

An Efficient approach for early detection of Cervical Cancer by using Classification Techniques

¹K. Shanthi, ²S. Manimekalai

¹Research Scholar, ²Head & Associate Professor

^{1,2}PG & Research Department of Computer Science, Affiliated to Thiruvalluvar University, TheivanaiAmmal College for Women (Autonomous), Villupuram
Tamil Nadu, India.

Email: softshanthi@gmail.com, mamekaroshini@gmail.com

Abstract

Cervical cancer is found to be the world's second most frequent cancer among women. It is also one of the prominent causes of mortality among women. The major goal of this research is to use several classification techniques to predict cervical cancer so that it would be possible to cure it. If cancer is discovered in its early stages, it can be entirely treated. The Pap Smear Test (PST) is a straightforward and reliable test to detect cervical cancer in its preliminary stages. It's also used to figure out what kind of cancers to deal with. The image of the Pap smear is classified using appropriate classifiers. These classifications help to find the abnormal cells and foresee the cancer cells. The prior and accurate identification of cells based on their normality and abnormality can lead us to find the chance of cancer which will further strengthen the medical imaging system to have a strong hold on disease prediction and treatment.

Keywords: Pap-Smear, Cervical cancer, Classifiers.

1. Introduction

Cancer is a well-known and recognized disease worldwide, which is known to be an illness without a cure. According to the Global Burden of Disease Cancer Alliance, 9.7 million people died due to cancer in 2017, making it the second biggest cause of death after cardiovascular illnesses [46]. Cervical cancer is a real-world problem that has become more prevalent in recent times. It has been acknowledged that about 70% of women who die from this disease have not been diagnosed yet.

The second most common cause of cancer deaths in the world is breast cancer. This disease is considered a primary and appropriate diagnostic method for women with this condition in developing countries; around 84% of cases of cervical cancer occur in the developing countries. This problem is largely due to the lack of proper screening and treatment. When the cells become uncontrollable and start reproducing with uncontrolled cell death and partition, it leads to the development of a malignant tumor. Sinuous overgrowth of these cells can cause further damage to the organs of the body. This type of cancer is known as CIN.

Human papillomavirus is a sexually transmitted infection that can affect most individuals in their lifetime. It is mainly caused by the strains of HPV that are most prevalent in males. So far, about 200 different kinds of human papillomavirus have been found. 13 and 18 are high-risk kinds. Approximately 70% of all cervical cancer cases are caused by these two. Having intercourse with several partners, weak immune system, smoking habits and using oral contraceptives are all risk factors [5]. Unusual bleeding, vaginal discharge, and pain during sexual intercourse are all common indications of cervical cancer. Cervical cancer can be treated if caught early and treated properly [13]. Cervical intraepithelial neoplasia occurs when cells in the cervical tissue begin to grow and multiply abnormally. Pap smears and liquid-based cytology are the most widely used tests to detect cervical cancer. These tests are done for women who have had an abnormal Pap smear. However, manual

screening can be very time-consuming and expensive. It involves the use of multiple slides and different orientation and overlapping cells.

Human Papillomavirus (HPV) can be used in conjunction with and as a separate Pap test. The specimen for this test is obtained from the cervix with a tiny brush. HPV testing are not performed for all types of sexually transmitted viruses. Only the strains with the highest risk of developing cancer are tested [16] [18]. Some of the articles which have the planned structured review and produce an investigational outcome are the ones that examine the effectiveness of the Pap smear test. The following articles are from the Cervical Cancer dataset, which includes approximately sixty articles. These articles are focused on evaluating the literature on the topic of cancer.

This dataset consists of articles related to Cervical Cancer that were published from 2001 to 2020. The selected articles were chosen according to the topic of the study and its future studies consequences.

The Computer Assisted Diagnosis (CAD) approach recognizes a few different aberrant morphological traits in cervical cancer precursor lesions. For clinical analysis, clinicians examine characteristics such as marginal form, the caliber of the blood carrying vessels, color, texture, capacity, inter capillary distribution and spacing, and contour [44].

The aim or the basis of this article is to provide a review of the various techniques used to classify Pap smear. The difference between a journal publication and a conference publication is often explained by the personal understanding of the author. Each publication has its own distinct characteristics. While book portions may take a long time to publish, some journal pieces have already been in the works for a long time.

