

## Machine Learning Approach for Prediction of Cervical Cancer

B Jyothi Priyanka<sup>a</sup>, Dr M S V S Bhadri Raju<sup>b</sup>

<sup>a</sup>M.Tech, Computer Science & Technology, S.R.K.R Engineering College, Bhimavaram, Andhra Pradesh

<sup>b</sup>Professor, Department of CSE, S.R.K.R Engineering College, Bhimavaram, Andhra Pradesh

<sup>a</sup>priyajagan632@gmail.com, <sup>b</sup>msramaraju@gmail.com

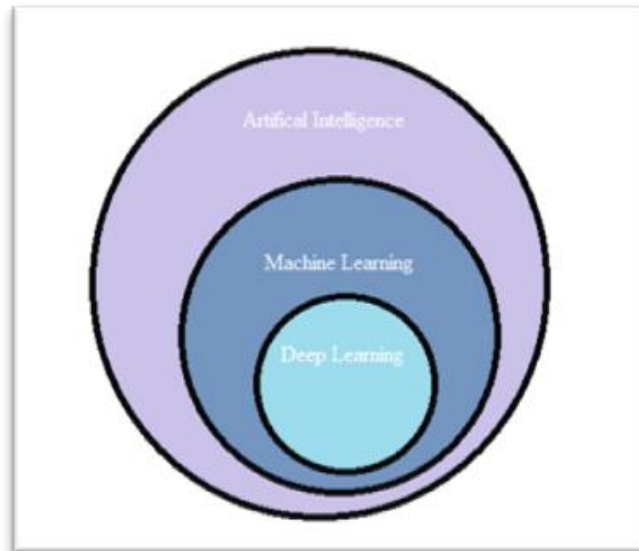
**Article History:** Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

**Abstract:** Women in the world are suffering from many diseases among those diseases Cervical Cancer is also mentioned globally. Every year many cancer cases are being registered throughout the world. Cervical cancer is ranked fourth of all the other common cancers according to WHO. Prediction of this cancer in its early stages can be cured, avoiding the death rate. Many people are less aware of this type of cancer as this disease is symptom less. Performance of screening test in regular bases cancerous cells can be detected in its early stages which reduces the mortality rate of people every year. There are many medical approaches for the prediction of this cancer like pap-smear test, colposcopy, biopsy, HPV test or HPV DNA test and other screening tests are performed. These medical methods are combined with the Artificial Intelligence for less false rate and more accurate results. This paper considers pap-smear test images for the prediction of cancerous cells combined with Deep Learning techniques for more efficient results. Convolution Neural Networks (CNN's) ResNet50 pre-trained model for the prediction of cancerous cells which produces accurate results. The proposed work classifies the cells from the inputted images. This cancer can be cured when it is in the initial stages, the identified abnormal cells helps us for the further treatment. The proposed methodology classifies all the classes with 74.04% of accuracy in prediction of cells for maximum number of epochs. Also in addition, it specifies the class of the testing image.

