

# Improvement of Multi Layer Perceptron Classification on Cervical Pap smear data with Feature Extraction

K. Hemalatha<sup>1</sup>, Dr. K. Usha Rani<sup>2</sup>Research Scholar, Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, Andhra Pradesh, India<sup>1</sup>Professor, Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, Andhra Pradesh, India<sup>2</sup>

**ABSTRACT:** Artificial Neural Network (ANN) is an effective technique of Soft Computing can model Computer-Aided Diagnosis (CAD) system efficiently. CAD system is an essential for the prediction of Malignancy in Cervical Cancer. Cervical Cancer can be cured if it is diagnosed in early stages. Hence, for the effective screening of cancer lesions in the Cervical cell images which are captured using Pap smear test the successful ANN structure Multi Layer Perceptron (MLP) is used in this study. MLP network is trained with Cervical Pap Smear images database with original features and then only with extracted features. Classification performance of MLP in the two cases is calculated and analyzed with the help of network measures such as Classification Accuracy, Recall, Precision, Mean Squared Error (MSE) and Time.

**KEYWORDS:** Artificial Neural Network, Computer-Aided Diagnosis, Multi Layer Perceptron, Cervical Pap smear image, Classification.

## I. INTRODUCTION

Soft Computing is one of the important Artificial Intelligence (AI) approaches which simulates the remarkable abilities of the human mind. Soft Computing is a collection of computing tools and techniques to handle the uncertain and imprecision situations. The major techniques in Soft Computing are: Artificial Neural Network (ANN), Fuzzy Logic (FL) and Genetic Algorithms (GA) [1]. ANN plays a vital role in the modeling of Medical Diagnosis System for several diseases. In contrast to traditional modeling methods which deal with linear techniques, ANN can process complex and non-linear data patterns successfully [2]. Because of the successful implementation of ANN in pattern recognition it is well suitable for Computer Aided Disease Diagnosis [3]. Computer Aided Diagnosis is significant in Cancer detection for effective screening of cancer lesions.

Cervical Cancer is the deadliest cancer occurs in woman ranked after Breast Cancer. It can be preventable if precancerous lesions are screened effectively in early stages then it can be treated adequately. Pap smear test is the simple and popular screening test to find out the Cervical Cancer [4]. Pap smear test will be conducted to the victims of Cervical Cancer by collecting the tissues from the cervix for further diagnosis. The collected tissues are subjected to dyeing process to prepare the slide of Cervical cells. Pathologists examine the Cervical cytology image cells through the microscope to find out whether the Cervical cell is Malignant or not [5]. The histo pathological images of Cervical cells on Pap smear are presented in figure 1.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

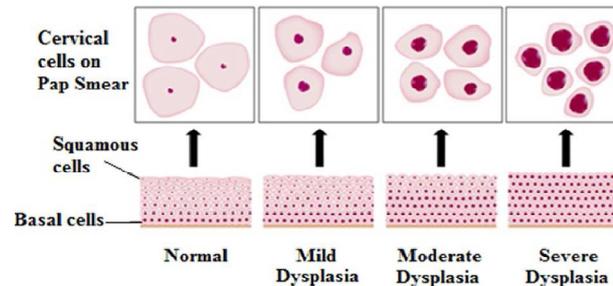


Fig. 1. Cervical Cytology cell images on Pap smear [6]

Figure 1 displays the cervical cytology cell images captured from the digital camera of the microscope. It presents the structure of the cervix cells in the specimen collected and the structure of cervical cells presented on the Pap smear slide.

The manual diagnosis of Cervical Pap smear images is time consuming. Technical and human errors may occur. Due to the limitations in the manual diagnosis Computer-Aided Diagnosis is necessary for efficient Cervical Cancer diagnosis. Hence, ANN is used to produce accurate results. The Cervical Pap smear images are classified into normal and abnormal classes using ANN.

This Paper is organized as follows: Section II describes review of work done related to classification of Cervical Cancer data using ANN. Details of the materials and methods used are given in Section III. Section IV presents experimental results showing Cervical Cancer data classification. Finally, Section V presents conclusion.