2. Related Work

The method of Image analysis is utilised to cervical images for cancer detection and classification have used various classifiers and algorithms. The most common method for cervix area detection is automatic detection. The ROI contains the most relevant information for appropriate cells and disease classification. The usual technique of retrieving medical images is based on five basic steps are Image Pre-processing, segmentation, feature extraction, feature selection and classification. These are the steps and the approaches used. The multiple categorization approaches utilized for cervix cell pictures are

described in this research. The classification approach aids in determining if a cervix cell is normal or abnormal.

N. Kumaresan, et al. [1] reviewed Pap's smear cell segmentation and classification procedures for cervical cancer and found that they were more accurate. A Survey of feature extraction for Leukemia diagnosis in blood microscopic pictures was discussed by P. Aishwariya, S. Manimekalai, et al. [45]. Chen Y. F., et al. [3] presented semi-automatic Pap smear cell segmentation and classification. Cervical Cancer cells Classification with Colposcopy Images by Deep Learning and its major characteristics was described by Vasudha, Ajay Mittal, et al [7]. The detailed analysis of Pap-Smear classification of image is well established and explained by K Shanthi, S. Manimekalai, et al. [43]. Dr. P. Santhi, et al. [9] presented an efficient System for Identifying Chemical Structures and its classification with the help of SVM, which improved the accuracy.

3. Classification Techniques

The SVM classification technique was explored by F. Asadi et al. [11]. The SVM is an approach to machine learning which aids in the classification and reversion exploration of data. SVM is utilized for text categorization, handwriting recognition, and science in addition to image classification. With the use of a dual wavelength approach, a system can separate the nuclei and cytoplasm of a cervix cell in the SVM method.

Linear discrimination analysis is utilized in this smear classification technique to categorize single cells as normal and others as pathological. SVM is divided into supervised and unsupervised learning. Supervised learning is the process of clustering data based on a specified condition. Unsupervised learning is a strategy that groups data in reverse and does not apply any conditions.

QaziMudassarIlyas et al. [23] established a strategy for detecting cervical cancer from dataset using image processing. The nearest neighbor classifier is a non-parametric approach for classifying objects based on the votes of its closest neighbors. If $K=1$, a user-defined constant, the object is assigned. It's a straightforward way of classifying distinct nearest neighbors. In the feature space, the 1-nearest neighbor classifier assigns a

point to the class of its clandestine neighbor. Assign a feature space point X to which $C_n \text{Inn}(x) = Y$. (1). The weighted closest neighbor classifier can assign a weight of $1/k$ while maintaining the weight of 0. The process of providing a weight to the i^{th} closest W_{ni} , with $w_n \sum_{i=1}^n w_{ni} = 1$.

M S MinuSanjudharan et al. [13] uses a Bayes classifier for the cervical classification. The Bayes Classifier is a classification technique that employs the Bayes theorem. Because they are simple to create, this form of classifier is commonly employed in machine learning methods. This classifier is known as simple Bayes or independent Bayes since it is straightforward to implement. The Bayes classifier maintains independence between data points with attributes and uses probability theory to categorize data.

K. Deepa et al. [14] used a medical image to detect the cervical cancer and discussed the ANN classifier. ANN stands for artificial neural networks. It is a structure-based model to calculate on biological neural network functions. Non-linear statistical data modeling is used to find patterns when there is a complex relationship between input and output. Three layers have been added to the artificial neural network. The input layer is the first layer, and it is used to provide input. The second layer is the median layer, which aids in the transmission of data between neurons.

The output layer, which sends the output neurons, is the final layer. ANN offers a number of topologies that can be used to detect the cervix cell with ease and accuracy. The ANN employed the Learning Vector Quantification approach to determine the c mean value of the extracted image after storing an input image and extracting the features to identify the cancer cells responsible for cervical cancer. It is a crucial component of multi-layer perception where ANN is used.

S. Arun Rajesh et al. [15] used the MLP classifier to develop a method for automatically diagnosing cervical cancer from blood sample pictures. MLP is a type of artificial network that consists of a number of layers and nodes. A neuron in the input node uses the nonlinear activation function. The supervised learning technique is a type of back propagation of training. Feed the MLP classifier's frontward neural network, which contains three layers: one input layer, concealed

layer, and output layer. The input layer controls which nodes have which feature vector dimension, while the output layer controls the classification problem. The rule of (Inputnode) & (Outnodes) is used to keep track of the number of hidden nodes. The sigmoid activation function uses a hidden layer, while the liner activation function uses an output layer.

