International Journal of Computer Science & Communication (ISSN: 0973-7391)
Special Issue, Page 60-66, December 2020 ------ available at www.csjournals.com

Prediction of Cervical Cancer with Ontology Based Deep Learning Approach

Divakar H R[1]
Asst. Professor, Dept. of MCA,
PES College of Engineering,
Mandya.

Dr. D Ramesh[2]
Professor & Head, Dept. of MCA,
Sri Siddhartha Institute of Technology,
Tumkur

Dr. B R Prakash[3]
Asst. Professor,
Govt. First Grade College,
Tumkur.

Abstract: Cervical cancer has been one of the world's most common diseases for women. There have been some issues such as pathologists' shortages which lead to an increase in the burden. It includes an appropriate and accurate procedure for diagnosing cervical cancer without human involvement. This research introduces an integrated method that is built with ontology driven methods and neural networks to diagnose cervical cancer. MATLAB deep learning toolbox is often used to classify the derived features from cytology images used to analyze different cancer stages. Core-to-cytoplasm relationship, shape and color strength along with a nucleus region, perimeter and eccentricity are dominant features used for diagnosis. This paper employs Scale Conjugate Gradient (SCG) Training Algorithm and reports the efficiency of 95.6% using deep leaning approach.

Keywords: Cervical cancer, deep learning, ontology, SCG

I. INTRODUCTION

Cervical cancer occurs in women, along with bosomal disease, one of the world's most commonly known tumours. In most cases, moderately elderly women between the ages of 40 and 55 are affected by this disease. Continuous cervical cancer has been extensively studied in around 50000 women and is responsible for more than 280000 deaths per year [2]. There is a wide difference in the cases of cervical disease around the globe these days. Risk factors which include with cervical is smoking, unsafe sex or HIV disease, poor use of contraceptive drugs. On the western side, the prevalence of this disease is slowly decreased by routine screening and early diagnosis. This disease is highly found in developing countries, 80 percent of the new cases are evidently found which counts for one quarter of the cervical conditions worldwide annually.

The human papillomavirus (HPV) virus is the main cause for this type of cancer. Cervical cancer develops if irregular cells in the cervix grow and become out of control more rapidly. The irregular modifications caused by the cervical cells turn it into a precarious condition known as 'Cervical Intraepithelial Neoplasia' (CIN). Such conditions are categorized as low CIN and high CIN depending on their intensity or degree [5].

The Government of India's National Cancer Control Plan (NCCP), initiated and funded by the Ministry of Health, emphasized the implementation of Community cervical detection programmes. The NCCP has given support to all countries for the implementation of a screening system for cervical cancer for cancer prevention [12].

There are two specific screening tests, which help to diagnose or prevent cervical cancer, (i) Pep tests are designed to look for changes in cervical pre-cancer cells. (ii) HPV check identify the cell-shifting HPV virus.

Another traditional screening method is liquid-based cytology (LBC). LBC is a means of preparing cervix samples to analyse and diagnose at laboratories. The rate of detection for LBC use is higher than the Pap testing [7]. Both of these methods mentioned are time-consuming and did not produce the wrong results.

This paper offers an important and knowledgeable method for the diagnosis of cervical cancer by the use of cytological tissue image analysis. The cell types are shown in figure1:

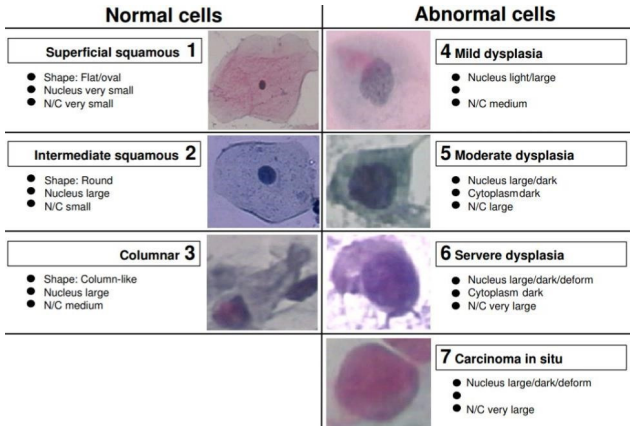

Figure1. Types of cells

II. RELATED WORK

Addressing morphological grading of cervical cell identifiers [1] which can be used to create a software program which can assist physicians in monitoring cervical cancer. It provides an unmonitored genetic algorithm which is based on the Fischer genetic operator model for cell part separation. Co-included are several algorithms to quantify the morphological properties of cells required for identification of the disease (ratio core / cytoplasm, cellular deformity). Each cell identified has characteristic vectors for the input into the popularity stage of the automatic data analysis method for trailing cervical cancer.

