AI based diagnosis & classification of lymph node Fine Needle Aspiration Cytology (FNAC)

Muskaan Beriwal
Dept of Computer Engineering
MPSTME, NMIMS
Mumbai, India
mberiwal13@gmail.com

Shiv Avashia
Dept of Computer Engineering
MPSTME, NMIMS
Mumbai, India
shivashia12@gmail.com

***Abstract*—In clinical practice, an essential technique that is used by pathologists to diagnose lymph node malignancies is the visual examination of cytopathological slides. Manual visual inspection of whole slide images is challenging, time-consuming, and subject to substantial inter-observer variability, which might result in sub-optimal detection of said malignancies. These False-positive and False-negatives can lead to further complications. Furthermore, upon understanding the current situation, we find that pathologists must review hundreds of slides on an everyday basis. Most of the reviewed slides are normal in nature while only a fraction being abnormal or malignant. The abnormal slides need to be submitted for further analysis. This makes it a very time- consuming and tedious process. Our research looks at pathology from a computational perspective. We leverage deep learning technology to learn features, patterns, and detect anomalies to aid in the detection of lymph node malignancies in whole slide images of fine needle aspiration cytology. An initial screening to classify reactive, infectious, and neoplastic slides would reduce the workload of a pathologist largely. Our main goal was to develop a dataset containing digitized images of reactive, infectious, and neoplastic FNAC slides of lymph nodes. Furthermore, we used this dataset to identify techniques for pre-processing these images such as clustering, thresholding and morphological operations. We then identified and implemented 3 classification models that were appropriate for our dataset, namely, Convolutional Neural Network (CNN), InceptionNet and Resnet50. On evaluation of the results obtained, we found that in terms of accuracy, Resnet50 performed the best amongst all three models with a training accuracy of 76.19%. However, further future improvements such as obtaining a larger dataset and performing feature extraction are necessary in order to increase the accuracy of the models.**

***Index Terms*—FNAC, Image Processing, Classification, Deep Learning, Computational Pathology, Cytology, Artificial Intelligence, Neural Network**

I. Introduction

Computational pathology is a combination of digital pathology, medical image analysis, computer vision, and machine learning. Digital pathology is an ideal application for advanced image analysis techniques due to the massive quantity of information and data accessible in multi-gigapixel cytopathology images. As a result, artificial intelligence and deep learning have proven to be successful in powering computational pathology research.

AI-based computational pathology has the potential to improve the efficiency of clinical workflow and diagnostic quality, and ultimately create personalized diagnosis and treatment regimens for patients. A significant advantage of computational pathology is to improve accuracy and reduce errors in diagnosis and classification. The traditional Pathological Diagnosis techniques practiced currently have the following limitations:

1) It is a time-consuming process.
2) There is a shortage of experienced pathologists and the limitation of global health care resources.

One of the diagnosis techniques is Fine Needle Aspiration Cytology (FNAC) of lymph nodes. Classifying the FNAC of lymph nodes is a tedious process as pathologists need to manually review hundreds of slides on an everyday basis. We aim to automate the process by leveraging deep learning technology. However, to accomplish this, we need to develop a dataset containing the required images (due to lack of any existing standard datasets). By successfully automating the process of classification, we hope to reduce the strain on medical experts.

The research objectives of our study include:

- Building a dataset containing digitized versions of stained whole slide images of Reactive, Infective and Neoplastic FNACs of lymph nodes, under the guidance of a medical expert (since a dataset containing labelled images of FNAC slides of lymph nodes is not available as of now).
- Identifying the techniques for pre-processing such images to highlight parameters such as nuclear size, shape, and texture. Image processing techniques can enhance important features while also reducing noise and filtering out unnecessary information.
- Identifying and implementing models that work best with these kinds of images. Not all deep learning models can perform the intricate classification that is required in this application. Thus, we need to analyze the architecture of different models and judge whether the model can perform the required task.

This paper provides a comparative analysis of 3 classification models, each with a differently augmented and pre-processed dataset. The rest of the paper is structured as follows: Section II gives an overview of relevant related works. Section III focuses on the methodology. Implementation and results are discussed in section IV. Finally, section V presents the conclusion and future scope.