Defeng Liu et al. [6] devised a cervical classification method. The texture characteristics and form features with contour signature were extracted from the image using this method and Random forest is used for classification. . It is a method of ensemble learning that includes a classifier, recession, and additional tasks. RF is building numerous decision trees in the form of classes, with the class and training time output by operate. It is used to develop analytical models for both classification and regression problems. RF classifiers are used in the medical or biological fields to combine high-dimensional data.

The use of good and bad cells as per their structure and color is utilized to produce the highest quality findings in cancer detection. Using Random Forest to create a larger number of multiple binary decision trees. When growing trees, training data helps to avoid the over-fitting problem in individual decision trees. Over-fitting refers to fine-tuning the classifier to be as close as possible to the training data, reducing the classifier's precision. It's a decision tree learner that everyone has.

The Least Square Support Vector Machine is used by V. Luis Rosado et al. [21] for classification. It is a Least Square form of the SVM approach, which is a collection of supervised learning methods for identifying patterns and scrutinizing data, which aids categorization and reversal analysis. The key advantage of LSSVM classification is its quick calculation time.

4. Methodology

4.1. Support Vector Machine

The SVM aids in the classification and reversion exploration of data. SVM is utilized for text categorization, handwriting recognition, and science in addition to image classification. With the use of the dual wavelength approach, a system can separate the cervix cell into nuclei and cytoplasm in the SVM method [10]. Linear discrimination

analysis is utilized in this smear classification technique to categorize single cells as normal and others as pathological. SVMs are divided into supervised and unsupervised learning. Supervised learning is the process of clustering data based on a specified condition. Unsupervised learning is a strategy that groups data in reverse and does not apply any conditions.

4.2. Nearest Neighbor Classifier

The nearest neighbor classifier is the type of the approach of non-parametric type for classifying objects based on the votes of its closest neighbors. If $K=1$, a user-defined constant, the object is assigned. It's a straightforward way of classifying distinct nearest neighbors. [19]. In the feature space, the 1-nearest neighbor classifier assigns a point to the class of its clandestine neighbor. Assign a feature space point X to which C_n $1_{nn}(x) = Y$. (1). The weighted closest neighbor classifier can assign a weight of $1/k$ while maintaining the weight of 0. The process of providing a weight to the i^{th} closest W_{ni} , with $W_{ni}=1$ and $n_i = 1$.

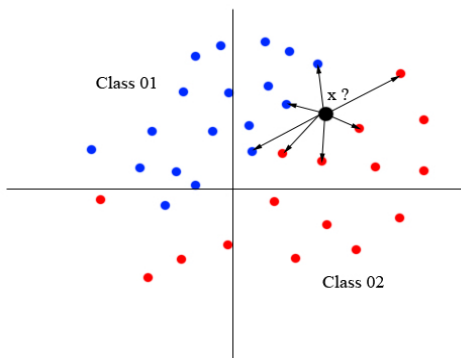


Fig.1. K-Nearest Neighbor Method

4.3. Bayes Classifier

The Bayes classifier is a classification technique that employs the Bayes theorem. Because they are simple to create, this form of classifier is commonly employed in machine learning methods. This classifier is known as simple Bayes or independent Bayes [12] because it is straightforward to implement. The Bayes classifier maintains independence between data points with attributes and uses probability theory to categorize data.

4.4. Artificial Neural Network (ANN)

ANN stands for artificial neural networks. It is the model which is entirely made on the model of structure with biological neural networks of function. Non-linear statistical data modeling is used to find patterns when there is a complex relationship between input and output. Three layers have been added to the artificial neural network. The input layer is the first layer, and it is used to provide input. The second layer is the median layer, which aids in the transmission of data between neurons. The output layer, which sends the output neurons, is the final layer. ANN offers a number of designs that may be used to recognize cervix cells with ease and accuracy [13] [27]. The ANN employed the LVQ (Learning Vector Quantification) approach to determine the c mean value of the extracted image after storing an input image and classifying the detection of cervical cancer. It is a crucial component of multi-layer perception where ANN is used.