A new approach for classifying the various malignancies in cervical images with acoustic shadowing was introduced by screening and then detection using acoustic shadowing. This approach used a Support Vector Machine (SVM) classification to identify cancer stages and help the pathologist diagnose cancer better. The image proposed was tested and proved effective (2013) with a set of images. The multiple extraction features discussed by the Gaussian Cervical Cytological image mixture model [2]. This approach differentiates between the nucleus and cytoplasm from single and multiple cellular pap smear cervical cytology images using the Gaussian Mixture Model. A research was undertaken in order to develop a system of automatic image grading through the regional classification of interests (ROI) [4]. Details of the patient's cuelcro are extremely difficult to identify irregularities and use only the form, size and gray level.

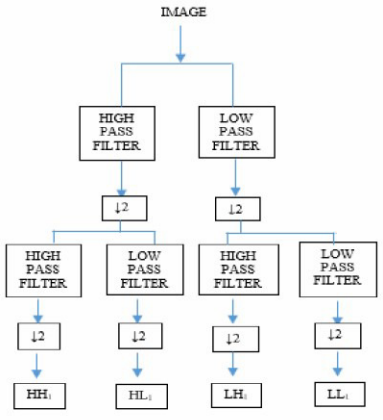

Figure2. Feature Extraction phase

By extracting and classifying features of the discussed image histology of cervical cancer [5]. This study provides an automated grouping by CIN category of vertical segmented epithel regions. The approach developed includes the determination of the medial axis, the determination of bounding boxes and the partitioning of the whole epithelium region into multiple vertical segments with medial axis consideration. Features will then be produced and carried out in all experimental procedures. The image of epithelium histology provides the finding for different methods of fusing classification data for each segment or for an alternate form of classification weight updating in neural network.

The detection of cervical cancer by thresholding and segmentation of watershed [6], this offers a realistic, effective algorithm for cervical cancer detection. The picture of the histogram has been further segmented by rim detection for the exact edge and shape of the tumour surface, with magnetic resonance imaging scans of cervical cancer. It is also determining the functional anatomical orientation and its effect on other cervical regions. MRI scans are used and processed as a .jpg in MATLAB in cervical cancer patients. Photographs will then turn into a gray picture based on the distribution of the pixel intensity of the image. To find edges of the tumour region, edge detection is applied to the image. A hybrid separation is carried out on the sampled image, i.e. thresholds and watershed segmentation are combined to eliminate the tumour-infected region. Then, an extracted image is used in order to determine exact tumour measurements using morphological information processing techniques.

Using Automatic cervical cancer identification by Radiating Gradient Vector Flow (RGVF) segmentation and SVM classification this Segmentation and classification of cervical cells can be considered as a key task for robust automated pap smear analyzes. This method suggested a system for automatic slide analysis with single cell paper smear. The RGVF snake method is used in segmentation phases into three regions. The SVM model is used to achieve 93,78% precision and 98,96% sensitivity and 96,69% specificity in the classification process [7].

A discussion on a photonic crystal-based bio-sensor has immediately tackled cervical cancer identification. The idea used an open source application for the use on a Linux operating system, MIT Electromagnetic Equation Propagation (MEEP). They also developed a two-dimensional crystal photonic bio-sensor for cervical cancer detection. They also developed a MATLAB application, which extracts the sensor's output attributes and classifies the data automatically [8].

The visual recognition algorithms for cervical cancer diagnosis is employed here to incorporates numerous classification strategies including SVM, Fuzzy processes, cell identification strategy and hierarchical clusters to research and analyze automated cervical cancer identification [9]. Automatic diagnostic support program has been developed for cervical cancer through image analysis. It suggested automatic screening approaches for cervical cancer to support cytopathologists, as pathologists were missing. The successful removal of nuclei is important for the automated diagnostic support program. Firstly, cells must be divided in blue and red cells as cells are coloured with various colours, based on the cell type. The nuclei are then separated into mega pixels by segmentation. At last, we classify malignant cells and distinguish between positive and negative stimuli by nuclear enlargement and nuclei colour density. The result was a 97% positive rate of discrimination and a 55% negative rate of discrimination [10].