**Keywords:** Deep Learning, Convolution Neural Network, ResNet50, Transfer Learning, Cancer of Cervix, Pap-Smear images

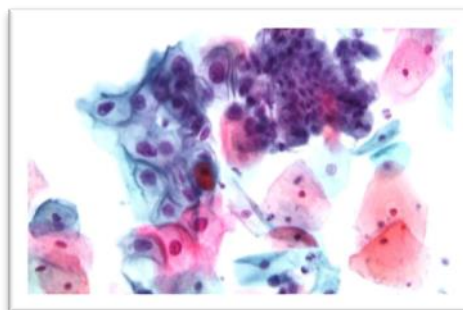
### 1. Introduction

Many centuries have been passing for the discovery of cancer disease, yet it's been a challenge to the Medical Sciences to abolish cancer cases throughout the world. But they succeeded in the prediction and cure of the disease in its early stages up to some extent. Cancer is the collection of diseases. Human body consists of trillions of cells. Normally, cells divide and form new cells when the cells become older they die and those dead cells are replaced by the new cells. This cells replacement is done even when the cells are damaged. The uncontrollable count of new cells generation leads to cancer in human body, production of unnecessary new cells lead to tumors. There are many types of cancers present in this world. Some cancers are malignant they spread to other tissues in the body [13]. Thus, it's better to predict the cancerous cells at the early stages. There are many medical methods to discover the cancer in its early stages which helps the patients to avoid the worst case. When compared to the olden days the Medical Sciences are much more advanced. As there is a saying, "Prevention is better than cure" it's better to prevent human from cancer disease by vaccinating early. There are vaccines for many diseases like polio, small pox, flu and other diseases similarly the advancement of Medical Sciences helped human races to invent vaccine for cancer disease which can prevent the cancer attacks to some extent [4]. Cancer disease is gender biased thus, both male and female act as victims. Most common cancers in women are breast, colorectal, endometrial, cervical, lung and skin cancers as per American Cancer Society [2]. As count of cancer patients are increasing in huge number it's been challenge to Medical Sciences. Many medical methods have been invented in the prediction and cure of cancer [1]. As the technology has wide spread its wings in every area many novel techniques are being implemented for faster results. Same way Medical Sciences are combined with Artificial Intelligence for better improvement in Medicine. Introduction of modern computing technologies can reduce the human errors and also help in faster results. Machine Learning has originated from Artificial Intelligence which mostly deal with Oncology department in Medicine and Deep learning is subpart of Machine Learning, fusion of Artificial Intelligence with human is said to lead a fruitful result in the ongoing centuries.



**Figure 1:** Interlink of computing technologies

Cervical cancer is, “the fourth most common cancer in women according to the survey by World Health Organization” [20]. This is dangerous cancer when compared to other cancers. Cervical cancer belongs to the carcinoma family. The occurrence of this cancer is by getting infected with HPV virus. HPV virus is found by group of scientists it is mostly transmitted through sexual contact. There are many types of HPVs present, cancer is caused by type16 and type18 [19]. These types are said to high-risk HPVs as these cause cancerous cells in the body whereas type6 and type11 are low-risk HPVs they cause genital warts on the skin. Normally, in more number of cases HPV infection disappears in two years without causing any abnormalities but in some cases, they are responsible for causing precancerous cells in the human body. Cervical cancer starts from the cervix this cancer doesn't have any symptoms till its early stages. This cancer differs there are two types it is observed when seen under a microscope. The first type is Squamous cell carcinoma starts in the cells that line the bottom of cervix. Another one is Adenocarcinoma it develops in the gland related cells that are present at the upper part of the cervix [3]. As this cancer doesn't have any symptoms regular screening tests are said to be done to stay free from the panic of having the cancerous cells. This screening includes Pap-smear test which results the presence or absence of precancerous cells. When the results are obtained if the result shows us the presence of abnormal cells they must be further treated or may require some additional tests to be done. If there are no such abnormal cells it doesn't require any further examination [12].



**Figure 2:** Image of Pap-smear test of abnormal cell

The images are obtained by performing pap-smear test those images are called “pap-smear images”. The cells are collected from cervix and screening is performed based on the sample cells taken. Those images are said to be processed by using the novel techniques to identify the cells whether they are normal or abnormal. As there is advancement in the technologies day-by-day Artificial Intelligence has wide – spread its wings in the field of healthcare. There are many algorithms and techniques to predict the cells, classify the cells and also to segment the cells. Deep Learning models in the prediction of cancerous cells is more efficient when compared to Machine Learning algorithms. Classification of cells is done faster as pre-trained techniques of CNN are used which doesn't require to be started from the scratch. In this paper pap-smear images are considered and based on those images cells are identified as cancerous and non-cancerous cells. The identified cells undergo 7 classes' namely

superficial squamous, intermediate squamous and columnar epithelial are set of normal cells. Mild, moderate and severe dysplasia, also carcinoma-in-situ are classes of abnormal cells.

## 2. Related work:

From the decades the cure of cancer has taken many paths, the complete abolishment may not be done but risk of effecting and also prediction of the disease can be done. Any disease can be cured when its being detected in its early stages same way even cancer can also be cured when it is being predicted in its initial stages. But cervical cancer is difficult to predict in its early stages as it doesn't have any early symptoms. Screening is only the way that it can be predicted thus regular screening is done for the prediction of cancerous cells. The screening results might be monitored false – positive sometimes or might be delayed to avoid such risks Artificial Intelligence has been introduced in the healthcare.