## II. RELATED WORK

Extensive studies have been carried out for the early diagnosis and classification related to Cervical Cancer using Soft Computing Techniques. In this section few studies related to the classification of Cervical Pap smear data having different features using ANN are presented.

Kangkana Bora *et.al* [7] proposed an intelligent system for automatic categorization of Pap smear images to detect Cervical dysplasia. In this system the classification is carried out based on the shape, texture and color features of two generated databases of Cervical Cancer: Herlev University database, Ayursundra Healthcare Pvt. Ltd. and Dr B. Borooah Cancer Institute database.

Edwin Jayasingh Mariarputham *et.al* [8] proposed Nominated Texture Based Cervical Cancer (NTCC) Classification system to classify the public image database of Herlev University Hospital, Denmark, with 917 Pap smear images into seven classes: Normal squamous, Normal Intermediate squamous, Normal Columnar, Abnormal Mild dysplasia, Abnormal Moderate dysplasia, Abnormal Severe dysplasia, Abnormal Carcinoma in situ. The NTCC system extracted the texture features of the Pap smear images and trained with SVM classifier. The results obtained in this study revealed that there is no unique set of feature suitable for all classes.

Siti Noraini Sulaiman *et.al* [9] extracted four features: Nucleus Size, Cytoplasm Size, Nucleus's grey level, Cytoplasm's grey level using feature extraction algorithm based on pseudo-coloring technique. Several new image processing algorithms to improve the performance of the Neural pap system are proposed. The Adaptive Fuzzy-k-Means (AFKM) Clustering algorithm is proposed to replace the Moving k-Means (MKM). Finally the extracted data set is classified by the Hierarchical Hybrid Multi Layered Perceptron (H<sup>2</sup>MPLP) network.

N. A. Mat-Isa *et.al* [10] proposed diagnosis confidence percentage and diagnosis confidence level analysis to strengthen the HMLP network performance. In this study Nucleus Size, Cytoplasm Size, Nucleus's grey level, Cytoplasm's grey level features of cervical cancer cells is used to classify the data into normal, Low-grade Squamous Intraepithelial Lesions, (LSIL) and High-grade Squamous Intraepithelial Lesions (HSIL).

### III. METHODOLOGY

#### Multi Layer Perceptron (MLP):

The Multi Layer Perceptron is the commonly used and efficient neural network for classification problems. It is a feed forward type network consists of consecutive input layer, hidden layers and output layer. The input layer task is to receive the information. This information will be processed and then transmitted to the hidden layer. The hidden layer performs the necessary intermediary computations and sends the results to the output layer [12]. The general structure of MLP is represented in the figure 2.

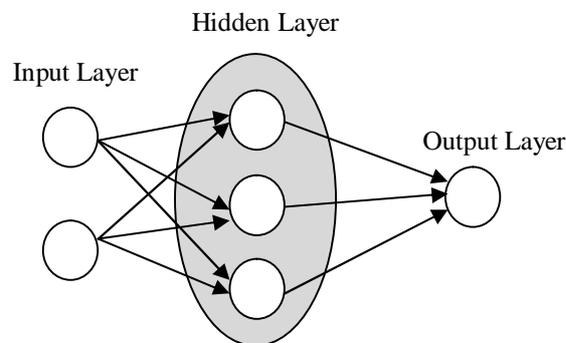


Fig. 2. General Structure of MLP

The MLP general structure consists of an input layer, a hidden layer and an output layer. The neurons presented in the input layer transmit the information to hidden layer. The number of hidden layers may change based on the problem considered. However one hidden layer is enough to solve simple problems. The resultant information obtained from the hidden layer will be transmitted to an output layer to produce the output. MLP uses the back propagation algorithm for training. The network weights and biases are initialized to a very small value and during the training process they will be adjusted to minimize the Mean Squared Error (MSE) [13].

#### Data Collection:

Cervical Pap smear images database of Herlev University Hospital, Denmark which is publicly available in web [11] with 917 Pap smear images with 20 features and seven classes is considered in this study. The resolution of the images captured with the help of microscope is  $0.201 \mu\text{m}/\text{pixel}$ . Two expert Cyto-technicians classified the images into seven classes: Superficial squamous, Intermediate squamous, Columnar, Mild dysplasia, Moderate dysplasia, Severe dysplasia, Carcinoma in situ [14]. These seven classes of images are presented in figure 3.