Automatic screening and monitoring of cervical cancer cells by images method were developed using an

ordinary approach to scan cervical cancer cells and eliminate errors by pathologists. The Pap smear picture is first pre-processed for colour adjustment with reference pictures to enhance the image profile in this article. Use the classification method k-means to specify the location and measure the field of the nucleus. For the measurement of the cytoplasm area, the cell was identified by geometric rotational reference to its position and edge profile. Eventually it became possible to quantify the area of the nucleus to cytoplasm and to show anomalous cells. Such cells have been registered for further diagnosis in medical records. For individual cell measurements, the precision of the imaging test was up to 79 percent. In case of cells without folding or combining, the sensitivity was 60% and the specificity 100% [11].

III. METHODOLOGY

Data Collection

The Herlev University Hospital has a series of single pap-smear images from a wide variety of glass slides. Skilled 0.201 m resolution cytotechnicians = pixel to grab the optical pictures by microscopic means of the individual cells. Then every cell picture is classified into 7 cell types manually. Separate cytotechnicians identify it twice for clarification. The screenshot will be removed if the clarification is unfavourable. The data set is divided accordingly

1) ***Normal Cells - 242 cells***
   - Superficial squamous epithelial,74 cells.
   - Intermediate squamous epithelial, 70 cells.
   - Columnar epithelial, 98 cells.

2) ***Abnormal Cells- 675 cells***
   - Mild squamous non-keratinizing dysplasia, 182 cells.
   - Moderate squamous non-keratinizing dysplasia, 146 cells.
   - Severe squamous non-keratinizing dysplasia, 197 cells.
   - Squamous cell carcinoma in situ intermediate, 150 cells.

The pre-processed data is collected by identifying twenty significant features from the image dataset and it is taken from Kaggle data repository. The twenty significant features and its abbreviations are as follows.

Table 1: Features of cell

Sl. No.	Feature	Name abbreviation
1	Nucleus area	Narea
2.	Cytoplasm area	Carea
3.	N/C ration	N/C
4.	Nucleus brightness	Ncol
5.	Cytoplasm brightness	Ccol
6.	Nucleus shortest diameter	Nshort
7.	Nucleus Longest diameter	Nlong
8.	Nucleus elongation	Nelong
9	Nucleus roundness	Nround
10	Cytoplasm shortest diameter	Cshort
11	Cytoplasm longest diameter	Clong
12	Cytoplasm elongation	Celong
13	Cytoplasm roundness	Cround
14	Nucleus perimeter	Nperim
15	Cytoplasm perimeter	Cperim
16	Nucleus position	Npos
17	Maxima in nucleus	Nmax
18	Minima in nucleus	Nmin
19	Maxima in cytoplasm	Cmax
20	Minima in cytoplasm	Cmin

Ontology Model

The development of ontology is one of the profound process to indentify the associations among attribute features which are considered for this study. The main aspect of the generation of ontology is to acquire values for all the significant features those are determined for this dataset. The ontology is constructed using Protege 4.3 ontology construction application and its reliability is verified using FaCT++ and HermiT reasoners.

