. The current gold standard technique includes histopathological diagnosis of tissues recovered in the biopsy process (Moccia et al.; 2018). However, this process is time-consuming, expensive, requires pre-operative preparation of patients and increased medical complications. These challenges have led to development of optical biopsy methods like Narrow Band Imaging (NBI) which is the current state-of-the-art for laryngeal cancer detection. Although NBI is used in endoscopic inspection there are few potential shortcomings like small mucosal changes not visible by human eye, alteration of mucosa vascular tree due to intraepithelial papillary capillary loops (IPCL) and longitudinal hypertrophic vessels (Moccia et al.; 2017). Also, NBI has a reduced sensitivity of 53% in detecting SCC if handled by inexperienced endoscopists. Further, use of such advanced techniques, require sufficient training time and highly skilled medical staff which is not affordable in under-developed countries or low-income areas. Furthermore, to avoid overlooking of cancer and natural anxiety, endoscopists often carry out unnecessary biopsy resulting in false cancer detection. This creates more financial and mental pressure on patients. Thus, it is crucial to examine the tissues for potential lesions and detect laryngeal cancer at an early stage. This will also avoid serious

[1] http://www.ncirl.ie

[2] https://www.cancer.org

[3] https://www.hse.ie

[4] https://seer.cancer.gov

medical complications like loss of vocal function and deterioration in quality of life.

The main motivation of this research is to overcome the above challenges like expensive diagnosis, need of expert medical supervision and mental stress of patients using deep learning methods. Application of deep learning in medical image classification has achieved significant progress in recent years. Deep learning models have been used in classification of skin cancer images, histologic gastric biopsy samples, radiation oncology, diabetic retinopathy, and classification of colorectal lesions.

Further in this paper we propose to use Multiple Instance Learning based Convolutional Neural Network (MIL-CNN) for classification of healthy tissue and other laryngeal abnormalities that can lead to laryngeal cancer (Quellec et al.; 2017). We have considered Ilse et al. (2018) as our state-of-the-art which classifies breast cancer and colon cancer images. Multiple instance Learning has been used in medical image screening and histopathological image classification (Bhattacharjee et al.; 2020). MIL was also used in melanoma detection and distinguish between common nevi images and melanoma (Astorino et al.; 2020). Similarly, a binary MIL model was adopted to detect breast cancer (Sudharshan et al.; 2019). Thus, inspired from the deep learning and MIL paradigm we would like to contribute to the machine learning community and biomedical sector by providing a better predicting model.This research will help medical experts in early detection of laryngeal cancer and improve the 5-year mortality rate of patients. Portable medical devices embedded with deep learning detection system can be assembled.

1.1 Research Question

"How well can a Multi-Instance Learning based Neural Network classify laryngeal tissues for early detection of laryngeal cancer?"

1.2 Research Objectives

Table 1: Research Objectives

Obj	Description	Parameters
1	Investigate current work done in the field of laryngeal cancer detection using machine learning and deep learning techniques	Model design, Issues, Results
2	Analyze the application of Multiple Instance Learning (MIL) in biomedical image classification, identify state-of-the-art and verify its feasibility to solve this research problem	state-of-the-art for cancer detection
3	Design and implement suitable MIL architecture.	Accuracy, Sensitivity, Specificity, F1-score
4	Design and implement Convolutional Neural Neural Network (CNN) from scratch	Accuracy, Sensitivity, Specificity, F1-score
5	Apply transfer learning technique using appropriate deep learning model.	Sensitivity, Specificity, F1-score
6	Analyze and compare the proposed MIL model with state-of-the-art and other implemented models	Accuracy, Sensitivity, Specificity, F1-score

Further the state-of-the-art Ilse et al. (2018) for histopathological image classification uses the BRATS dataset for breast cancer detection and 100 HE images for colon cancer detection. Each image was divided into patches called instances which were represented as a bag. An attention-based MIL pooling mechanism was used in the state-of-the-art which we intend to incorporate in our research. The MIL technique was suitable for our laryngeal tissue dataset as the the tissues itself are patches derived from original endoscopic image frames and thus can be considered as bags. Also, the state-of-the-art suggests of using the novel MIL technique in different kinds of cancer detection. These factors highly motivated us to perform a first hand research to classify laryngeal tissue images using the novel MIL technique.

Rest of the research is divided into different sections. section 1 provides the literature review of recent work carried out in detection of laryngeal cancer, areas where multiple instance learning is used and pre-processing carried out in medical imaging. The next section focuses on the proposed methodology for carrying out the research including model building, implementation and evaluation. The last section provides conclusion, limitation and future scope.

2 Related Work

Alzheimer's disease detection was carried out using multiple instance learning (MIL). Local intensity image patches were considered as key features. MIL solved the problem of ambiguous training labels. The classification accuracy obtained was about 89%. It performed better than state-of-the-art linear SVM model (Tong et al.; 2014). Multiple instance learning model was used to classify mild and severe cerebral small vessel (CVD) disease. The model extracted intensity patches from regions with cancer probability. The model achieved accuracy of 75%. The model results were better than eight machine learning models (Chen et al.; 2015). A convolutional neural network based multiple instance learning architecture was used to solve multi-task problem. This MIL based CNN benefited using transfer learning approach. The proposed model outperformed benchmark models like pretrained VGG (Zeng and Ji; 2015).

Texture-based machine learning algorithms were used to classify laryngeal images for early detection of laryngeal cancer. This was a four-class classification problem with disease classes as tissue with IPCL-like vessels, leukoplakia, tissue with hypertrophic vessels, and healthy tissue (Moccia et al.; 2017). A multiple instance based deep neural network was applied on two real-world medical datasets for natural image classification and pathological lung cancer detection. Each image was considered as bag of multiple instances. The MIL model worked on pooling/voting strategy. The proposed model outperformed SVM, CNN and their combinations (Quellec et al.; 2017).

A system to select informative frames from endoscopic video to assist better prediction of laryngeal cancer was developed. The feature set comprised of features from existing state-of-the-art used in medical domain and newly trained features. A binary SVM classifier with one vs one scheme was used for classification. 3-fold cross validation was performed with recall of 91% (Moccia et al.; 2018). Squamous Cell Carcinoma (SCC) of the vocal cords was detected using the Confocal Laser Endomicroscopy (CLE) images. A novel patch-based convolution network was proposed for CLE image classification. This method was based on pretrained inception V3 network (Aubreville et al.; 2018).