Many algorithms are being used and also many more tools and techniques are used and being in use for the faster and less false – positive rate in the prediction of cancerous cells. Many researchers have predicted the cervical cancer by using different algorithms.

Shimizu et al., [17] has expressed views on approach of AI in Oncology. Artificial Intelligence (AI) works similar to the human intelligence many computational algorithms are applied for recognition, prediction, classification and many other. Machine Learning (ML) is evaluated from AI and Deep learning is a part of ML. Deep Learning is applied in medicine for the purpose of diagnosis of diseases, classification of cells based on images obtained from medical tests done to predict the type of disease. AI has drawn its own specialization in the field of Medicine which made changes in workflow of clinical approaches towards the diseases. Accuracy, specificity, sensitivity and efficiency of medical results have been increased. In upcoming decades it is expected to have computational knowledge even in Medical Sciences for more accurate results with in less time span.

Hinton et al., [8] has a belief that Deep learning has changed the health-care in the upcoming decades. Deep learning has deep networks with input, output and many intermediate layers. Deep learning deals with complicated models in the field it is being used. It extracts features for the further workflow and has successfully exceeded the human experts in outcomes with less flaws. Neural Networks add an alternative “hidden layers” which deals the problems in more efficient way. The number of hidden layers are limited in order to avoid the complexity in the developed model. As new algorithms are discovered every time the potential of applications in various fields will increase rapidly.

Geeitha et al., [6] has proposed Machine Learning models in analyzing gene expression combined with different Data Mining techniques. Also the importance of feature selection and classification techniques which classify benign and malignant cervical cells. According to the survey, it is said that Machine Learning can deal with large volumes of data and can easily provide accessibility to the detection of cancer. Comparative analysis of Machine Learning models such as hold out method, Backward Elimination Hilbert-Schmidt Independence Criterion (BAHSIC), Singular Value Decomposition Entropy gene selection (SVD Entropy), SMOTE technique, J48, Random tree forest, Boosted Decision Tree, Sequential Elimination approach and classifiers such as SVM, Naïve Bayes, Logistic Regression, KNN, ANN, MLP, WEKA segmentation classifier etc., are implemented. SVM provides higher prediction accuracy when combined with KNN and SMOTE which is developed for imbalanced data. Even though the prediction rate is higher the feature extraction and selection is a bit challenging thus researchers have decided in future to carry out the work in Deep Learning which can deal with number of large datasets with dimensionality reduction

Rayavarapu et al., [15] used two popular Machine Learning techniques Voting and Deep Neural Network (DNN) classifiers to predict growth of cancerous cells. By applying supervised algorithms prediction is done. The preprocessed data is given to the classifiers which divide the data into training and testing data. The result is given to the voting classifier based on the majority voting the class label is acknowledged. DNN is applied whereas when compared voting classifier is highly accurate. The feature extraction can be improved in extinction as DNN takes more time complexity

Devi et al., [5] expressed views on prediction of cells by applying Artificial Neural Networks (ANN) from the screening method results. ANN uses different architectures to obtain high accuracy rate in prediction and classification of cancerous cells. In future combination of Gene Algorithm with Feed Forward Neural Networks is used to predict cervical cancer cells from the pap and liquid-based cytology test images.

Kurnianingsih et al., [10] has expressed views that segmentation of cells and classification of cells play a major role in prediction of cancerous cells. In the proposed work Mask R-CNN is used for segmentation. The classification phase uses VGG-like Net. For the prior information ResNet10 is used in addition with Mask R-CNN. Future work mainly focuses on the implementation of deeper network models to increase the performance.

Athinarayanan et al., [16] used an automated detection system to find the cells from the input images produced at the time of testing the patient. Texture features are premeditated in the decision making system along with SVM. The experimental results are compared with other classifiers KNN and ANN which resulted SVM is better classifier. This proposed system helps the physician to make decision faster for further treatment of the patient as the classification of cells is done accurate.

Ghoneim et al., [7] has drawn a conclusion from previous cases that computer-aided systems are more accurate and Deep learning based systems are more efficient. Cervical Cancer prediction here is dealt with Convolutional Neural Networks (CNN) model followed with Extreme learning machine (ELM) based classifier. CNN extract features from raw input and that data is fed in to ELM-classifier, here two CNN models are used "VGG-16 Net and CaffeNet" for the efficient extraction of features and then output is fed into ELM-classifier which results the detection of cervical cancer cells.