II. Related Works

Kashyap A. et al. [1] explored the statistical techniques, clinical applications, potential challenges, and future possibilities in computational pathology. The paper comes to the conclusion that artificial intelligence is a powerful tool that can be used to classify cells with a high accuracy, but first, certain challenges must be overcome like integrating raw data from various sources, limitation of hardware processing capacity, and a lack of specific training programs.

The aim of the work done by Aeffner F. et al. [2] is to develop an intelligent diagnosis system for breast cancer classification. Classification of FNAC of benign and malignant breast tumours is done using artificial neural networks and support vector machines. Multilayer perceptron (MLP) with back-propagation algorithm, probabilistic neural networks (PNN), learning vector quantization (LVQ), and support vector machine were the four classification models employed (SVM). The authors found that probabilistic neural network and support vector machine are better classifiers than the other models. In the study conducted by Cui M. et al. [3], various Cytological breast aspirates were subjected to nuclear morphometry in order to examine how well it could discriminate between benign and malignant breast lesions and to establish the right cut-off values for each category.

The paper by George et al. [4] studies the basics of tissue image analysis with the aim to provide pathologists, with information such as the features, applications, and general workflow of these tools. The authors also look at the advantages and limitations of tissue image analysis and how they can be used for artificial intelligence and machine learning.

The study conducted by Dr. Kamlesh lenka et al. [5] dealt with the challenges that come with identifying malignant cells in whole slide images of FNAC of breast lesions with regards to consistency and reproducibility. The authors discuss the various deep convolutional neural network architectures based on fine-tuned transfer learning to classify cells into benign and malignant. They examined VGG16, VGG19, ResNet-50 and GoogleLeNet-V3 (inception V3) on a dataset consisting of 212 images (90 benign and 113 malignant) which were augmented and cleansed while pre-processing.

The goal of the study by Hafez NH et al. [6] was to find out various causes of lymph nodes diseases and study the approach towards lymph node cytology to help in classifying malignant lymph nodes. In another study, authors Guan Q. et al. [7] aimed to find out the reliability and diagnostic accuracy of FNAC as the preliminary screening of lymph nodes by keeping track of Diagnostic sensitivity, specificity, positive predicted values (PPV), negative predicted values (NPV), accuracy, and discordance rate.

Authors Amartya Ranjan Saikia et al. [8] analyzed clusters of nucleus in the sub-band image of a FNAC sample generated using ”complex discrete wavelet transform” and developed a new method which helped them to compute the variability and other nucleus-statistical textural features. The calculated variance parameters and texture features were then used to characterize the nucleus clusters into normal and malignant.

The study by Sanyal P. et al. [9] aimed to create an artificial neural network (ANN) that can discriminate between papillary and non-papillary cancer thyroid on micro photographs from thyroid FNAC smears. Authors R. Beaulah Jeyavathana et al. [10] provided a comparative study between different pre-processing techniques. The survey gave information on automatic segmentation techniques, with a focus on CT images. The goal was to address the challenges involved in segmenting CT scans as well as the relative benefits and drawbacks of the technologies currently used to segment medical images.

III. Methodology

Fig. 1. Proposed Methodology

Figure 1 shows the proposed methodology. The first step is Data Acquisition. Since no standard datasets were available for FNAC of lymph nodes, we built our own dataset by scraping images from textbooks, websites, and online slides, belonging to the three classes – Reactive, Infective and Neoplastic. Since we had limited data, we then performed Data Augmentation to increase the size of the training data. Next, we used pre-processing techniques to highlight the point of interests and remove the noise present in the images. We then identified models that would be most suitable for our dataset and then trained those models on the pre-processed dataset and the training and testing accuracy curve was plotted. This gave us insights on what hyperparameters to tune. We varied different hyperparameters to fine-tune the model and improve the performance of the model.

A. *Dataset Acquisition*

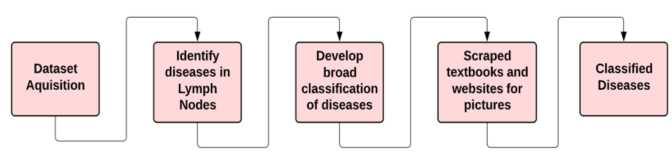

Fig. 2. Dataset Acquisition Diagram

With the help of a medical expert, we identified and listed out all the possible diseases that occur in lymph nodes. These diseases were then broadly classified into three categories – Reactive, Infective and Neoplastic. Images belonging to each class were then scraped from medical textbooks, websites, and virtual pathologies. Figure 2 represents the process flowchart.