A computer aided diagnostic to detect laryngeal cancer was developed using Deep Convolutional Neural Network (DCNN). The system classified laryngoscopic images into three classes that are normal tissues (NORM), benign laryngeal tumours (BLT) and precancerous laryngeal lesions (PRELCA). GoogLeNet Inception v3 transfer learning was applied. The model achieved sensitivity of 73%, AUC of 92% and specificity of 92% (Xiong et al.; 2019). Learned and handcrafted features were used to design an early laryngeal squamous cell carcinoma (SCC) system. Texture-based descriptors (Local Binary patterns) and deep learning-based descriptors were used. Pretrained convolutional neural networks including ResNet v2, Inception-v4 and Inception-ResNet v2 were adopted. The texture-based feature classification achieved recall of 94%. Whereas, the CNN architectures achieved recall between 93% to 95% (Araújo et al.; 2019). Artificial intelligence was used to diagnose oesophageal cancer (Squamous Cell Carcinoma) and adenocarcinoma. The AI-based system was developed using Single Shot Multi-Box Detector neural network architecture. The sensitivity of the AI-based model was 98% (Horie et al.; 2019). Transfer learning approach was used to classify laryngoscopic video frames for early detection of laryngeal cancer. Inception v3, ResNet-50 and SqueezeNet state-of-the-art transfer learning architectures were explored. All the three models were pre-trained on ImageNet dataset and then fine tuned for the video frames data. It was observed that SqueezeNet performed relatively better (Galdran et al.; 2019). Availability of high-power computers and GPUs have benefited the use of deep learning methods to analyse and classify medical images. A Region-based Fully Convolutional Network (R-FCN) was used to detect the larynx and classify laryngeal lesions. Multi-organ target images were used to classify the lesions using target detection technique (Luan et al.; 2019). A convolutional neural network architecture was proposed to improve the traditional clinical radiomic framework for detecting head and neck squamous cell carcinoma. The study was based on training deep CNN model from scratch (Ma et al.; 2019). An auto encoder based neural network was proposed to classify benign and malignant tumours using the hyperspectral laryngeal images. The network achieved sensitivity of 92.32% and specificity of 91.31%. This method can be used in non-invasive devices to detect head and neck cancers (Ma et al.; 2019). A fully convolutional neural network (FCN) was used to classify healthy epithelium from cancerous head and neck lesions. Semantic segmentation with FCN was analysed which achieved average recognition rate of 86.70% (Rodner et al.; 2019).

Accurate detection and segmentation of laryngeal leukoplakia was performed using a novel multi-scale recurrent fully convolutional neural network. The design consisted of multi-scale input layer, double U-shaped network, and a side-output layer (Ji et al.; 2020). A combined architecture comprising of Deep Boltzmann Machine (DBM) and Support Vector Machine (SVM) was proposed to classify oral and neck cells using hyperspectral images.Minimum spanning tree was used to extract high-level features. The model achieved accuracy between 91.55% and 94.75% (Jeyaraj et al.; 2020). Multiple instance learning was used to classify weakly supervised data. 16 state-of-the-art MIL models are investigated based on their performance. MIL paired with SVM, KNN and GRAPH based networks are studied and found suitable to be used for classification tasks (Bhattacharjee et al.; 2020).

Artificial neural network based genetic algorithm was used to predict laryngeal cancer. The genetic model was combined with neural network for better performance. The precision obtained was between 80% to 84%. Transfer learning was used to solve biomedical domain problems. DenseNet121 pre-trained model was used for detection of lung cancer. The model achieved mean sensitivity, specificity and accuracy of around 74% (Ausawalaithong et al.; 2018). Similar work was performed to detect skin cancer using DensNet121 where it achieved validation accuracy of 90% (Rahi et al.; 2019). DenseNet121 transfer learning approach was also implemented in two-step skin cancer classification. The accuracy obtained in both the steps was 85% and 80% respectively (Moldovan; 2019).

The table 2 and table 3 provide summary of few important papers which discuss objectives, the design approach, techniques implemented and key findings in laryngeal cancer detection and multiple instance learning domain.

Table 2: Summary about detection of laryngeal/head and neck cancer related work using deep learning and machine learning techniques

Author(s)	**Objectives**	**Research Design**	**Keywords**	**Findings**
Xiong et al. (2019)	Investigate use of texture-based descriptors and deep-learning based descriptors	Gaussian smoothing for pre-processing, Use of pre-trained networks - ResNet and Inception v4	Tissue classification, texture analysis, CNN	The transfer learning approach achieved sensitivty of 73%, specificity of 92% and AUC of 92%.
Araújo et al. (2019)	Classify laryngoscopic images into normal, benign and pre-cancerous tissues.	Transfer learning using GoogleNet, Inception v3 and using saliency maps for model interpretation	Deep CNN, transfer learning, saliency map	The Deep learning approach outperformed machine learning methods with recall between 93% to 95%.
Horie et al. (2019)	To differentiate between superficial and advanced SCC.	Use of single shot multi-detector AI architecture with 0.0001 learning rate	Artificial intelligence, SCC, esophageal cancer	The model achieved sensitivity of 98%. Unable to detect inflammed mucosa cancer.
Moccia et al. (2017)	To classify laryngeal tissue images into four classes - healthy, leukoplakia, hypertropic vessels and IPCL vessels.	Anisotropic filtering, feature extraction using LBP and GLCM, multi-class SVM for final classification	LBP, GLCM, tissue classification, SVM	Performed better than K-means and naive-bayes with classification recall of 93%.
Galdran et al. (2019)	Train three state-of-art models using optimized weights.	ResNet-50, Inception v3 and Squeeze-Net models were analyzed	larygoscopy, CNN, real-time classification	Squeeze-Net performed relatively better precision 94%, recall 95% and F1-score 98%.
Luan et al. (2019)	Detect lesions in larynx using high computation power, investigate if R-FCN can confirm region of interest.	The R-FCN design consists of CNN, region proposed network, ROI pooling and position sensitive prediction layer	R-FCN, ROI, laryngeal lesions	R-FCN achieved precision of 72.03% with IoU set at 0.3%. The future work focuses on increasing classification precision by using improved IoU.
Diamant et al. (2019)	To investigate if CNN improves head and neck cancer detection using CT-images.	The CNN was based on convolution, non-linearity, pooling and classification with kernel size of 5 X 5 with ReLu activation.	CT-images, CNN, deep learning	The deep-CNN achieved sensitivity of 86% and specificity of 89%.
Aubreville et al. (2018)	To automate classification of CLE images and analyse patch-based classification for better performance.	Pre-processing include flipping, cropping and polar transformation. Pretrained inception v3 network was designed. Individual patch-based and image-level classification was analysed.	GPU, inception v3, CLE images	Both methods performed well with AUC between 94% and 95%.
Rodner et al. (2019)	Differentiate between healthy epithelium and cancerous lesions in the head and neck region.	Standard histopathlogical images were taken as input. Segmentation and detection was carried out using FCN	image segmentation, FCN, head and neck cancer	FCN was trained on 114 images with classification accuracy of 86.70%.
Ma et al. (2019)	Adaptive classification of bening and cancerous tissues.	Wavelength bands are used for training of autoencoder networks. Pixel-wise image information is extracted for classification purpose. The network improves prediction adaptively.	tissue classification, head and neck cancer, autoencoder	The model achieved specificity of 91.31% and sensitivity of 92.32%.