Taha et al., [14] has focused mainly on the classification step as in the prediction of cancer classification of cells play a crucial role. Rather designing classification algorithm from scratch CNN combined with SVM is used which saves time consumption and is more accurate. CNN here is used for feature extraction as it studies the pap images layer by layer and SVM is used for the decision making purpose which helps to classify normal cells from the cancerous cells.

Ungarpalli et al., [11] performed two stage of classification first is it classifies number of cells present in the input images and the second is the class of the cell. CNN is applied for the categorization of cells. Proposed work can predict two classes from the seven classes where as in the future it is expected to predict all the seven classes also provide the cell segmentation for the all classes in the given data accurately.

### 3. Proposed Work:

#### Classification:

As images are considered in the proposed work pre-processing is the primary step performed. Classification of cells is said to be done as first step in the proposed work. The data that has been considered consists of 7 classes such dataset is partitioned into training set and validation set. The pre-trained model of CNN ResNet50 is used in the proposed work for the purpose of the classification and identification of cancerous cells from given image. Classification process under goes other sequence of steps as following:

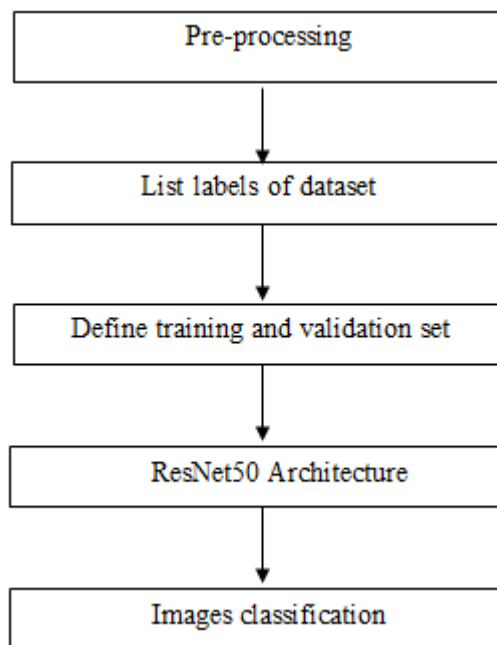


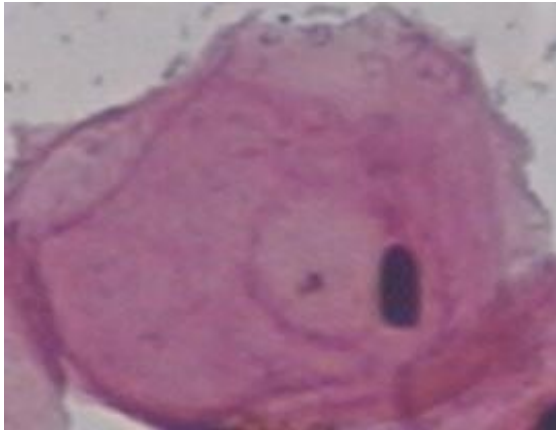
Figure 3: Classification process

#### Pap-Smear Data Collected:

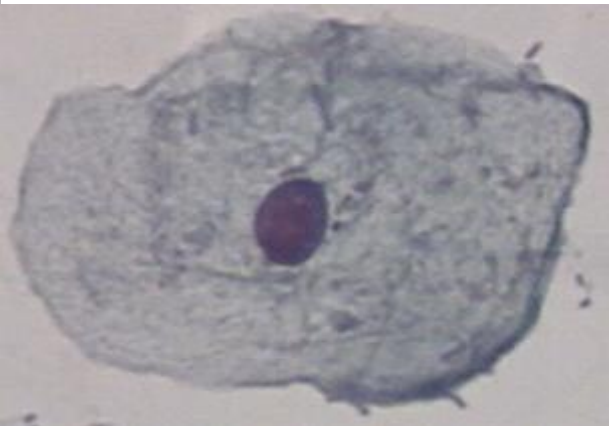
Database images are collected from the "Herlev University Hospital" database which contains 7 different classes and 20 different features. The sample cells are collected from the cervix of the human body by screening method. The screening test done is pap-smear test, it is done to identify the cancerous cells in its initial stages

which can avoid the heavy risk of cancer. The 7 classes of cells present in the dataset are partitioned as normal and abnormal. The names of the cells that are classified in the dataset are mentioned below.

