Screening of Cervical Cancer by Artificial Intelligence based Analysis of Digitized Papanicolaou-Smear Images

Rajeev Gupta1, Abid Sarwar2, Vinod Sharma3

ABSTRACT

Introduction: Cervical cancer is one of the main causes of female mortality due to cancer in the world. Researchers have suggested that by regular cervical cancer screening its incidence and mortality a can be reduced by about 84%. Study aimed to apply multiple artificial intelligence based algorithms on the task of screening of cervical cancer and compare their working efficiency and to select the best performing algorithm and use that for developing intelligent software tool that can aid in the prognosis of cervical cancer.

Material and methods: Fifteen different machine learning algorithms were implemented for screening of cervical cancer by analysis of digitized Papanicolaou-smear images using a primary dataset for training. Slides of Pap smear test of about of about 200 clinical cases were collected from multiple health care institutions in north India and images were captured using a digital microscope. Images of 8091 cervical cells were obtained from these slides, which were quantitatively profiled and calibrated to prepare a primary database that could reflect the morphological features of the cells of cervix. All the algorithms were trained using this database and were tested using a subset of unseen cells. The generalizing ability of algorithms was tested and its performance was evaluated against the human pathologist using percentage of correctly classified cells and sensitively.

Results: Almost all the algorithms performed well for identifying the cells infected by cancer by classifying the cells according to Bethesda system of classification. Among all the algorithms used, multiple backpropagation neural networks presented an efficiency of about 78.0%, whereas the efficiency of all other algorithms was in the range of 69% to 76%.

Conclusion: The results demonstrate that the artificial intelligence based techniques can effectively be used for developing tools for mass level screening of cervical cancer.

Keywords: Artificial Intelligence, Artificial Neural Networks, Cervical cancer, Medical Database, Pap-smear test.

INTRODUCTION

Timely and precise diagnosis of cervical cancer is an important real-world medical problem. Cervical cancer has turned into one of the primary causes of mortality among women around the globe. This being the fourth most common women cancer in the world, researchers throughout the world are gravely concerned for finding a solution to it so that an early and timely diagnosis of this disease could be realized and as such the overall mortality rate could be mitigated. The total number of cases of cervical cancer in 2012 reported by World Health Organization (WHO) was 528000 and the number of deaths equal to 266000. Most of these cases about 84% occur in underdeveloped and developing countries as against the developed countries; the reason being reduced access to proper screening, treatment and follow-up services.1 Cervical cancer is a tumor, malignant in nature which occurs when the cervical cells begin to multiply with uncontrolled cell division and death. Since these cells are uncontrolled and unmanaged, they start multiplying enormously and get transformed into lump of mass called tumor. These cancerous cells can infect the other parts of body too by flowing through the blood stream. The infected cells are called as intraepithelial neoplasia or CIN. Usually the CIN remains constant and is purged by the response of the immune system, but in very small proportion of cases, if CIN is not timely treated, it grows to become a cancer. Scientific investigation has confirmed that CIN results from human papillomavirus (HPV) which is usually transmitted sexually. In the initial stages Cervical cancer can usually doesn’t show-up any symptoms. As the cancer progresses to become an invasive cancer, symptoms start to appear in the form of abnormal vaginal discharge, abnormal vaginal bleeding, pain during vaginal intercourse. For women who have had menopause, new vaginal bleeding may occur. Papanicolaou test also called as Pap-Smear test has been a primary test used for screening of cervical cancer since past many decades and in many developed countries where Pap-Smear test is regularly done has also lead to steep lowering of the number of diagnosed cases.2,3 The main aim of this test is to report any unusual developments in that may transform into cervical dysplasia. In many developed countries regular screening through pap-smear test is recommended to women who have a history of having sex with several partners. After obtaining the sample, it is stained using Papanicolaou technique, which helps to differentiate the cervical cells from other bodily secretions. Research in past many years has witnessed an increasing use of Computer based intelligent techniques for solving complex problems in the domain of medical science such as Liver Pathology, Perinatology, Urology, Thyroid disorders, Oncology, Gynecology, Cardiology, etc.4-10 The main aim of using Artificial Intelligence techniques in medical science is development of such computer based intelligent systems as can support human experts in making more precise and timely diagnosis. Development of such diagnostic systems has also led to a considerable advancement in our understanding of clinical expertise, transformation of such expertise into abstract models, and further conversion of those abstract models into...
result oriented diagnostics systems. So present study aimed to apply multiple artificial intelligence based algorithms on the task of screening of cervical cancer and compare their working efficiency and to select the best performing algorithm and use that for developing intelligent software tool that can aid in the prognosis of cervical cancer.