Table 3: Summary about detection of laryngeal/head and neck cancer related work using deep learning and machine learning techniques

Author(s)	**Objectives**	**Research Design**	**Keywords**	**Findings**
Ji et al. (2020)	To analyse if the novel boldface-M-Net can segment and detect leukoplakia lesions.	The BM-Net design consists of multi-scale input layer, double U-shaped network, and a side-output layer.The convolutions are modified with combined multi-scale U-Net and recurrent convolutions.	U-Net, segmentation, leukoplakia, Recurrent neural network	The BM-Net outperformed FCN, U-Net, M-Net and ensembles. An accuracy of 99%, sensitivity of 25% and specficty of 99.50% were obtained.
Jeyaraj et al. (2020)	Classify cancerous tissue and normal tissues of the oral region.	A combination of Deep Boltzmann Machine (DBM) and SVM classification is considered. Cancer detection is done using background mix-pixels.	Tissue classification, oral cancer, Deep Boltzmann Machine, SVM	Classification accuracy of 94.75% was observed.
Quellec et al. (2017)	Investigate the performance of deep learning based MIL model in medical and natural image classification.	Each image is considered as multi-instance bag. Instance prediction is based on voting strategy. Novel global contrast pooling is used for better prediction.	Lung cancer, MIL, deep learning, image classification	MIL model achieved accuracy, precision and recall of 99.60%.
Tong et al. (2014)	To solve ambiguous training image labels and improve predictions.	Each image is represented using graph which shows relationship of patches. Leave one-out cross validation is performed.	Alzheimer disease, MIL, graphs, MRI images	The classification accuracy of MIL was 89% based on 834 images.
Chen et al. (2015)	Classify absent/mild CVD and moderate/severe CVD using MIL.	Extraction of intensity patches from cancerous regions of the image. The patches act as input to MIL model	Cerebral small vessel disease, MIL, patches, image classification	The MIL model accuracy was found to be 75%. MIL outperforms other machine learning models.
Zeng and Ji (2015)	To solve the ambiguous problem between labels and images.	Multi-task images are taken as input. The images are connected with shared-CNN model. The transfer learning based CNN helps in better prediction with MIL	MIL, multi-task learning, bioinformatics	The CNN-based MIL performed better than pretrained VGG. Higher performance was achieved with less features.
Bhattacharjee et al. (2020)	To investigate if weakly supervised MIL method is suitable for biomedical domain analysis.	A novel genetic pooling algorithm based on multiple instance learning and genetic algorithm is applied. Bag-level losses are minimized. Bag level pooling layer is replaced with random initialization of weights.	MIL, bioinformatics, genetic algorithm	The method achieved accuracy between 93% and 94%.

Based on analysis of the literature review it was understood that machine learning techniques were relatively less effective than deep learning methods. However, majority of the deep learning methods have focused on using transfer learning models to solve the laryngeal cancer detection problem. This creates a research gap to investigate other novel architectures that have the ability to predict effectively than the traditional transfer learning approach. On the other hand, the multiple instance learning has gained much popularity in solving image classification problem of biomedical domain. This highly motivated us to incorporate a multiple instance learning based neural network approach in our research problem.

3 Methodology

This research is related to biomedical domain and after detailed analysis it was found that The Cross Industry Standard Process for Data mining (CRISP-DM) is most suitable for this research project

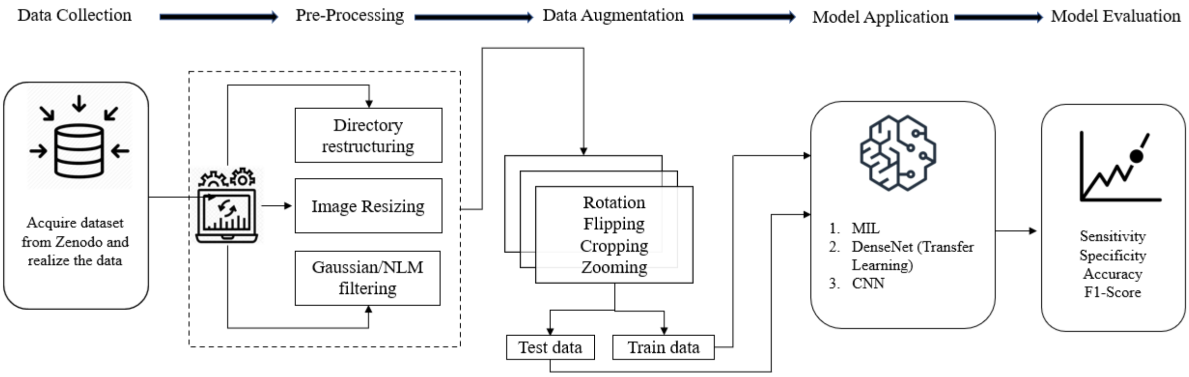


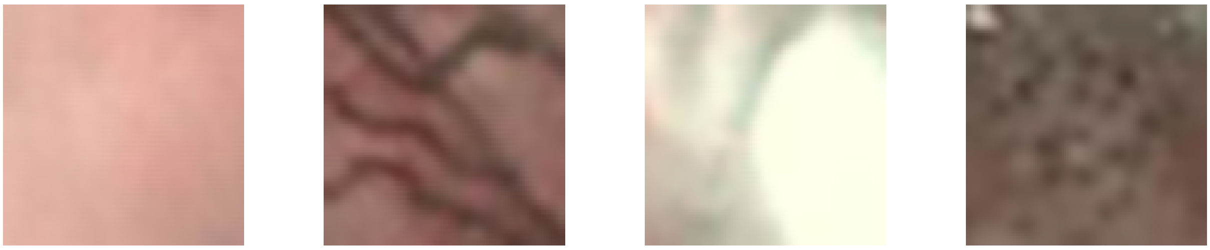
Figure 2: Process flow of research method

3.1 Project Understanding

The present gold standard methods used in screening and diagnosis of laryngeal cancer are tissue biopsy and visual endoscopy. However, these processes have drawbacks in terms of operational cost, time and need of expert supervision. Also, the state-of-the-art machine learning models depend on manual selection of feature set by medical professionals and lack self-learning ability. On the contrary, deep learning methods equipped with memory efficient and adaptive capabilities can be leveraged in early detection of laryngeal cancer. Moreover, embedded devices powered with deep learning models can prove advantageous in countries where medical experts are not available. These portable devices can provide cost-effective and real-time diagnosis. This will help the biomedical domain in assisting doctors with effective and accurate predictions of laryngeal cancer at an early stage. Therefore, this research aims at developing a Multiple-Instance Learning based deep learning method for classification and early detection of laryngeal cancer.