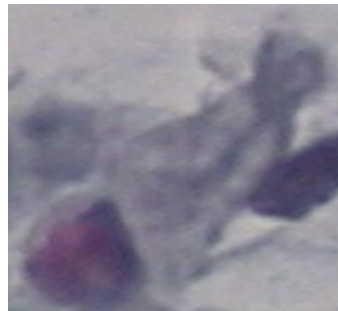
Normal Cells:



**Figure 4:** Superficial Squamous Epithelial cell

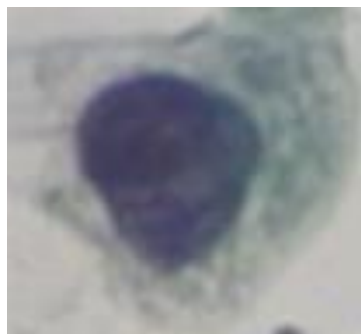


**Figure 5:** Intermediate Squamous Epithelial cell

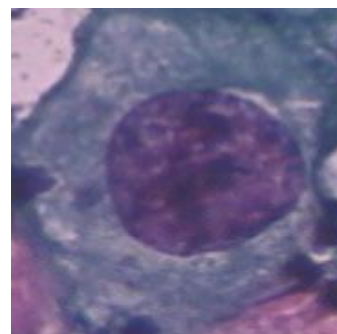


**Figure 6:** Columnar Epithelial cell

Abnormal Cells:



**Figure 7:** Mild Dysplasia cell



**Figure 8:** Moderate Dysplasia cell



Figure 9: Severe Dysplasia cell

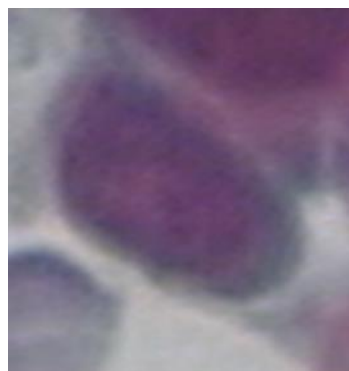


Figure 10: Carcinoma – in – situ

**ResNet50:**

Convolutional Neural Networks are hidden layers which are used for the purpose of dealing with the images. Training of these models take lesser time compared to other Machine Learning algorithms. Classifiers used here are pre-trained in advance with millions of images thus, no need to start from the scratch which optimizes the training time of the model. There are various trained models in CNN “Residual Neural Networks (ResNet)” is considered for the prediction of cervical cancer cells [18]. ResNet has many variants in it, the proposed work uses ResNet50 for the prediction. All the variants work the same but the number of layers vary. ResNet50 is used for the classification of images. The researchers have studied that compared to the plain networks residual networks are more beneficial as they provide efficient performance even they are added with large number of layers. In CNN architecture addition of layers followed one after the other causes the problem of degradation where it can be solved by introducing ResNet [9] as it solves the problem by introducing residual blocks which generates residual function which helps us adjust the input features to upgrade the high-level features.

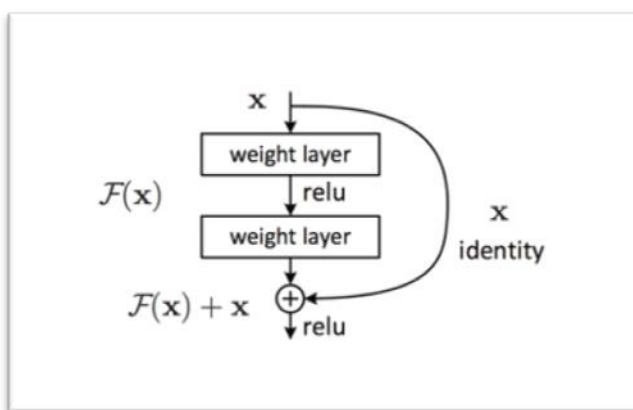


Figure 11: Structure of Residual Neural Networks (ResNet)

The identity link represents the shortcut connection, where the ResNet uses the activations from previous layers as it skips few layers which helps us to train the model faster. Similar to CNN, ResNet50 has stages in it which includes an identity block and convolution blocks. The identity block and convolution block has other 3 convolution blocks in them.