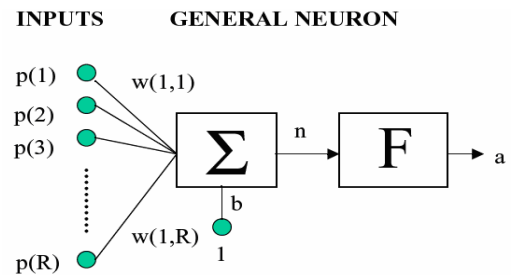


Fig. 2. Model of an Artificial Neuron

4.5. Multi-Layer Perception (MLP)

MLP is a type of artificial network that consists of a number of layers and nodes. A neuron in the input node uses the nonlinear activation function. The supervised learning technique is a type of back propagation of training. Feed the MLP classifier's frontward neural network, which contains three layers: one input layer, one concealed layer, and one output layer. The input layer controls which nodes have which feature vector dimension, while the output layer controls the classification problem. The rule of (Input node) (Outnodes) is used to keep track of the number of hidden nodes. The sigmoid activation function uses a hidden layer, while the linear activation function uses an output layer.

4.6. Random Forest Classifier

The Random forest is a method based on ensemble learning method which includes classifiers, regressions, and other tasks. RF is building numerous decision trees in the form of classes, with the class and training time output by operation. The system is to develop analytical models for both classification and regression problems. RF classifiers are used in the medical or biological fields to combine high-dimensional data. It will make use of the variation dissimilarities of normal & abnormal cells to deliver the reliable results and findings for cancer detection. Using Random Forest to create a larger number of different binary decision trees [6]. When growing trees, training data helps to avoid the over-fitting problem in individual decision trees. Over-fitting refers to fine-tuning the classifier to be as close as possible to the training data, reducing the classifier's precision. It's a decision tree learner that everyone has.

5. Result and Discussions

This research examines the various classification strategies and investigates the many classifiers used to detect cervical dysplasia. This classifier aids in the classification of Pap smear cells as normal or abnormal.

Classification Method	Reference No	Accuracy
Support Vector Machine	[11]	94
Nearest Neighbor Classifier	[15]	82.9
Bayes Classifier	[13]	87
Artificial Neural Network	[14]	94.3
Multi-Layer Perceptron	[15]	92.03
Random Forest	[25]	93

Table 1. Shows a comparison of different classification methods.

The cervical classification systems depending on the accuracy of classifier are described in this section. Number of papers are considered in this study for the investigation of cervical classification methods. Below the graph is

based on the detailed evaluation of accuracy of different classifiers.

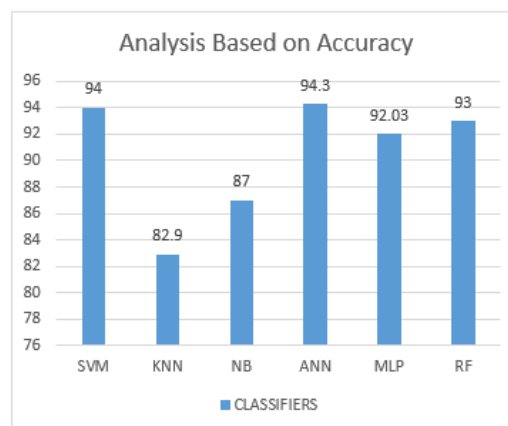


Fig. 3. Analysis Graph Based on Accuracy

6. Conclusion

The area of cervical cancer and how cervical dysplasia can be detected was explored in this paper. Cervical cancer can be detected in its early stages with the Pap smears test technique of cervical screening. Preprocessing, segmentation, extraction, selection, and classification are all common procedures in medical image processing. The most crucial goal of the picture retrieval technique is to determine the classifier. This paper aids in the investigation of numerous classification methods, and when compared to these strategies, the ANN classifier provides greater accuracy for future work. The efficient utilization of ANN can be blended accurately with current medical imaging systems to provide the additional accuracy and reliability of obtained results.

7. References

- [1] N. Kumaresan and D. Somasundaram, (2018), Review of Pap Smears Cell Segmentation and Classification Techniques for Cervical Cancer Analysis, *Ethno Med*, Vol.12, pp. 96-105.
- [2] Bora K., Chowdhury M., Mahanta L. B., Kundu M. K., Das A. K.,(2017), Automated classification of Pap smear images to detect cervical dysplasia, *Computer Methods and Programs in Biomedicine*, 138, pp. 137-145.
- [3] Chen Y. F., Huang P. C., Lin H. H., and Wang L. E., (2014), Semi-automatic segmentation and classification of pap smear cells, *Biomedical and Health Informatics, IEEE Journal of*, 18, pp. 94– 108.
- [4] Sakthivel K, Jayanthiladevi A ,Kavitha C, (2016), Automatic detection of lung cancer nodules by employing intelligent fuzzy c means and support vector machine, *Biomedical Research, Special Issue*, PP: : S123-S127.
- [5] Kangkana Bora, Lipi B Mahanta, Anup Kumar Das,(2018), Fuzzy NSCT based Feature Extraction Method for Automated Classification of Pap Smear