3.2 Dataset Description

The data in this research is sourced from Zenodo which is owned by researches that host datasets for use under their nascent open data policy. The laryngeal dataset is formed by collecting healthy and early stage laryngeal cancerous tissues. The tissue samples are derived in the form of patches with size 100 X 100 pixels. The patch extraction process in manually done by using 33 laryngoscopic images obtained from Narrow Band Imaging (NBI). These images were obtained from 33 patients who suffered from laryngeal squamous cell carcinoma. Further, the dataset contains four tissue classes as shown in 3 namely, He (healthy tissue), Le (tissue with leukoplakia), Hbv (tissue with hypertrophic blood vessels) and IPCL (tissue with interpapillary capillary loops). The dataset contains total of 1320 tissue images. The dataset is balanced with 330 images per class.



(a) He (b) Hbv (c) Le (d) IPCL

Figure 3: Laryngeal Tissue Classes - Moccia et al. (2017)

3.3 Pre-Processing

3.3.1 Noise Reduction

For better efficiency in image analysis, high quality of images is a vital parameter. The main processing in deriving high quality images include noise reduction, contrast/brightness correction, sharpening etc. In case of endoscopic images, noise reduction plays a crucial role because noise is the most common issue in processing of endoscopic images. Lack of brightness in the endoscopic video frames is the primary reason behind the noise (Obukhova et al.; 2018).

As our research contains endoscopic images of the laryngeal tract we investigated different methods discussed by Obukhova et al. (2018) The methods we analysed include edge-preserving, denoising and smoothing algorithms like Gaussian filter, Median filter, total variation filter, wavelet filter, non-local means filter (NLM). We also performed image sharpening and brightness adjustment to obtain more clarity than the raw images. The effectiveness of each filter was assessed using Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) as mentioned in (Pedersen et al.; 2017). It was observed that Gaussian filter provided better quality score and hence was used for image denoising in this research. The detailed overview of filtering is explained in section. 5

3.3.2 Gaussian Filter

Gaussian blur filter is a low pass image smoothing filter which preserves the edge-related information. It minimizes the noise by preserving low spatial frequency and eliminating negligible details in the image. It convolves the image using a specific kernel size represented by equation 1 where, σ is standard deviation and x and y represent local indices (Umapathi and Narayanan; 2014). In this research the images were filtered using 3 x 3 kernel size.

$$G(x, y, \sigma) \;=\; \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2\;+\;y^2}{2\pi^2}} \tag{1}$$

(a) BRISQUE Score - 110

(b) BRISQUE Score -71

Figure 4: Gaussian Filter applied to (a) raw image to generate (b) denoised image

For better robustness and generalization of the model the data needs to be augmented with sufficient training data. Based on the work carried out by Taylor and Nitschke (2018) we adopted various geometric transformations which include rotation, zooming and flipping.

3.4 Modelling

3.4.1 Baseline 1 - Convolutional Neural Network

Deep learning can facilitate such discovery by learning simple points and then building more complex layers over it. The Convolutional Neural Network is one such deep learning technique which can be used for image classification, segmentation, and registration. CNNs learn by using spatial and image configuration information which are vital informative parameters in biomedical images.

3.4.2 Baseline 2 - Transfer Learning

The main idea behind transfer learning is to use a predefined architecture based on natural image data like IMAGENET along with pretrained weights. The weights are then fine-tuned using the images from our dataset. This procedure is universally adopted in different biomedical areas. Inspired from Ma et al. (2019); Ausawalaithong et al. (2018); Rahi et al. (2019) we have adopted DenseNet121 model trained on IMAGENET.

3.4.3 Proposed Method - Multiple Instance learning

Multiple instance learning is characterized as weakly supervised algorithm where the data is arranged in sets. These sets are commonly known as bags and the data within the bags are referred as instances. During the training phase only the bag labels are provided and not the individual labels of the instances. A bag can be referred as an image. The primary assumption of the MIL problem is to consider that all negative bags contain only negative instances and there is at least one positive instance in the positive bags. This concept is demonstrated in Figure 5.

Thus, we convert our problem into a binary classification problem with two classes as 0 - Healthy tissues and 1 - pre-cancerous tissues which can lead to laryngeal cancer. The proposed MIL model will be able to predict if the tissue sample is normal or an early sign

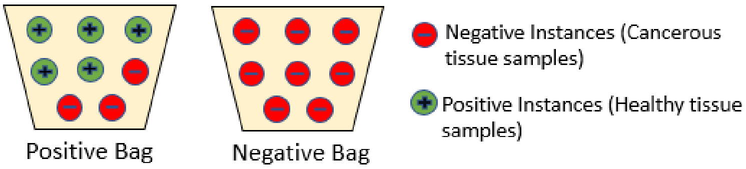

Positive Bag
Negative Bag
Negative Instances (Cancerous tissue samples)
Positive Instances (Healthy tissue samples)

Figure 5: Basic Idea Behind Multiple Instance Learning

of laryngeal cancer. The early signs include leukoplakia (Le), tissues with hypertrophic blood vessels (Hbv) and tissues with Interpapillary capillary loops (IPCL).

4 Design Specification

In this research we have adopted a novel multiple instance learning (MIL) technique. This sections gives detailed overview of the proposed MIL method