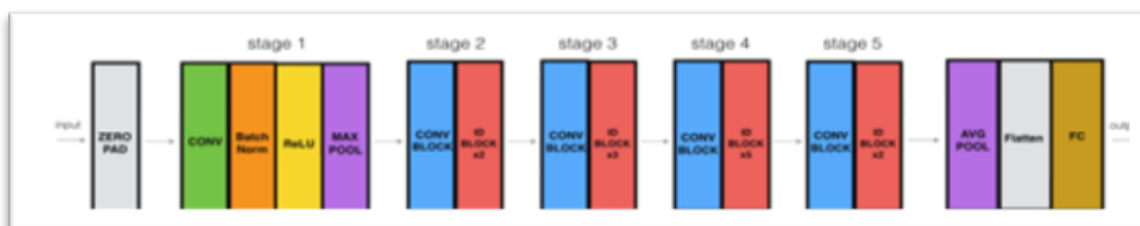
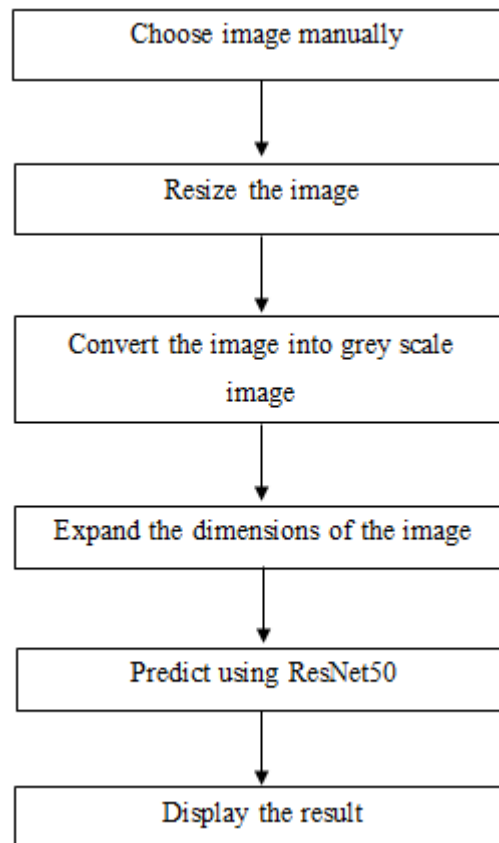


Figure 12: ResNet50 Architecture

In the proposed work ResNet50 is used by providing the pap-smear images dataset for the identification of cancerous cells from the images. As known transfer learning models are trained on a problem and are used for the similar problems. The images are considered and are feed into the proposed work for prediction. Proposed work includes training and testing set which are considered by partitioning the dataset. The number of epochs are mentioned manually and are changed if required. The test image is also fed to the proposed work manually. By using the Adam optimizer in the proposed methodology cell classification is performed. The proposed methodology uses softmax function as the activation function. Softmax function present in the output layer predicts the multinomial probability distribution. In the proposed work all the layers are considered as they are except the last layer it is trained according to the requirement. The work flow for the prediction of cells is shown in the below figure.



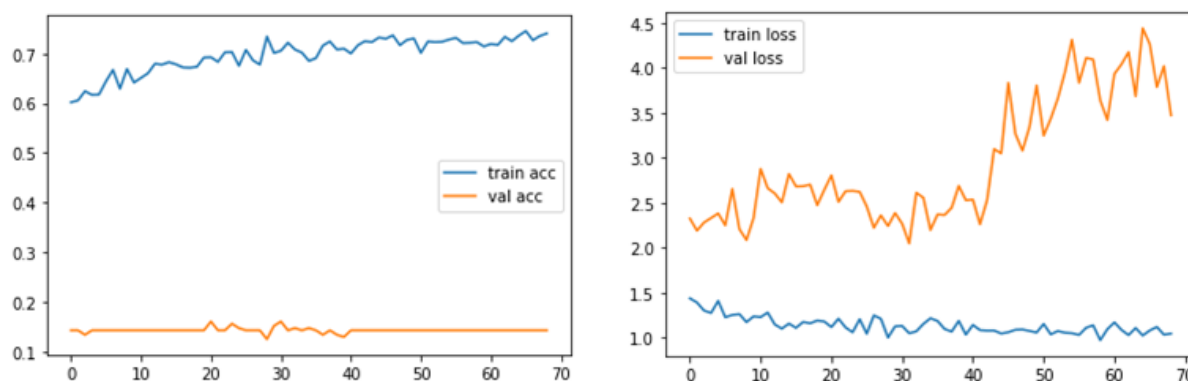
**Figure 13:** Workflow of prediction of cells