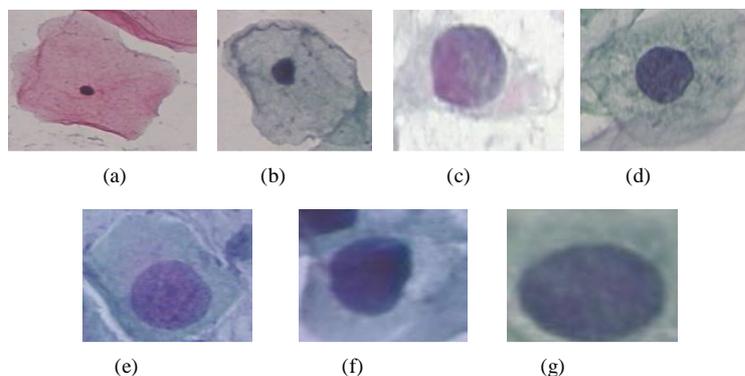


Fig. 3. Sample image of each class of Cervical cell. (a) Superficial squamous (b) Intermediate squamous (c) Columnar (d) Mild dysplasia (e) Moderate dysplasia (f) Severe dysplasia (g) Carcinoma in situ

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

The sample images of cervical cells from each class of Herlev database is shown in the above figure. The sample images are the RGB images obtained from the glass slide containing Pap smears. The first three classes are comes under the normal cells category whereas the remaining four classes are considered as the cancerous or abnormal cells. The list of 20 features of the Herlev database is mentioned in the table 1.

TABLE 1. LIST OF FEATURES IN CERVICAL PAP SMEAR DATASET

Feature No	Description	Feature No	Description
1	Nucleus area	11	Cytopl. longest diam
2	Cytoplasma area	12	Cytopl. Elongation
3	N/C ratio	13	Cytopl. Roundness
4	Nucl. Brightness	14	Nucl. Perimeter
5	Cytopl. Brightness	15	Cytopl. Perimeter
6	Nucl. shortest diam	16	Nucl. relative position
7	Nucl. longest diam.	17	Maxima in Nucl.
8	Nucl. Elongation	18	Minima in Nucl
9	Nucl. Roundness	19	Maxima in Cytopl
10	Cytopl. shortest diam	20	Minima in Cytopl

These twenty features mentioned in the above table are the morphological features such as size, shape and textural features of Cervical Pap smear cells. These features are extracted from the nucleus and cytoplasm regions of the Cervical cell images to identify the normal and cancerous cells.

### Feature Extraction:

In Image processing, a feature is information chosen from the image. The feature may be Morphological feature or a second order feature depends on the intensity of the image. Feature Extraction reduces the amount of details required to classify a large data [5]. Several researchers classified the Cervical Cancer data using the significant four features: Nucleus Size, Cytoplasm Size, Nucleus's grey level and Cytoplasm's grey level which proved the improvement in the classifier performance. Some of the studies are presented in section.II. In our previous work [15] the Cervical Pap smear images of Herlev database are segmented into Cytoplasm and Nucleus regions to extract the above said four important features by using a different approach i.e. Fuzzy Edge Detection Method and a new Modified Cervical Pap smear (MCPS) Dataset is generated. The generated dataset is also experimented with MLP.

## IV. EXPERIMENTAL SETUP AND RESULTS

Matlab R2015a tool is used to perform the experiments. In this study, first the Herlev dataset of 917 Cervical Pap smear images with 20 features mentioned in Table 1 is used to train the most commonly used and successful ANN structure i.e. MLP. MLP network classified the dataset into normal and abnormal classes. The MLP network is constructed with an input layer of 20 neurons as the number of features is 20, a hidden layer with 10 neurons (default) and an output layer with a neuron.

Later the same experiment with extracted 4 features i.e. MCPS dataset is done to train the MLP. In this case the MLP network is constructed with an input layer of 4 neurons as the number of features is 4, a hidden layer with 10 neurons (default) and an output layer with a neuron.