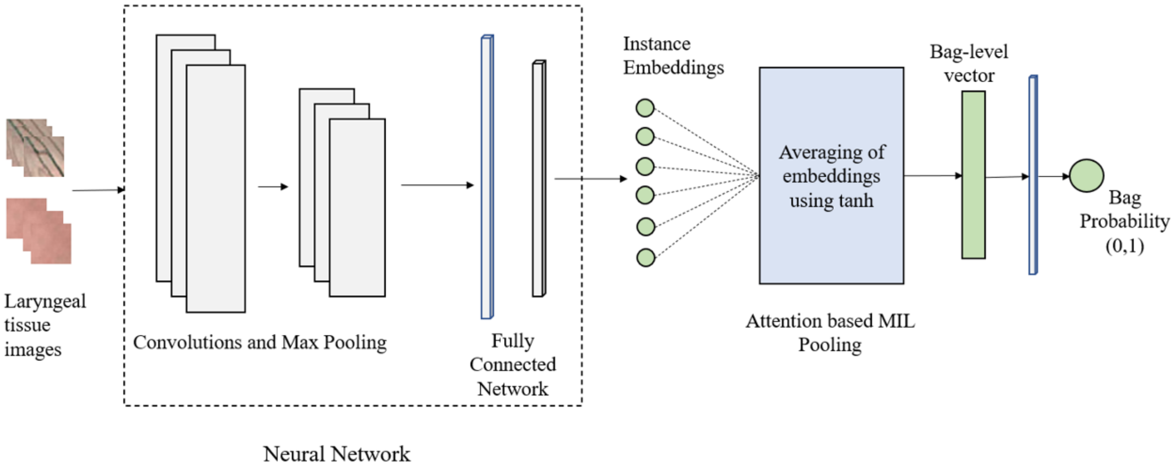


Figure 6: Attention-based MIL Architecture

The proposed MIL architecture shown in figure 6 learns the bag or image label based on its Bernoulli distribution. The convolutional neural networks configuration is used to fully parameterize bag label probability $\Theta(X)$. To classify a bag (image) of instances the MIL model follows three steps – (i) transform the instances using some function f. (ii) Combine these transformed instances using a permutation invariant function sigma. (iii) Transform these combined instances by using function g. The selection of operators for f, g, sigma depends on type of approach adopted. In this research we have used the embedded approach to design the attention-based MIL architecture which provides individual bag probabilities independent of the number of instances. The embedded approach is preferred because as compared to instance level approach it does not introduce any biases. This benefits the final classification phase to correctly differentiate between cancerous and healthy bags (Ilse et al.; 2018).

4.1 Neural Networks

The image tensors are fed to the neural networks which is a set of 2D convolutions and max pooling layers. The goal of neural network is to transform and parameterize the instances. Precisely, they convert the features or instances into low dimensional embedding. It performs the task stated in step (i). Advantage of using neural network for parameterization is it makes this architecture flexible which can be trained using backpropagation (Ilse et al.; 2018).

4.2 Attention Mechanism

The attention mechanism is a MIL pooling operator based on embedded level approach. It is highly adaptive and flexible than other pooling operators discussed in (Ilse et al.; 2018). It calculates weighted average of the instance weights obtained from the neural networks. Alternatively, the attention layer aggregates the individual instance scores or features from previous layers to represent a bag level score. For better gradient descent a hyperbolic tangent tanh function is used.

5 Implementation

The implementation section provides a detailed overview of approach and techniques used in transforming, pre-processing and model building. It also highlights the technology stack and other important features of the model building phase.

5.1 Data Transformation

The data was sourced in a zip file with a complex directory structure. The first phase of transformation comprised of arranging the images into four directories represented by the class labels. The next step involved, arranging all the image data into two folders which represent healthy and cancerous tissues for binary classification. To achieve these transformations, we used standard python libraries including os, shutil and tarfile.

5.2 Data Pre-processing

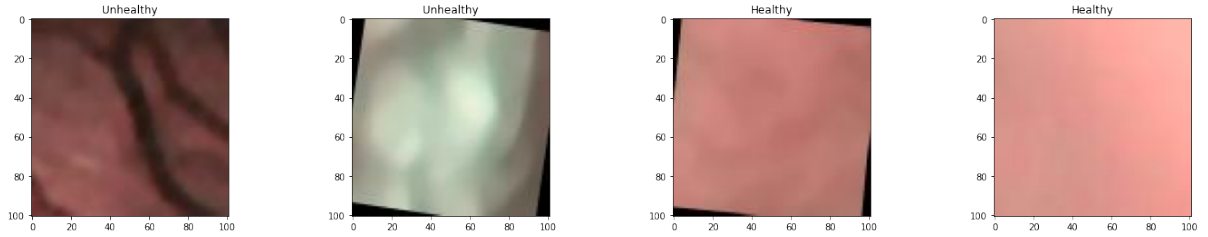
The first stage involved resizing the images from the dataset to a unique size. The original images were in 101 x 101 and 102 x 102 dimensions. So, we resized the images to a closest dimension value of 100 x 100 ensuring that minimum information loss occurs in the resized images. The next step involved deriving better quality images. The dataset was of endoscopic images and hence contained noise. We investigated different pre-processing techniques explained in section 5.2. The final stage involved, denoising the images using Gaussian filter. The images were visualized using matplotlib and pylab image libraries in python. The filter codes were readily obtained from Sci-kit image [5] and OpenCV python

[5] https://scipy-lectures.org/packages/scikit-image/index.html

libraries of computer vision [6]. Few helper functions from standard libraries including PIL, numpy, os and shutil were also used.

5.3 Data Augmentation

The Augmentor python library was used to perform data augmentation. The augmentation pipeline was constructed using pre-defined library functions like rotate, zoom, top-bottom flip and left-right flip. In image rotation the probability value was set to 0.7 along with along with left and right rotation values set to 10 degrees. For image zooming the probability was 0.5 and the zooming factor value ranged from minimum of 1.1 to maximum of 1.5. Lastly, for all the flipping operations the probability was set to 0.5. The original dataset contained total of 1350 images distributed across 4 classes. We augmented this dataset by setting the sample value to 3000 for each class.



(a) Flip (b) Rotate Right (c) Rotate Left (d) Zoom

Figure 7: Data Augmentation using Zooming, Flipping and Rotation

5.4 Convolutional Neural Network

We have designed and trained a convolutional neural network without using pre-trained weights. It is a sequential model comprising of 2D convolutional layers, dense layers and pooling operator. The model is compiled using SGD optimizer with learning rate set to 0.001. For all the convolutional blocks we have used ReLu activation function [7]. The model was implemented using functions from the keras deep learning library for python. The detailed design configuration of our CNN model is shown in table 4

Table 4: CNN model configuration

Blocks	**Configuration**
Block 1	Convolutional layer of size (3X3). Number of kernels = 32. Max Pooling size of 2x2 with Dropout of 0.5.
Block 2	Convolutional layer of size (3X3). Number of kernels = 64. Max Pooling size of 2x2 with Dropout of 0.5.
Block 3	Convolutional layer of size (3X3). Number of kernels = 128. Max Pooling size of 8x8.
Dense Layer	Size = 2
Output	Softmax classifier with 2 outputs

[6] https://pypi.org/project/opencv-python/

[7] https://www.tensorflow.org/tutorials/images/cnn

5.5 Transfer Learning

For transfer learning we implemented DenseNet-121 on our dataset. The model architecure was based on (Huang et al.; 2017). Pre-trained weights were used by training the model on ”ImageNet” data - ILSVRC. Preliminary model configurations were sourced from keras api documentation [8]. The parameters were fine tuned based on our dataset using input dimension of 100X100. We have used average pooling operator and ”SGD” optimizer with learning rate set to 0.001.