The prediction process is done only by selecting the image manually from the taken dataset. The image is resized in to particular fixed size as it is easier to extract features from the images when all are of same size. In the next step the colored images are converted to grey scale images. The training and testing sets are divided from the given dataset and the manually selected image is converted into pixels by applying the filters present in the pre-trained model the image is identified. The image and class which it belongs to is displayed as the result.

#### 4. Experimental results:

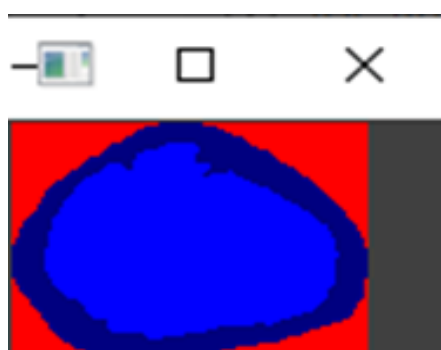
The proposed method works on the dataset containing 3 class normal and 4 class abnormal pap-smear screening images. The ImageDataGenerator class is defined in the model to increase the training set for efficient results. Preprocessing is done by converting the images into pixel values a threshold value is said to be maintained if the value obtained are higher than the threshold value they are considered as 1 and remaining as 0. Batch size is considered as 32 and optimizer used is Adam optimizer. “Categorical Cross entropy” is used as the loss function and “Softmax function” is used as activation function. The system with i3 processor is used for running the model.

The results of the proposed work is taken for different epochs are tested randomly and the accuracies vary for every epoch. The proposed work can classify and predict with an accuracy of 74.04%. The efficient results are represented in a graph below:



**Figure 14:** Efficiency of the proposed method

The result obtained also includes the image which is selected manually for the prediction of the type of class it belongs. The image is displayed in a separate window as below:



**Figure 15:** Output Image after prediction

## 5. Conclusion:

The proposed work can predict all the 7 classes present in the dataset by using “ResNet50” with an accuracy of 74.04%. Compared to CNN, pre-trained model ResNet50 provides more accurate results in prediction and classification of cells present in the given data whereas CNN provide 44% of accuracy in prediction of all the classes. The proposed method is easy to train, as transfer learning is used where it need not be trained from the scratch. The proposed method also predicts the class of the image either it belongs to carcinoma – in – situ, severe dysplasia, moderate dysplasia, mild dysplasia, columnar epithelial, intermediate squamous epithelial and superficial squamous epithelial are the abnormal and normal cell classes present in the given dataset. Thus, the proposed work provide good results by predicting the type and class of the cells. In addition, to the proposed method experiments can be done on the more advanced image dataset. Also can use high end pre-trained techniques like Inception, VGG-16, Yolo v3 etc., for efficient accuracy.

## References

1. All about Cancer, Cancer Society of Finland, Available at: <https://www.allaboutcancer.fi/facts-about-cancer/detection/#8667b054> (Accessed date 25.07.2020).
2. American Cancer Society, Cancer Facts for Women, Available at: <https://www.cancer.org/healthy/find-cancer-early/womens-health/cancer-facts-for-women.html>(Accessed date 25.07.2020).
3. American Cancer Society, What is Cervical Cancer?, Available at: <https://www.cancer.org/cancer/cervical-cancer/about/what-is-cervical-cancer.html> (Accessed date 25.07.2020).
4. Cancer.Net, ASCO.org, What are Cancer Vaccines? Available at: <https://www.cancer.net/navigating-cancer-care/how-cancer-treated/immunotherapy-and-vaccines/what-are-cancer-vaccines> (Accessed date 25.07.2020).
5. Devi, M. A., Ravi, S., Vaishnavi, J., & Punitha, S. (2016). Classification of Cervical Cancer Using Artificial Neural Networks. *Procedia Computer Science*, 89, 465–472. <https://doi.org/10.1016/j.procs.2016.06.105>
6. Geeitha, S., & Thangamani, M. (2020). A cognizant study of machine learning in predicting cervical cancer at various levels-a data mining concept. *International Journal on Emerging Technologies*, 11(1), 23–28.