B. Dataset Augmentation

Data augmentation is useful to improve performance and outcomes of deep learning models by artificially increasing the amount of data by creating new data points from existing data. After carefully analysing the images, we identified two techniques of augmentation, which are shown in Figure 3. These techniques keep in mind the nature of the images and aim to reduce the chances of over fitting the model.

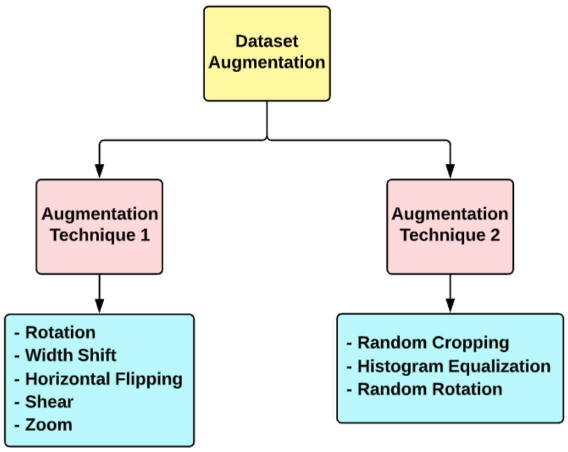

Fig. 3. Dataset Augmentation Process

Technique 1:

- Rotation: We rotate the images with a range of 0-30 degrees. This changes the orientation of the slides thus making the image ready for further augmentation techniques to be applied.
- Width shift: This technique is used to shift the image to the left or right (horizontal shifts).
- Horizontal-Flipping: To ”flip” or ”mirror” an image in the horizontal direction (left-right).
- Shear: Shear is used to shift one part of an image, a layer, a selection or a path to a direction and the other part to the opposite direction.
- Zoom: To specify points of interest.

Technique 2:

- Random Cropping: Random crop is a data augmentation technique wherein we create a random subset of an original image. This helps in better generalization of our model.
- Histogram equalization: It is used to enhance contrast in images. It accomplishes this by effectively stretching out the intensity range of the image.
- Random rotation: Rotation is a feature that allows us to turn an image in a clockwise or counterclockwise direction. If applied on the original image, the model could be over-fit as the degree of rotation does not matter while working with images of slides. So, for this to be effective, we apply random rotation after random cropping.

C. Dataset Pre-processing

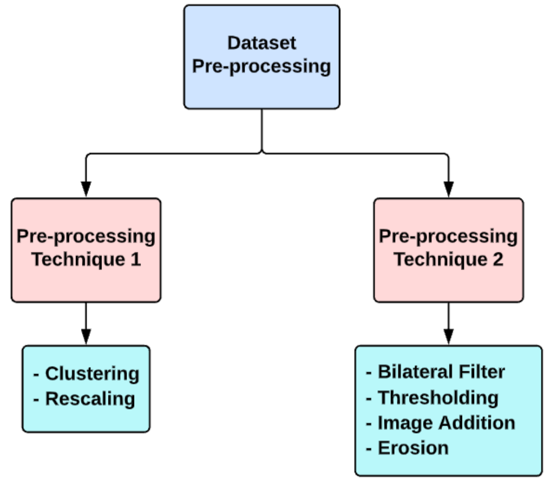

Fig. 4. Dataset Pre-processing Pipeline

Figure 4 shows the dataset pre-processing pipeline.

Technique 1: Here we used clustering as an edge-based segmentation method to identify the cells (highlight them) and rescale them. This technique highlights each cell distinctly while also reducing the background noise. Hence, the model can learn features such as cell shape, size, texture, and spread of the cells with ease. The pre-processing pipeline is as follows-

- Clustering: Clustering helps us in edge detection. Edge detection is used to identify points in a digital image with discontinuities, or sharp changes in the image brightness. The edges of the image are those regions where the brightness of the image fluctuates dramatically. To process the learning data, we employ the K-means algorithm on a set of centroids that were chosen at random and served as the starting points for each cluster.
- Rescaling: Rescaling is essentially normalizing the pixel values or clipping the pixel values between 0 and 1. This is done to prevent the exploding gradient problem.