5.6 Attention-based Multiple Instance Learning

The first stage is loading the images for binary classification. This is achieved by a custom function which loads the images from two folders labelled as ”healthy” and ”cancerous”. We have implemented 3-fold cross validation for better model generalization. Hence, the custom function randomly splits the images into train bag and test bag for each split. The next step involves batch creation for model training. Here we read each image with its label using cv2 python library, resize the dimension as 64x64 and store normalized the images into a batch. The last dense layer in table 5 generates weighted features which are then used by attention-based MIL pooling layer Wang et al. (2018) to for averaging and generating output.

Table 5: MIL model configuration

Blocks	**Configuration**
Input Layer	Image dimension = 64x64. datatype = float32.
Block 1	Convolutional layer of size (4X4). Number of kernels = 36. Activation function = ReLu. Max Pooling size of 2x2.
Block 2	Convolutional layer of size (3X3). Number of kernels = 48. Max Pooling size of 2x2.
Dense Layer	Size is 512 with Dropout of 0.5. Activation function = ReLu
Dense Layer	Size is 512 with Dropout of 0.5. Activation function = ReLu
Attention MIL Pooling Layer	Size is 128.
Output Layer	Sigmoid.

6 Evaluation

We investigated the performance of our novel attention-based MIL model to classify healthy and cancerous tissues. We compared our MIL approach with transfer learning models and CNN without using pretrained weights. Extensive evaluation was carried out using various standard metrics along with sensitivity and specificity which are more significant in the biomedical domain.

[8] https://keras.io/api/applications/densenet/

6.1 Baseline Approach 1: Detection of Laryngeal Cancer using Convolutional Neural Network

We have trained our CNN model on the laryngeal dataset without using any pre-trained weights. Training was performed using 15840 images and 1980 images were reserved for validation. The model parameters were calculated based on batch size, train and validation set. The model achieved train accuracy of 0.88 and validation accuracy of 0.89 observed in figure 8. The test accuracy was 0.89 which is comparable to test accuracy and rules out chances of model overfitting. Further model evaluation is shown in table 6.

Table 6: CNN model evaluation

Test Accuracy	Validation Accuracy	Sensitivity	Specificity	F1-Score
0.89	0.89	0.94	0.84	0.90

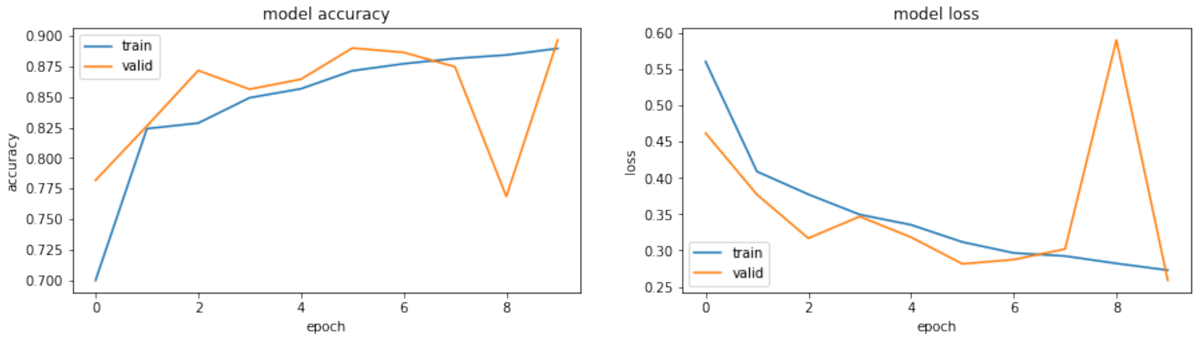
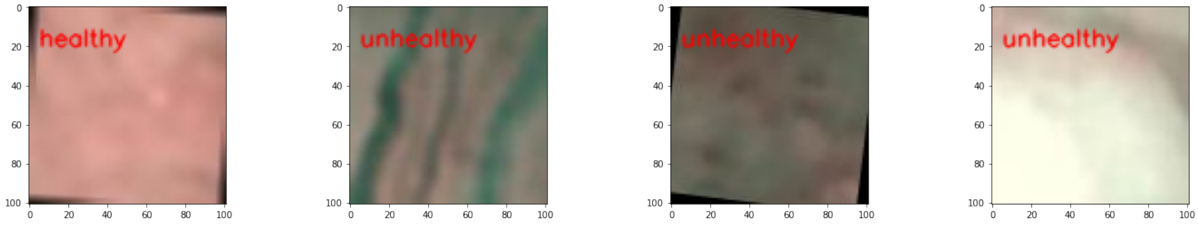


Figure 8: Accuracy and Loss Curves for Training and Validation Phase



(a) He (b) Hbv (c) IPCL (d) Le

Figure 9: Healthy or Unhealthy/Cancerous Tissues Prediction

The predictions done by CNN model are shown in figure 9. It is observed that model accurately predicts the diseased tissues as Unhealthy/Cancerous and normal tissues as healthy. These prediction are done by using model weights obtained in the training phase.

6.2 Baseline Approach 2 : Detection of Laryngeal Cancer using Transfer Learning

We evaluated the effect of transfer learning using Dense-Net121 state-of-art architecture for image recognition Huang et al. (2017). The model gained train accuracy as 0.99 but validation accuracy of 0.75. This significant difference is also observed in loss graph of train and validation phase shown in figure 10 and is sign of model overfitting. Thus, we infer that this transfer learning approach is not suitable for laryngeal cancer detection. The detailed evaluation metrics are shown in table 7 .

Table 7: CNN model evaluation

Test Accuracy	**Validation Accuracy**	**Sensitivity**	**Specificity**	**F1-Score**
0.75	0.75	0.83	0.67	0.75

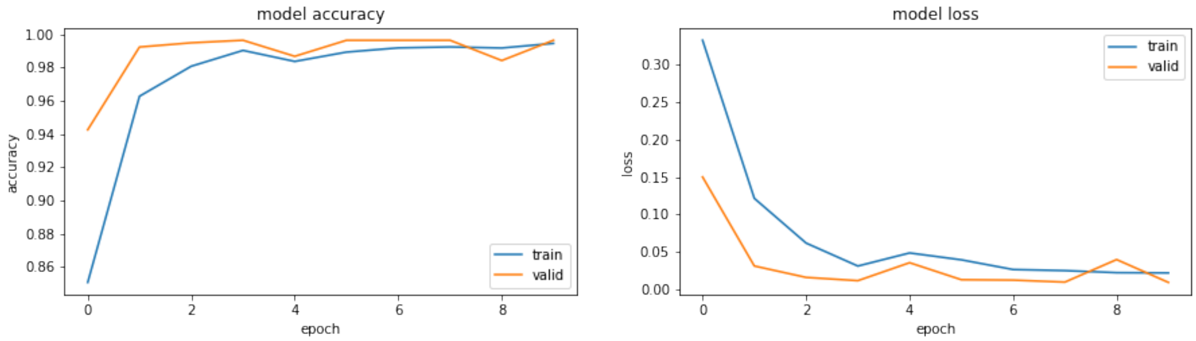


Figure 10: Accuracy and Loss Curves for Training and Validation Phase

6.3 Newly Proposed Approach : Laryngeal Cancer Detection Using Attention-Based Multiple Instance Learning

The Attention-based MIL model was evaluated using 3-fold cross validation by including the "earlystop" callback for better computation time, model generalization and performance than the state-of-the-art (Ilse et al.; 2018). The "earlystop" mechanism has patience level of 5 epoch and monitors the validation bag loss.