7. Ghoneim, A., Muhammad, G., & Hossain, M. S. (2020). Cervical cancer classification using convolutional neural networks and extreme learning machines. *Future Generation Computer Systems*, 102, 643–649. <https://doi.org/10.1016/j.future.2019.09.015>
8. Hinton, G. (2018). Deep learning—a technology with the potential to transform health care. *JAMA - Journal of the American Medical Association*, 320(11), 1101–1102. <https://doi.org/10.1001/jama.2018.11100>
9. Jeremy Jordan, Common Architectures in Convolutional Neural Networks, Available at: <https://www.jeremyjordan.me/convnet-architectures/> (Accessed date: 12.01.2021).
10. Kurnianingsih, Allehaibi, K. H. S., Nugroho, L. E., Widyawan, Lazuardi, L., Prabuwo, A. S., & Mantoro, T. (2019). Segmentation and classification of cervical cells using deep learning. *IEEE Access*, 7, 116925–116941. <https://doi.org/10.1109/ACCESS.2019.2936017>
11. Manasa Ungrapalli, N. S., & Myna, A. N. (2019). Classification of pap smear images for cervical cancer using convolutional neural network. *International Journal of Innovative Technology and Exploring Engineering*, 9(1), 2801–2807. <https://doi.org/10.35940/ijitee.J1226.119119>
12. National Cancer Institute, Cervical Cancer Screening, Available at: <https://www.cancer.gov/types/cervical/patient/cervical-screening-pdq> (Accessed date 25.07.2020).
13. National Cancer Institute, What is Cancer? Available at: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer> (Accessed date 24.07.2020).
14. N, T. . B. D. . J. (2017). Classification of Cervical Cancer Using Pap-Smear Images: A Convolutional Neural Network Approach. *Department of Electrical and Computer Engineering*, 1(d), 698–706. <https://doi.org/10.1007/978-3-319-60964-5-23>
15. Rayavarapu, K., & Krishna, K. K. V. (2018). Prediction of Cervical Cancer using Voting and DNN Classifiers. *Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies, ICCTCT 2018*, 1–5. <https://doi.org/10.1109/ICCTCT.2018.8551176>
16. S., A., & M.V., S. (2016). Classification of Cervical Cancer Cells in Pap Smear Screening Test. *ICTACT Journal on Image and Video Processing*, 06(04), 1234–1238. <https://doi.org/10.21917/ijivp.2016.0179>
17. Shimizu, H., & Nakayama, K. I. (2020). Artificial intelligence in oncology. *Cancer Science*, 111(5), 1452–1460. <https://doi.org/10.1111/cas.14377>
18. Towards data science, A comprehensive guide to Convolutional Neural Networks, Available at: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53> (Accessed date 12.01.2021).
19. UICC Global Cancer Control, Cervical cancer elimination, Available at: [https://www.uicc.org/what-we-do/thematic-areas-work/cervical-cancer-elimination?gclid=CjwKCAiA57D\\_BRAZEiwAZcfCxZnopy3PUaLF0PmKco4esxdZ0pHfEybpWuHmJQUAQPFrtukdCFbUVhoChSsQAvD\\_BwE](https://www.uicc.org/what-we-do/thematic-areas-work/cervical-cancer-elimination?gclid=CjwKCAiA57D_BRAZEiwAZcfCxZnopy3PUaLF0PmKco4esxdZ0pHfEybpWuHmJQUAQPFrtukdCFbUVhoChSsQAvD_BwE) (Accessed date 24.07.2020).
20. World Health Organization, Cervical Cancer, Available at: [https://www.who.int/health-topics/cervical-cancer#tab=tab\\_1](https://www.who.int/health-topics/cervical-cancer#tab=tab_1) (Accessed date 24.07.2020).