Technique 2: In this technique, we apply a bilateral filter on the image followed by thresholding to increase the intensity of low-frequency information. This is added to the original image and erosion is performed on the summation. This technique almost completely removes background information while preserving the most important foreground information. This reduces the number of visual features that the model has to learn thus simplifying the classification process. The pre-processing pipeline is as follows-

- Bilateral Filter: A bilateral filter is a non-linear, edge-preserving, and noise-reducing smoothing filter for images. It replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels.

After applying this filter, the edges in the images become clearer therefore highlighting the cells and reducing the amount of unnecessary background information in the image.

- Thresholding: Thresholding is a method for performing image segmentation. Thresholding can be done in a certain range to binarize the image, thus reducing the amount of information in it and highlighting edges and other high-frequency features. After thresholding, the resultant image highlights the low-frequency information.
- Image Addition: Image addition enables us to combine the information contained in two or more images by adding the corresponding pixel from each input image. After thresholding, the low-frequency information is added to the original image. Image addition enhances the high-frequency information while also contributing to the enhancement of low-frequency information in the original image. The image now outlines all clusters of cells in the image while also highlighting the most prominent cells in each cluster.
- Erosion: Erosion is a fundamental morphological operation which makes use of a kernel to shrink the image. Erosion removes pixels on object boundaries. In our application, we used an elliptical morphological kernel smaller than the cells/clusters of cells while applying erosion. The output image clearly shows the foreground while reducing the unnecessary background information to a large extent.

D. *Classification Models*

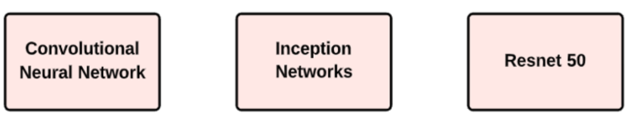

Fig. 5. Proposed Classification Models

*1) **Convolutional Neural Network**:* A Convolutional Neural Network (CNN) is a Deep Learning algorithm that can take in an input image, give importance to various elements and objects in the image (learnable weights and biases) and be able to distinguish one from the other. It is a supervised learning algorithm. Convolutional neural networks are capable of learning important features that are required for a classification. This can be achieved via tuning hyperparameters and training. As the nature of our network is to classify data, we can use either a sigmoid or a SoftMax activation layer in the final fully connected layer depending on the classification that we need to make.

*2) **Inception Networks**:* As opposed to stacking layers as done in a CNN, an inception network is a heavily engineered neural network which helps build custom classifiers that are optimized in both speed and accuracy. Inception networks contain inception layers which have filters of different sizes to identify objects of varying sizes. These filters are generally 1x1, 3x3 and 5x5 convolutions along with 3x3 max pooling. This inception layer reduces the cost of computation due to the dimensionality reduction caused by the 1x1 convolutions. Inception networks also make use of auxiliary classifiers to prevent the neurons from dying out.

*3) **Resnet 50**:* A residual network is an artificial neural network which utilizes skip connections to jump over some layers. These architectures concatenate the output of a layer (example: layer i) with the output of a later layer (example: layer i+2 or layer i+3). We believe that residual networks will be more effective for the classification problem at hand as skip connections avoid the problem of vanishing gradients. This ensures that the model does not fail while training the feature-wise normalized images wherein the mean of each image is set to 0. Residual networks also help mitigate the Degradation (accuracy saturation) problem wherein adding more layers to a suitably deep model leads to higher training error. Thus, we can have a deeper model to extract complex features.

IV. Implementations & Results

A. *Implementation*

Due to the limited work done in this specific field, our work involved a significant amount of experimentation in terms of augmentation, pre-processing, model selection, hyper-parameter tuning etc.

The dataset was augmented using two different pipelines as shown in Table 4.1. Each pipeline performed two kinds of data balancing- 1) oversampling all classes and 2) under sampling the skewed class while oversampling the other two classes.

Augmentation Technique	Processes Involved	Library Used
Aug1	Rotation, Width Shift, Horizontal Flipping, Shear, Zoom	Keras
Aug2	Random Cropping, Histogram Equalization, Random Rotation	Augmentor

Table 4.1 Dataset Augmentation

This resulted in 4 datasets given in Table 4.2. Each of these datasets were pre-processed with the aforementioned two pre-processing techniques. We then implemented models that were suitable for the classification task- CNN, InceptionNetV1, ResNet50.