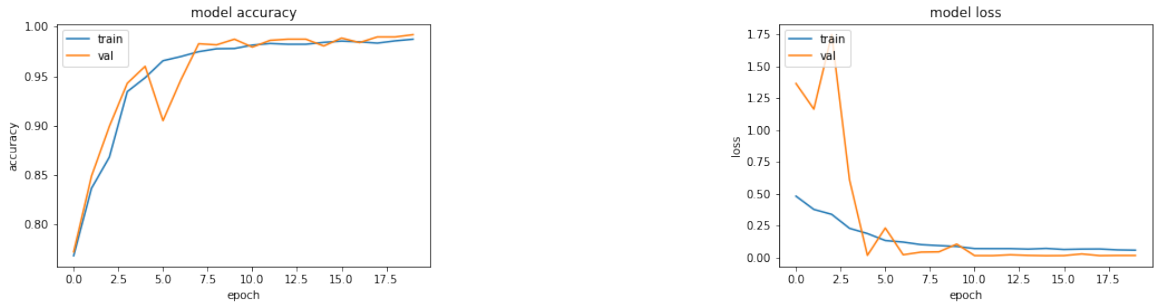


Figure 11: Accuracy and Loss Curves for Training and Validation Phase

The average training time for each fold was about 22 minutes and per epoch time is shown in table 8. The model performance was evaluated using train and validation bags whose loss and accuracy for one fold is shown in figure 11. The loss was monitored using "binary_crossentrophy" from the keras module. From the accuracy graph it is evident that the validation and training accuracy curves show stable increase with small spike in validation curve due to bias in 5th epoch. The training loss curve gradually decreases and becomes stable after 10th epoch. The validation loss curve also gradually becomes parallel with the train loss curve with a small gap. There was no evidence of overfitting, underfitting or biases which makes MIL as a best fit model for detection of laryngeal cancer.

Table 8: Evaluation of Attention-based MIL model

Fold	Test Accuracy	Validation Accuracy	Sensitivity	Specificity	Train time (min)
1	0.98	0.99	0.90	0.93	27.50
2	**0.98**	**0.98**	**0.93**	**0.96**	**23.25**
3	0.84	0.97	0.90	0.89	17

Sensitivity and specificity are two crucial metrics in evaluating any model trained on biomedical images. The table 8 summarizes these values for each fold along with test accuracy and training time. The average sensitivity and specificity values of the model were 0.90 and 0.93 respectively. It also performed relatively better as compared to other models which is evident in figure 12.

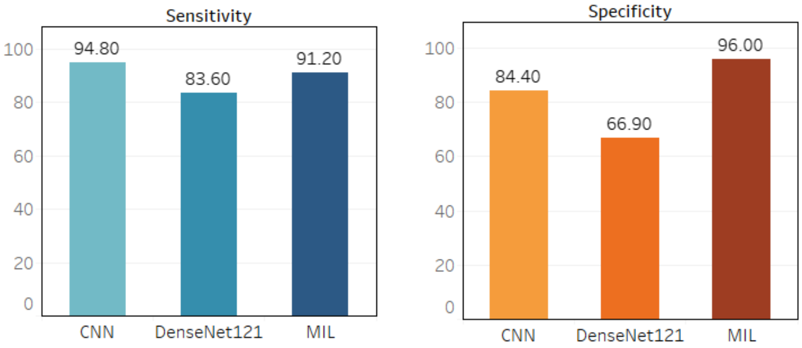


Figure 12: Basic Idea Behind Multiple Instance Learning

After overall analysis of evaluation it is observed that MIL approach achieves remarkable performance in classification and detection of laryngeal cancer at early stage.

6.4 Comparison and Discussion

Table 9: Evaluation of Attention-based MIL model

Models	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-Score
MIL-Net	**0.98**	**0.91**	**0.96**	**0.86**	**0.93**	**0.93**
CNN	0.90	0.94	0.84	0.90	0.90	0.90
Dense-Net121	0.75	0.84	0.67	0.76	0.75	0.75

The research discusses the Multiple Instance Learning (MIL) approach using convolution neural network for early detection of laryngeal cancer. To achieve this a novel attention-based MIL neural network was implemented. The MIL framework had performed well state-of-the-art biomedical image classification problems that include colon cancer and breast cancer classification. We incorporated it in laryngeal cancer detection which is done for the first time in laryngeal cancer domain. It achieved maximum accuracy of 0.98 with sensitivity of 0.91 which outperforms the current state-of-the-art in MIL based image classification.The proposed MIL model showed exemplary results when compared with CNN and Dense-Net121 depicted in table 9.

From the table 9 our attention-based MIL approach shows exceptional results with maximum accuracy of 0.98 in comparison with other models.

7 Conclusion and Future Work

The primary research objective was to study the effectiveness of Multiple Instance Learning in classification of laryngeal tissues. The secondary objective was to successfully enhance the image quality for better predictions. Finally evaluate the proposed model against proven models for image classification. It was found that the attention-based MIL model achieved outstanding results which are better than state-of-art. This model can assist the medical professionals for first line diagnosis of laryngeal cancer. Nevertheless, more better results could have been obtained with high quality images. Also, the constraint of dataset size has limited the scope of this research.

The future can involve multi-class classification using multiple instance learning which can give exact information about type of unhealthy tissue. Also, other types of tissues can be included for increasing model robustness.

Acknowledgement

I would like to offer my special thanks to my supervisor Dr. Rashmi Gupta for her constant encouragement and valuable feedback throughout my research . It was indeed a great experience to work under her guidance.

References

Araújo, T., Santos, C. P., De Momi, E. and Moccia, S. (2019). Learned and handcrafted features for early-stage laryngeal scc diagnosis, *Medical & Biological Engineering & Computing* **57**(12): 2683–2692.

Astorino, A., Fuduli, A., Veltri, P. and Vocaturo, E. (2020). Melanoma detection by means of multiple instance learning, *Interdisciplinary Sciences: Computational Life Sciences* **12**(1): 24–31.