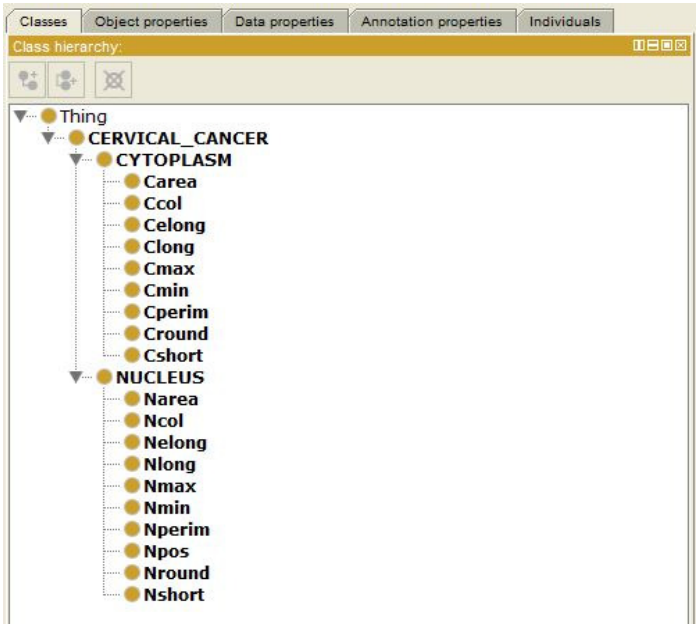

Figure 3. Class hierarchy of features of cervical cancer

The figure3 depicts asserted class hierarchy of generated ontology for the features considered for cervical cancer. The important attributes those are considered for this study are significant features identified by Nucleus and Cytoplasm of cancer cell. The figure4 shows the asserted class and figure5 represents the CCOWL tree structure produced by portege 4.3.

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Special Issue, Page 60-66, December 2020 ------ available at www.csjournals.com

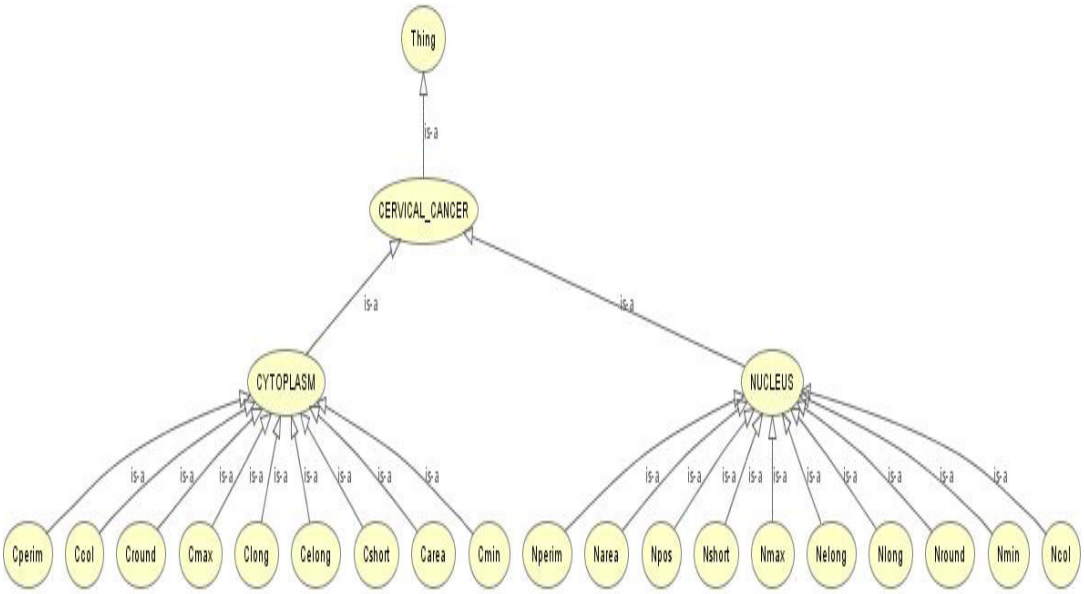

Figure 4. Asserted class hierarchy for cervical cancer features.

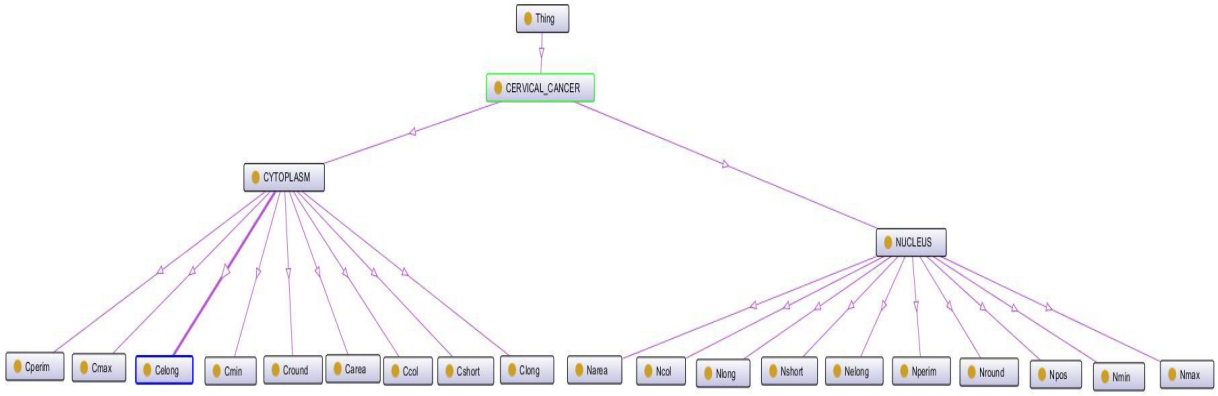

Figure 5. CCOWL tree structure produced by portege 4.3.