Classification Accuracy, Recall, Precision, MSE and Time parameters are calculated to predict the performance of MLP in two experiments. Accuracy refers to the degree of correctness of the output value. Recall represents the number of relevant instances from the retrieved data. Precision indicates the number of retrieved instances that are

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

relevant to the data. MSE is the average squared difference between outputs and targets. Lower values are better. Zero means no error.

To analyse the MLP network performance the above said measures of both the experiments are considered and tabulated in Table 3.

TABLE 3. NETWORK PERFORMANCE MEASURES OF MLP WITH TWO EXPERIMENTS

Experiment	MLP with No. of Features	Network Measures				
		Classification Accuracy (%)	Recall (%)	Precision (%)	MSE	Time (sec)
Experiment-1	20	85.05	78.94	60.72	0.0627	0.67
Experiment-2	4	<b>92.03</b>	78.09	<b>90.43</b>	0.0616	<b>0.265</b>

From the results presented in the table 3, it is observed that the Classification Accuracy and Precision are high when MLP is trained with extracted 4 features than the original features of the dataset. Recall value is reduced in experiment-2 when compared with experiment-1 value. Similarly time taken for the classification is also reduced obviously in experiment-2. The same is represented graphically in figure 4.

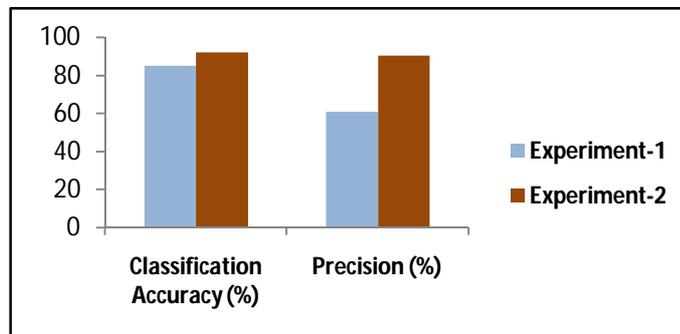


Fig. 4. Comparison of Classification Accuracy and Precision values of MLP.

From figure 4 the increment in classification accuracy and precision values of MLP can be observed clearly in experiment-2. In experiment-1, both the classification accuracy and the precision values are less when compared with the values of experiment-1.

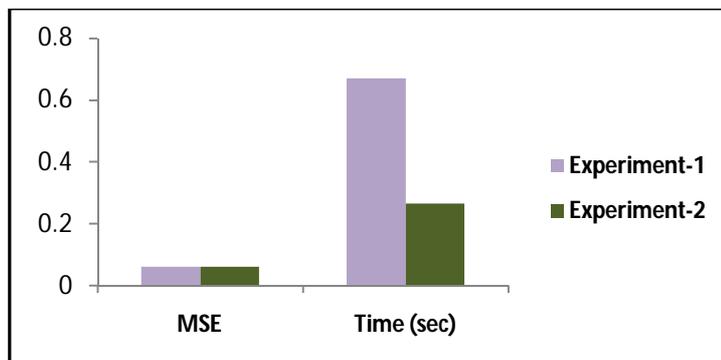


Fig.5. Comparison of MSE and Time values of MLP.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

The comparisons of MSE and time values of MLP in two experiments are represented in figure 5. From Figure 5 it can be observed that the Time taken for the classification in experiment-2 is less than the time taken by the MLP with experiment-1. The MSE is considerably low in both the experiments which is desirable.

## V. CONCLUSIONS

MLP Network which was proved a successful ANN structure by previous researchers is experimented to classify the Cervical dataset into normal and abnormal classes. Feature Extraction is the dimensionality reduction approach which can extract the prominent features of high-dimensional datasets. In this study Cervical Pap smear images with 20 features as first case and 4 extracted features as second case are considered for the study of MLP classification performance. The network performance measures in both the cases are calculated and compared. The results present that MLP perform well with 4 features than original features with an improvement of 6.98% classification accuracy and 29.71% of precision. Improvement of accuracy indicates the increase in the degree of correctness of the output. Whereas the precision enhancement signify the increment in the number of retrieved instances that are relevant to the cervical data. Hence, MLP classification performance is improved with Cervical Pap smear database by considering the extracted four important features rather than twenty features of original dataset.