- images, *International Journal of Applied Engineering Research*, Vol.13, pp. 6709-6716,
- [6] Defeng Liu.,(2021), Optimization and evaluation of random forest model in the efficiency prediction of chemo radiotherapy for advanced cervical cancer based on randomness signature from high resolution T2 weighted images, *Springer*, Vol. 303, pp. 811-820.
- [7] Vasudha, Ajay Mittal, MamtaJuneja,(2018), Cervix Cancer Classification using Colposcopy Images by Deep Learning Method, *International Journal of Engineering Technology Science and Research*, Vol.5, Issue 3, pp. 2394 – 3386.
- [8] Ling Zhang, Le Lu, Senior Member, IEEE, Isabella Nogue, Ronald M. Summers, Shaoxiong Liu, and JianhuaYao,(2018), DeepPap: Deep Convolutional Networks for Cervical Cell Classification.
- [9] Dr.P. Santhi, K.Deepa,(2017), Classification System for Identifying the Chemical Structure Using Support Vector Machine, *International Journal of Emerging Trends in Science and Technology*, Vol.3, No.1, pp. 2349- 7866.
- [10] S. Thilagamani, N. Shanthi, (2011), Object Recognition Based on Image Segmentation and Clustering, *Journal of Computer Science*, Vol. 7, No.11, pp.1741-1748,.
- [11] Umashankar G., (2018), Classification of cervical cancer from MRI images using Multiclass SVM classifier, *IJET*.
- [12] Seeja K., Shweta, Microarray (2011). Data Classification Using Support Vector Machine, *International Journal of Biometrics and Bioinformatics*, Vol. 5, pp. 10-15.
- [13] M. S MinuSanjidharan.,(2019), Cervical cancer prediction using Naïve Bayes Classification, *Journal of Science and Technology*.
- [14] Cristiana Braga.,(2019),A review of computational method for cervical cells segmentation and abnormality classification, *International journal of molecular science*, Vol. 20.
- [15] Meenakshi Sharma.,(2016), Classification of clinical dataset of cervical cancer using KNN, *Indian journal of science and technology*, Vol. 9.
- [16] HemalathaKulala., (2017), Improvrment of Multi-Layer Perceptron classification on cervical Papsmear data with feature extraction.
- [17] NazimRazali.,(2020), Risk factor of cervical cancer using classification in data mining, *Journal of physics*.
- [18] M S MinuShanjudharan, (2019),Cervical cancer prediction using Naive Bayes classification.
- [19] WR. D. Aldian, E. Purwanti and M. A. Bustomi,(2013), Applied Computing Based Artificial Neural Network for Classification of Cervical Cancer, pp. 4-7.
- [20] P. Sobrevilla, E. Montseny, F. Vaschetto, and E. Lerma,(2010), Fuzzy-based analysis of microscopic color cervical Pap smear images: nuclei detection, *International Journal of Computational Intelligence and Applications*, vol. 9, no. 3, pp. 187–206.
- [21] N. M. Harandi, S. Sadri, N. A. Moghaddam, and R. Amirfattahi,(2010), An automated method for segmentation of epithelial cervical cells in images of ThinPrep, *Journal of Medical Systems*, vol. 34, no. 6, pp. 1043–1058.
- [22] M. Wang, X. Zhou, F. Li, J. Huckins, R. W. King, and S. T. C. Wong,(2008), Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy, *Bioinformatics*, vol. 24, no. 1, pp. 94–101.
- [23] Kangana Bora., (2017),Automated classification of pap Smear image to detect cervical Dysplasia, *Computer Methods and Programs in Biomedicine*.
- [24] Luis Rosado.,(2019), A Review of computational methods for cervical cells Segmentation and Abnormality Classification, *MDPI*, Vol. 20, pp. 811-820.
- [25] Mamta Arora.,(2021), Improved performance of machine learning algorithm for prognosis of cervical cancer, Vol. 6, pp. 191-205.
- [26] Zhang J, Liu Y, (2014), Cervical Cancer Detection Using SVM Based Feature Screening. In: *Proceeding of the 7th International Conference on Medical Image Computing and Computer Aided Intervention (MICCAI'04)*, Saint-Malo, France, pp. 873–880.
- [27] K Deepa, Dharani.,(2021), A Journal on cervical cancer prediction using ANN, *Turkish Journal of Computer and Mathematics Education*, Vol. 12, pp. 1085-1091.
- [28] QaziMudassarIlyas.,(2021) An Enhanced ensemble diagnosis of cervical cancer: A pursuit of machine intelligence towards sustainable health, *IEEE*.
- [29] Ghoneim, A. Muhammad, G. Hossain, M.S, (2020) Cervical cancer classification using convolutional neural networks and extreme learning machines. *Future Generation, Computer System*, Vol. 102, pp. 643–649.
- [30] S. K. S. P. A. V. M. M. Sharma, (2016), Classification of clinical dataset of cervical cancer using KNN, *Indian Journal of Science and Technology*, vol. 9, no. 28, pp. 1-5.
- [31] S. Khraya, Shika Agarwal, (2014), Naive Bayes Classification: A Probabilistic Detection Model for breast cancer, *International Journals of Computer Applications*.
- [32] N. T. U.A. ThanatipChankong,(2014), Automatic cervical cell segmentation and classification in Pap smears, *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2.
- [33] H.S. JieSu,(2016),Automatic Detection of Cervical Cancer Cells by a Two-Level Cascade Classification System, *Analytical Cellular Pathology*, may 2016.
- [34] P. K Malli.,(2017), Machine learning Technique for detection of Cervical Cancer using k-NN and ANN, *International Journal of Emerging Trends & Technology in Computer Science*.
- [35] Y. Ling Zhang, (2017),DeepPap: Deep Convolutional Networks for Cervical Cell Classification, *IEEE*.
- [36] K. D. KangkanaBora,(2017),Automated classification of Pap smear images to detect cervical dysplasia, *Computer Methods and Programs in Biomedicine*, vol. 138, pp. 31-47.
- [37] M Anousouya Devi, S.Ravi , J.Vaishnavi and S.Punit,(2016),Classification of cervical cancer using Artificial Neural networks, *Twelfth international Multi conference on information Processing*, pp. 465-472.
- [38] AabhaPhatak, DrAshwini.B, (2015), Classification of MR Images of Cervical Cancer Using SVM and ANN, *International Journal of scientific research*, pp.3-6.
- [39] B.Ashok, Dr.Aruna,(2016),Comparison of Feature selection methods for Cervical cancer diagnosis using SVM classifier, *Journal of Engineering Research Applications*, Vol.6, Issue 1, pp.94-99.
- [40] Marina R.Plissiti.,(2011), Automated detection of cell Nuclei in Pap Smear Images Using Morphological Reconstruction and Clustering, *IEEE Journal of Transaction on Information Technology in Biomedicine* , Vol.15, No 2, pp.233-241.