Deep neural network algorithms

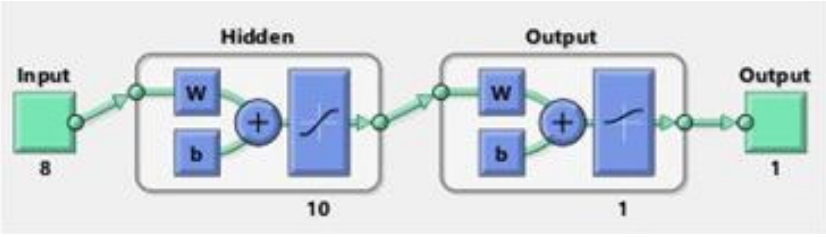

Figure 6. Deep **neural network**

The figure 6 sums up the techniques utilized in the proposed fill in as a model. This shows the progression of examination study directed and delineates the progression of work experienced.

Scale Conjugate Gradient (SCG) Training Algorithm

The Scale conjugate gradient algorithm is the idea combines the model trust region approach, known from Levenberg-

Marquardt algorithm with the conjugate gradient approach introduced by MØller [13]. This method is describes as in equation 1. Where s is the Hessian matrix approximation, E is the total error function and È is the gradient of E, scaling factors λ_k and σ_k are introduced to approximate the Hessian matrix and initialized by user at the beginning of the algorithm such that $0 < \lambda_k < 10\text{-}6$ and $0 < \sigma_k < 10\text{-}4$. For SCG, β_k factor calculation and direction of the new search can be described in equation 2 and equation 3.

$$s_k = \frac{\dot{E}(w_k + \sigma_k p_k) - \dot{E}(w_k)}{\sigma_k} + \lambda_k p_k \quad \text{Equation 1}$$

$$\beta_k = \frac{(|g_{k+1}|^2 - g_{k+1}{}^T g_k)}{g_k{}^T g_k} \quad \text{Equation 2}$$

$$p_{k+1} = -g_{k+1} + \beta_k p_k \quad \text{Equation 3}$$

Design parameters for each iterations are updated independently by user, which is one of crucial process for the success of the algorithm to get the efficient result.

IV. RESULTS AND DISCUSSION

The deep learning model using the scale conjugate gradient algorithm uses cervical cancer dataset derived from Herlev University Hospital and preprocessed data from kaggle data repository are fed in to this model as input with all twenty considerable features. The trained supervised deep neural network predict output matched with 1, depicts the presence of cervical cancer cells and 0 depicts as healthy cells. The dataset contains 917 samples consisting of twenty attributes, in 971 sample 641 samples are considered for training, 138 samples for validation and 138 samples for testing, it resembles 70% for training, 30% each for validation and testing.

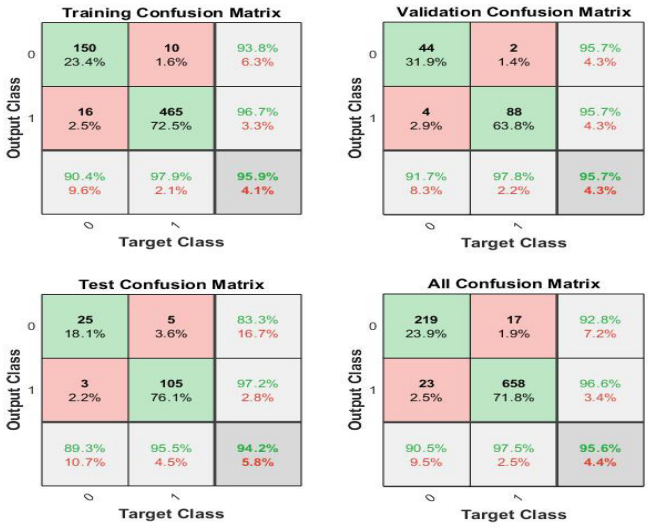

Figure 7. Confusion matrix of obtained using BP-SCG.