Aubreville, M., Goncalves, M., Knipfer, C., Oetter, N., Würfl, T., Neumann, H., Stelzle, F., Bohr, C. and Maier, A. (2018). Transferability of deep learning algorithms for malignancy detection in confocal laser endomicroscopy images from different anatomical locations of the upper gastrointestinal tract, *International Joint Conference on Biomedical Engineering Systems and Technologies*, Springer, pp. 67–85.

Ausawalaithong, W., Thirach, A., Marukatat, S. and Wilaiprasitporn, T. (2018). Automatic lung cancer prediction from chest x-ray images using the deep learning approach, *2018 11th Biomedical Engineering International Conference (BMEiCON)*, IEEE, pp. 1–5.

Bhattacharjee, K., Pant, M., Zhang, Y.-D. and Satapathy, S. C. (2020). Multiple instance learning with genetic pooling for medical data analysis, *Pattern Recognition Letters* .

Chen, L., Tong, T., Ho, C. P., Patel, R., Cohen, D., Dawson, A. C., Halse, O., Geraghty, O., Rinne, P. E., White, C. J. et al. (2015). Identification of cerebral small vessel disease using multiple instance learning, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 523–530.

Diamant, A., Chatterjee, A., Vallières, M., Shenouda, G. and Seuntjens, J. (2019). Deep learning in head & neck cancer outcome prediction, *Scientific reports* **9**(1): 1–10.

Galdran, A., Costa, P. and Campilho, A. (2019). Real-time informative laryngoscopic frame classification with pre-trained convolutional neural networks, *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, pp. 87–90.

Horie, Y., Yoshio, T., Aoyama, K., Yoshimizu, S., Horiuchi, Y., Ishiyama, A., Hirasawa, T., Tsuchida, T., Ozawa, T., Ishihara, S. et al. (2019). Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks, *Gastrointestinal endoscopy* **89**(1): 25–32.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. (2017). Densely connected convolutional networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

Ilse, M., Tomczak, J. M. and Welling, M. (2018). Attention-based deep multiple instance learning, *Conference Proceeding* .

Jeyaraj, P. R., Panigrahi, B. K. and Samuel Nadar, E. R. (2020). Classifier feature fusion using deep learning model for non-invasive detection of oral cancer from hyperspectral image, *IETE Journal of Research* pp. 1–12.

Ji, B., Ren, J., Zheng, X., Tan, C., Ji, R., Zhao, Y. and Liu, K. (2020). A multi-scale recurrent fully convolution neural network for laryngeal leukoplakia segmentation, *Biomedical Signal Processing and Control* **59**: 101913.

Luan, B., Sun, Y., Tong, C., Liu, Y. and Liu, H. (2019). R-fcn based laryngeal lesion detection, *2019 12th International Symposium on Computational Intelligence and Design (ISCID)*, IEEE, pp. 128–131.

Ma, L., Lu, G., Wang, D., Qin, X., Chen, Z. G. and Fei, B. (2019). Adaptive deep learning for head and neck cancer detection using hyperspectral imaging, *Visual Computing for Industry, Biomedicine, and Art* **2**(1): 1–12.

Moccia, S., De Momi, E., Guarnaschelli, M., Savazzi, M., Laborai, A., Guastini, L., Peretti, G. and Mattos, L. S. (2017). Confident texture-based laryngeal tissue classification for early stage diagnosis support, *Journal of Medical Imaging* **4**(3): 034502.

Moccia, S., Vanone, G. O., De Momi, E., Laborai, A., Guastini, L., Peretti, G. and Mattos, L. S. (2018). Learning-based classification of informative laryngoscopic frames, *Computer methods and programs in biomedicine* **158**: 21–30.

Moldovan, D. (2019). Transfer learning based method for two-step skin cancer images classification, *2019 E-Health and Bioengineering Conference (EHB)*, IEEE, pp. 1–4.

Obukhova, N., Motyko, A., Pozdeev, A. and Timofeev, B. (2018). Review of noise reduction methods and estimation of their effectiveness for medical endoscopic images processing, *2018 22nd Conference of Open Innovations Association (FRUCT)*, IEEE, pp. 204–210.

Pedersen, M., Cherepkova, O. and Mohammed, A. (2017). Image quality metrics for the evaluation and optimization of capsule video endoscopy enhancement techniques, *Color and Imaging Conference*, Vol. 2017, Society for Imaging Science and Technology, pp. 20–27.

Quellec, G., Cazuguel, G., Cochener, B. and Lamard, M. (2017). Multiple-instance learning for medical image and video analysis, *IEEE reviews in biomedical engineering* **10**: 213–234.

Rahi, M. M. I., Khan, F. T., Mahtab, M. T., Ullah, A. A., Alam, M. G. R. and Alam, M. A. (2019). Detection of skin cancer using deep neural networks, *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, IEEE, pp. 1–7.

Rodner, E., Bocklitz, T., von Eggeling, F., Ernst, G., Chernavskaia, O., Popp, J., Denzler, J. and Guntinas-Lichius, O. (2019). Fully convolutional networks in multimodal nonlinear microscopy images for automated detection of head and neck carcinoma: Pilot study, *Head & Neck* **41**(1): 116–121.

Sudharshan, P., Petitjean, C., Spanhol, F., Oliveira, L. E., Heutte, L. and Honeine, P. (2019). Multiple instance learning for histopathological breast cancer image classification, *Expert Systems with Applications* **117**: 103–111.

Taylor, L. and Nitschke, G. (2018). Improving deep learning with generic data augmentation, *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, pp. 1542–1547.

Tong, T., Wolz, R., Gao, Q., Guerrero, R., Hajnal, J. V., Rueckert, D., Initiative, A. D. N. et al. (2014). Multiple instance learning for classification of dementia in brain mri, *Medical image analysis* **18**(5): 808–818.

Umapathi, V. and Narayanan, V. S. (2014). Medical image denoising based on gaussian filter and dwt swt based enhancement technique, *International Journal of Soft Computing and Artificial Intelligence* **2**(2).

Wang, X., Yan, Y., Tang, P., Bai, X. and Liu, W. (2018). Revisiting multiple instance neural networks, *Pattern Recognition* **74**: 15–24.

Xiong, H., Lin, P., Yu, J.-G., Ye, J., Xiao, L., Tao, Y., Jiang, Z., Lin, W., Liu, M., Xu, J. et al. (2019). Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images, *EBioMedicine* **48**: 92–99.

Zeng, T. and Ji, S. (2015). Deep convolutional neural networks for multi-instance multi-task learning, *2015 IEEE International Conference on Data Mining*, IEEE, pp. 579–588.