- [41] F Asadi., (2020), Supervised Algorithm of machine learning for the prediction of cervical cancer, Journal of Biomedical, pp. 513-522.
- [42] Chandraprabha R, Dr.Seema Singh, (2016), Artificial Intelligent system for Diagnosis of cervical cancer :a brief review and future, International Journals of Latest Research in Engineering and Technology, pp.38-41.
- [43] K. Shanthi, S. Manimekalai. (2021). A Survey of Feature Extraction for Pap-Smear Image Classification. REVISTA GEINTEC-GESTAO INOVACAO E TECNOLOGIAS , 11 (4), 3468-3476.
- [44] K. Shanthi, Dr. S. Manimekalai. (2021). An Efficient Cervical Image Segmentation Method Using Lagrange Dual Dictionary Learning in Convolutional Neural Network. Annals of the Romanian Society for Cell Biology, 25(2), 1944–1957.
- [45] P. Aiswariya, S. Manimekalai. (2021). A Survey on Feature Extraction for Leukemia Detection in Blood Microscopic Images. REVISTA GEINTEC-GESTAO INOVACÃO E TECNOLOGIAS , 11 (4), 3477-3484.
- [46] P. Aiswariya, Dr. S. Manimekalai. (2021). Global and Local Entropy Based Segmentation Model for Detecting Leukemia in Blood Images. Annals of the Romanian Society for Cell Biology, 25(2), 1914–1926.