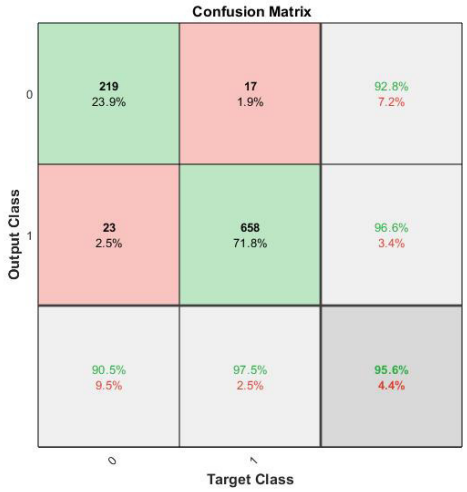

Figure 8. Overall Confusion matrix of obtained using BP-SCG

The figures 7 and figure 8 depicts the confusion matrix obtained from BP-SCG deep neural network where it reports 95.9%, 99.7% and 94.2 accuracy in training, validation and testing processes. Finally it reports 95.6% accuracy for overall data.

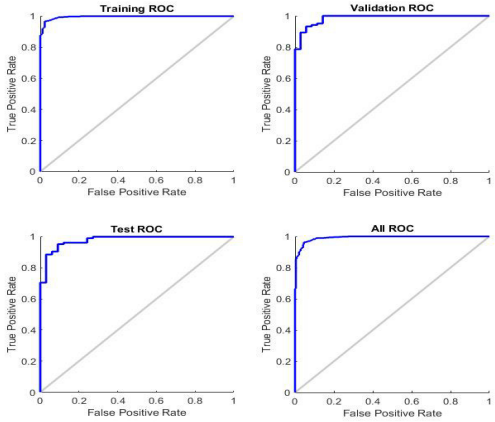

Figure 9. ROC of Scaled Conjugate Gradient (trainscg)

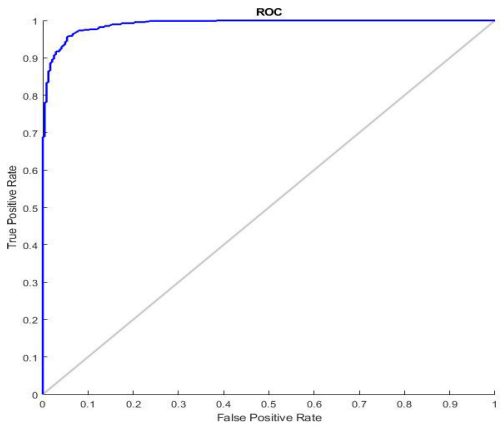

Figure 10. Overall ROC of Scaled Conjugate Gradient (trainscg).

An ROC (receiver operating characteristic) curve is a graph showing the performance of a classification model at all classification thresholds. ROC curve can be used to select a threshold for a classifier which maximizes the true positives is 0.956, while minimizing the false positives as shown in figure 9 and figure 10. ROC curves also give us the ability to assess the performance of the classifier over its entire operating range.

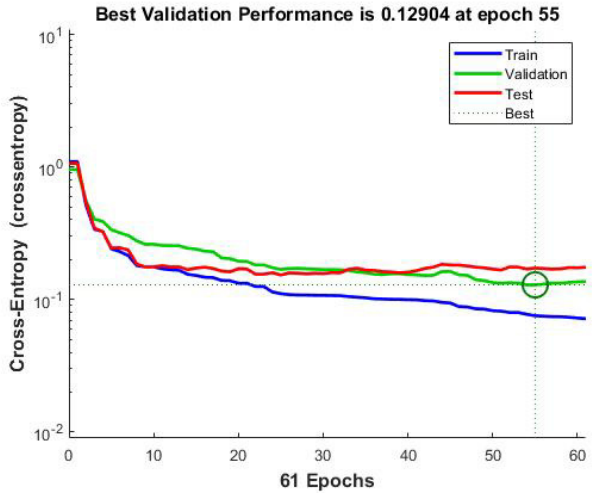

Figure 11: Performance plot of BP-SCG for classification

Mean square error (MSE) is a network performance function and measures the network's performance according to the mean of squared errors. The performance plot in figure 11 shows the value of the performance function versus the iteration number. It reports the best validation performance of 0.12904 at 55th epoch for BP-SCG deep learning algorithm.