## REFERENCES

- [1] Amit Konar, "Artificial Intelligence and Soft Computing Behavioral and Cognitive Modeling of the Human Brain", CRC Press, 2000.
- [2] Abraham Pouliakis, Efrossyni Karakitsou, Niki Margari, Panagiotis Bountris, Maria Haritou, John Panayiotides, Dimitrios Koutsouris and Petros Karakitsos, "Artificial Neural Networks as Decision Support Tools in Cytopathology: Past, Present, and Future", *Biomedical Engineering and Computational Biology*, 7, pp 1–18, 2016. doi:10.4137/ BECB. S31601.
- [3] Babak Sokouti, Siamak Haghypour, Ali Dastranj Tabrizi, "A framework for diagnosing cervical cancer disease based on feed forward MLP neural network and ThinPrep", *Neural Comput & Applic*, 24, pp. 221–232, 2014.
- [4] B. Ashok, Dr. P. Aruna, "Comparison of Feature selection methods for diagnosis of cervical cancer using SVM classifier", *Int. Journal of Engineering Research and Applications*, Vol. 6, Issue 1, (Part - 1) , pp.94-99, January 2016.
- [5] G. Karthigai Lakshmi and K. Krishnaveni, "Feature Extraction and Feature Set Selection for Cervical Cancer Diagnosis", *Indian Journal of Science and Technology*, Vol 9(19), May 2016.
- [6] Prof. S.Maheswari, K.Jayasudha, R.Revathy, K.Yogalakshmi, "Predicting the Severity of Cervical Cancer Using Image Processing Techniques", *International Journal for Research in Applied Science & Engineering Technology*, Volume 3 Issue I, January 2015.
- [7] Kangkana Bora, Manish Chowdhury, Lipi B Mahanta, Malay Kumar Kundu, Anup Kumar Das, "Automated classification of Pap smear images to detect cervical dysplasia", Preprint submitted to *Computer Methods and Programs in Biomedicine* August 19, 2016.
- [8] Edwin Jayasingh Mariarupham and Allwin Stephen, "Nominated Texture Based Cervical Cancer Classification", *Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine* Volume 2015.
- [9] Siti Noraini Sulaimana, Nor Ashidi Mat-Isab, Nor Hayati Othman, Fadzil Ahmada, "Improvement of Features Extraction Process and Classification of Cervical Cancer for the Neural Pap System", *Procedia Computer Science* 60, pp. 750 – 759, 2015.
- [10] N. A. Mat-Isa, M. Y. Mashor, N. H. Othman, "Classification of Cervical Cancer Cells Using HMLP Network with Confidence Percentage and Confidence Level Analysis.", *International Journal of The Computer, The Internet and Management*, Vol. 11, No.1, pp. 17 – 29, 2003.
- [11] MDE management and decision engineering laboratory-Pap smear Herlev database <http://labs.fme.aegean.gr/decision/downloads>
- [12] Zribi M, Boujelbene Y, "The Neural Networks with an Incremental Learning Algorithm Approach for Mass Classification in Breast Cancer", *International Journal of Biomedical Data Mining*, Vol 5(1), 2016.
- [13] Shiva Kumar · P, Srinivasa Pai · B, R. Shrinivasa Rao, G. S. Vijay, "Prediction of performance and emission characteristics in a biodiesel engine using WCO ester: a comparative study of neural networks", *Soft Computing*, Vol 20, pp 2665–2676, 2016.
- [14] Dalwinder Singh, Dr. Amandeep Verma, "Computer Aided Segmentation And Classification Of Cervical Cells", July 2015.
- [15] K.Hemalatha, Dr.K.Usha Rani, "Feature Extraction of Cervical Pap Smear Images using Fuzzy Edge Detection Method", *IC3T International Conference, Vijayawada*, 3-5 November, 2016. (Accepted for publication in AISC Series, Springer Publications, January,2017)