1) **Convolution Neural Network -** The CNN accepts input images of size 512x512. This model trains 13 million parameters. The loss function used is categorical cross-entropy as we are performing multi-class classification. We trained the model with Stochastic Gradient Descent which adjusts the learning rate using a learning rate scheduler as well as Adam Optimizer.
2) **Inception Network (Inception v1) -** Our inception network trains 10 million parameters. The network has two auxiliary outputs and one final output. The loss function used is categorical cross-entropy along with

stochastic gradient descent optimizer. This optimizer has a learning rate scheduler which decays the learning rate according to the validation loss.

3) **Resnet50 -** The model accepts an RGB image of 224 x 224. It trains 23 million parameters while keeping approximately 53 thousand parameters non-trainable. The final Dense layer consists of 3 neurons and a softmax activation function to perform 3-class classification. We obtain the probabilities of the image belonging to each class as output. The loss function used is categorical cross-entropy along with Adam optimizer.

Dataset Name	Number of Images	Dataset Features
Dataset 1	Total: 5740 Reactive: 1768 Infective: 1843 Neoplastic: 2129	Done using Aug1. All classes were oversampled.
Dataset 2	Total: 1440 Reactive: 416 Infective: 454 Neoplastic: 570	Done using Aug1. Only Reactive and Infective classes were upsampled.
Dataset 3	Total: 6000 Reactive: 2000 Infective: 2000 Neoplastic: 2000	Done using Aug2 All classes were upsampled.
Dataset 4	Total: 900 Reactive: 300 Infective: 300 Neoplastic: 300	Done using Aug2. Neoplastic class was downsampled and reactive and infective class were upsampled.

Table 4.2 Dataset Generated

B. *Results & Discussion*

The results were obtained after conducting experiments with different augmentation techniques, pre-processing, and classification models. These results are indicative of which model can best learn the required visual features given certain pre-processing. Based on the analysis of the results obtained, we can hypothesize which models would perform best in terms of accuracy, precision, and recall. That model could be applied in a real- world scenario provided that a standardized dataset is obtained.

Table 4.3 compares the results obtained on the basis of training and validation accuracy. From these results, we can see that augmenting the dataset to 2000 images (dataset 1 & 3) does not work to our advantage as it increases the training time while also making the model prone to over fitting.

The dataset with 400 images (dataset 2 & 4) performs considerably well when compared to the 2000 images dataset, which is evident by the accuracy values. Both techniques of pre-processing led to good results. This could be attributed to the similar nature of what was being accomplished by pre-processing i.e., reducing background noise while retaining or enhancing high-frequency information in the image.

Pre-processing technique 1 (dataset 1 & 2) worked better with the CNN while pre-processing technique 2 (dataset 3 & 4) worked better with ResNet. In terms of model, the inception net performed poorly in all scenarios. This was mainly due to the limited dataset. Although the dataset was augmented, the features more-or-less remained the same which was detrimental to the performance of the inception net.

After observing the predictions of the InceptionNet we also noticed a vanishing gradient problem as all probabilities were close to 0 for example, an image had the following probabilities - [0.09, 0.02, 0.0083]. The ResNet performed uniformly regardless of the pre-processing technique. However, in terms of training time, training the ResNet on the 2000 image dataset took up a long time.

In terms of accuracy, ResNet performed the best amongst all three with a training accuracy of 76.19% and validation accuracy of 47.78%.

Classification Model	Dataset used	Training Accuracy	Validation Accuracy
CNN	Dataset 1	37.10 %	37.0%
	Dataset 2	73.89%	62.69%
	Dataset 3	42.88%	36.14%
	Dataset 4	60.95%	37.03%
InceptionNet	Dataset 1	37.09%	37.10%
	Dataset 2	39.59%	39.55%
	Dataset 3	47.17%	37.22%
	Dataset 4	43.0%	40.0%
Resnet50	Dataset 1	54.81%	39.68%
	Dataset 2	56.59%	45.01%
	Dataset 3	61.23%	40.40%
	Dataset 4	76.19%	47.78%

Table 4.3 comparison of results based on accuracy

V. Conclusion & Future Scope

A. *Conclusion*

AI-based computational pathology is an emerging discipline. It is a branch of pathology that deals with information extraction from digitized pathology images, typically using deep learning. One of the diagnosis techniques used in preliminary classification of cancer cells is Fine Needle Aspiration Cytology of lymph nodes. Classifying the FNAC of lymph nodes is a tedious process. Our aim was to automate the process by leveraging deep learning technology. Since no standard datasets were available for FNAC of lymph nodes, we built our own dataset by scraping images belonging to the three classes:

- Reactive
- Infective
- Neoplastic

Due to limited data, data augmentation was done to increase the size of training data. We used pre-processing techniques like clustering and bilateral filtering to highlight the point of interests and remove the noise present in the images. We applied three models - CNN, InceptionNetV1 and ResNet50 for training the pre-processed dataset. Based on the analysis of the results obtained, the dataset with 400 images (dataset 2 & 4) performed considerably well when compared to the 2000 images dataset. Both techniques of pre-processing led to good results. In terms of classification models, the inception net performed poorly in all scenarios. In terms of accuracy, ResNet performed the best amongst all three with a training accuracy of 76.19% and validation accuracy of 47.78%.

B. *Future Scope*

- The dataset we acquired was small as compared to the general size of datasets neural networks require. Therefore, we would like to acquire a larger dataset which is standardized, i.e., have uniform magnification (5x, 20x, 40x, 100x) and develop a classification model based on this.
- Pathologists note features such as size of cell, ratio of nucleus to cytoplasm, distribution of cells, texture etc. to perform differential analysis while classifying FNAC of lymph nodes thus making feature extraction essential to getting an accurate classification model. We aim to extract these features and perform classification based on these parameters. Extracting such parameters opens the possibilities of the models that can be applied to this study. Models such as decision trees, support vector machines, logistic regressors can be applied.
- We plan to use models such as Transformers to perform classification after obtaining a larger dataset. Transformers have an attention mechanism that are capable of focusing only on the most important parts of the image while classifying it.

REFERENCES

[1] Kashyap A, Jain M, Shukla S, Andley M. Study of nuclear morphometry on cytology specimens of benign and malignant breast lesions: A study of 122 cases. J Cytol. 2017 Jan-Mar;34(1):10-15. doi: 10.4103/0970-9371.197591. PMID: 28182052; PMCID: PMC5259923.

[2] Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, Lujan GM, Molani MA. Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association. J Pathol Inform. 2019 Mar 8;10:9. doi: 10.4103/jpi.jpi82 18.

[3] Cui, M., Zhang, D.Y. Artificial intelligence and computational pathology. Lab Invest 101, 412–422 (2021). https://doi.org/10.1038/s41374-020-00514-0

[4] George, Yasmeen Elbagoury, Bassant Zayed, Hala Roushdy, Mohamed. (2012). Breast Fine Needle Tumor Classification using Neural Networks. IJCSI International Journal of Computer Science. 9. 247-256.

[5] Dr. Kamlesh lenka, Approach To lymph Node Cytology, Jul. 30, 2019, https://www.slideshare.net/KamaleshLenka/approach-to-lymph-node-cytology-part-1

[6] Hafez NH, Tahoun NS. Reliability of fine needle aspiration cytology (FNAC) as a diagnostic tool in cases of cervical lymphadenopathy. J Egypt Natl Canc Inst. 2011 Sep;23(3):105-14. doi: 10.1016/j.jnci.2011.09.009. Epub 2011 Oct 24. PMID: 22776815.

[7] Guan Q, Wang Y, Ping B, Li D, Du J, Qin Y, Lu H, Wan X, Xiang J. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. J Cancer 2019; 10(20):4876-4882. doi:10.7150/jca.28769.

[8] Amartya Ranjan Saikia, Kangkana Bora, Lipi B. Mahanta, Anup Kumar Das, Comparative assessment of CNN architectures for classification of breast FNAC images, Tissue and Cell, Volume 57, 2019, Pages 8-14, ISSN 0040-8166, https://doi.org/10.1016/j.tice.2019.02.001.

[9] Sanyal, P.; Mukherjee, T.; Barui, S.; Das, A.; Gangopadhyay, P. "Artificial intelligence in cytopathology: A neural network to identify papillary carcinoma on thyroid fine- needle aspiration cytology smears" J. Pathol. Inform. 2018, 9, 43.

[10] R. Beaulah Jeyavathana, R. Balasubramanian, and A. A. Pandian, "A Survey: Analysis on Pre - processing and Segmentation Techniques for Medical Images," Int. J. Res. Sci. Innov., vol. III, no. June, pp. 2321–2705, 2016.