Error histogram of figure 12 shows the distribution of errors for the training, validation and test subsets for BP-SCG deep neural network algorithms for 15 hidden layers.

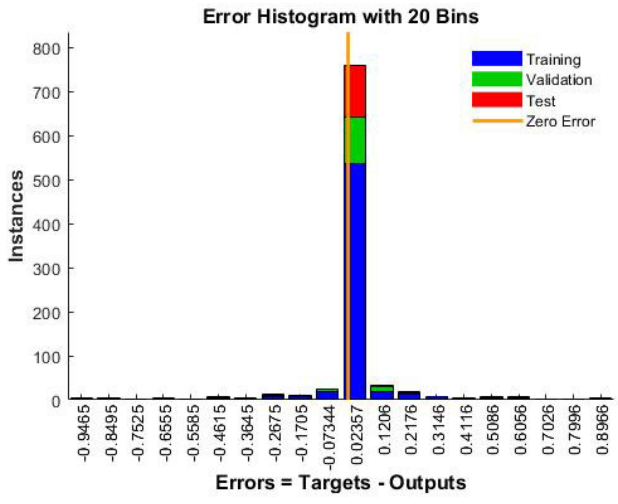

Figure 12: Error Histogram of BP-SCG

V. CONCLUSION

The prediction of cervical cancer occur in female patients is one of the major problem in field of medicine to recognise and identify the disease accurately. In this regard, we develop a sell defined model which forecast the occurrence of cervical cancer by considering the specific twenty features which are drawn from dataset available in kaggle repository. Experiments are conducted on available dataset and experimental results determine the accuracy of 95.6% using scaled conjugate gradient algorithm. This work can be unmitigated and enhanced for the automation of prediction of cervical cancer employing deep learning algorithms.

REFERENCES

[1]. N Lassouaoui, L Hamami, and N NNouali, Morphological Description of Cervical Cell Images for Pathological Recognition, Proceedings of world academy of science, engineering and technology, 2005.

[2]. N.SakthiPriya, Cervical Cancer Screening and Classification Using Acoustic Shadowing, International Journal of innovative research in computer and communication engineering, 2013.

[3]. G Karthigai Lakshmi, K Krishnaveni, Multiple feature extraction from cervical cytology images by gaussian mixture model, world congress on computing and communication technologies, 2014.

[4]. Shipra Roy, R.P Chauhan, G.K. Verma,Cervical Cancer Detection from MR Images based on multiresolution wavelet analysis, 11th International conference on industrial and information systems (ICIIC), 2016.

[5]. Peng Guo, Cervical cancer histology image feature extraction and classification, Missouri university of science and technology, 2014.

[6]. SetuGarg, ShabanaUrooj, RituVijay,Detection of Cervical Cancer by Using Thresholding & Watershed Segmentation, 2nd International Conference on Computing for Sustainable Global Development (INDIACom), 2015.

[7]. Sajeena TA, Jereesh A S, Automated Cervical Cancer Detection through RGVF segmentation and SVM Classification, International Conference on Computing and Network Communications (CoCoNet), 2015.

[8]. Nithin S, Poonam Sharma, Vivek M, PreetaSharan,Automated Cervical Cancer Detection using Photonic Crystal based Bio-Sensor, IEE International Advance Computing Conference (IACC), 2015.

[9]. M. Anousouya Devi, S. Ravi,J. Vaishnavi, SPunitha, Detection of Cervical Cancer using the Image Classification Algorithms, Procedia Computer Science, 2016.

[10].Ayaka Iwai, Toshiyuki Tanaka, Automatic Diagnosis Supporting System for Cervical Cancer using Image Processing, 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 2017.

[11].ManasSangworasil, ChayanisaSukkasem, Suvicha Sasivimolkul, PhitsiniSuvarnaphaet, SueijitPechprasarn, RujiradaThongchoom, MetiniJanyasupab, Automated

Screening of Cervical Cancer Cell Images, 11thViomedical Engineering International Conference (BMEiCON), 2018.

[12].ParthaBasu, Debjani Chowdhury, Cervical cancer Screening and HPV Vaccination: A Comprehensive Approach to Cervical Cancer Control, Indian Journal of Medical Research, 2009.

[13].Martin Fodslette Møller, "A scaled conjugate gradient algorithm for fast supervised learning, Neural Networks," Volume 6, Issue 4, 1993, Pages 525-533.