**MATERIAL AND METHODS**

The performance of any machine learning system is largely dependent on the data with which it is trained. It is said that any human behavior can be copied into a machine provided that it is trained using a suitably precise database. The database used in this work to train the algorithms is made-up of 8,091 cervical cells which have been profiled on the basis of 40 morphological properties. Out of these 40 morphological attributes 19 belong to cytoplasm, 19 belong to nucleus, one attribute consists of the ratio of area of Nucleus and Cytoplasm, and one last attribute contains the cells label depicting as to which bethesda class the cell belong to. The data for study i.e slides of Pap-smear test were obtained from three leading health care institutions from different geographical locations in northern India viz Government Medical College, Jammu, Acharaya Shree Chandra College of Medical Science and Hospital, Jammu and Nijjer Pathology Laboratory, Amritsar. These slides were observed under different levels of magnifications under a microscope with a high resolution digital camera attached to it (NIKON Nikon Eclipse E400 DS-F12). Images of the area of interest on the slides were captured at 40x magnification followed by multiple rounds of image preprocessing subroutines to enhance visibility of individual discrete cells. Once the cells were distinctly visible, the individual cells were cropped off from cells clusters and unique-ids were assigned to them so as to identify each cell distinctly in the database. Once a considerable set of cervical cells were obtained, these cells were analyzed, inspected and tagged by multiple pathologists so as to ensure the accuracy and consistence in the database. Figure 1 show the block diagram of how the clinical data was transformed into a digital database.

For profiling of the cells so obtained, as open source software “CellProfiler” was used which evaluated the cells in the basis of 39 morphological features (figure-1). These profiled measurements of the cells were then arranged in a spreadsheet in such a way that each tuple represented a cervical cell. Various other properties of database were also integrated so that digital data could be easily correlated with the corresponding clinical case. A detailed analysis of morphological features is pre-requisite for any automated cervical screening system as such the features set chosen for quantitative evaluation have to be such that they should be precisely measurable in specific time with good accuracy. Computer techniques based on various image processing algorithms provide simple and proficient formulation of morphological features as compared to the manual interpretation by cyto-pathologists which are subjective, laborious and many times prone to human errors.

**Algorithms used and there implementation**

The algorithms considered were Random subset space, Decorate, Ensemble of Nested dichotomies also called

---

**Figure-1:** Block diagram of various steps showing how digital database was prepared
END, Back propagation artificial neural network, J48 graft, Radial basis function network (RBF), Random Committee, Decision Table, PART, Naïve Bayes, Multi-class classifier, Filtered Classifier (FC), Rotation Forest, Random forest and Bagging. These Algorithms were selected due to their reported excellent performance in recently published literature. Bagging partitions the dataset into several subsets which are the used to train several models and the final output model is obtained by cumulative results from all the training models. In random forest, multiple decision trees are made and for each input tuple one path is traced in each if the decision trees constructed and this is repeated for all the training tuples. The final classification class for an instance is considered by taking the votes from all the decision trees.

Random subspace is a general form of random forest algorithm, in random forest, decision trees are used as base classifiers whereas in random subspace any type of classifier can be used as base classifier. The system of nested dichotomies recursively transforms the multiclass categorization problem into a relatively easy two class categorization problem which can easily be represented using binary trees. The resultant classification is obtained by aggregating the results obtained by the individual binary trees. In random committee, an assembly of individual multiple tree classifiers is made, each of is trained and tested with a common dataset but using a different randomness generating seed value. The DECORATE algorithm iteratively adds a new learning classifier to the current ensemble at each iteration. The data used for training is build by adding some artificial data to the original data. A new classifier is added only if on addition of new classifier, the total error is reduced. Rotation forest works by creating an ensemble or classifiers each trained using a different subset of attributes from the main dataset. J48-Graft creates a special decision tree which is grafted by a technique that adds fresh nodes on the places of either the old leaf nodes or in between the leaf nodes such that the accuracy in prediction is improved. Naïve Bayes classifier is based on the theory of conditional independence, which assumes that each attribute has its independent roll in the prediction of the final outcome and is also independent of the presence or absence of other attributes. The multiclass classifier works on the principal of divide and conquer technique and transforms the complex problem into simple two class problem which can be solved using one verses one path is traced in each if the decision trees constructed and this is repeated for all the training tuples. The final classification class for an instance is considered by taking the votes from all the decision trees.

The DECORATE algorithm iteratively adds a new learning classifier to the current ensemble at each iteration. The data used for training is build by adding some artificial data to the original data. A new classifier is added only if on addition of new classifier, the total error is reduced. Rotation forest works by creating an ensemble or classifiers each trained using a different subset of attributes from the main dataset. J48-Graft creates a special decision tree which is grafted by a technique that adds fresh nodes on the places of either the old leaf nodes or in between the leaf nodes such that the accuracy in prediction is improved. Naïve Bayes classifier is based on the theory of conditional independence, which assumes that each attribute has its independent roll in the prediction of the final outcome and is also independent of the presence or absence of other attributes. The multiclass classifier works on the principal of divide and conquer technique and transforms the complex problem into simple two class problem which can be solved using one verses one path is traced in each if the decision trees constructed and this is repeated for all the training tuples. The final classification class for an instance is considered by taking the votes from all the decision trees.

Multiple tree classifiers is made, each of is trained and tested with a common dataset but using a different randomness generating seed value. The DECORATE algorithm iteratively adds a new learning classifier to the current ensemble at each iteration. The data used for training is build by adding some artificial data to the original data. A new classifier is added only if on addition of new classifier, the total error is reduced. Rotation forest works by creating an ensemble or classifiers each trained using a different subset of attributes from the main dataset. J48-Graft creates a special decision tree which is grafted by a technique that adds fresh nodes on the places of either the old leaf nodes or in between the leaf nodes such that the accuracy in prediction is improved. Naïve Bayes classifier is based on the theory of conditional independence, which assumes that each attribute has its independent roll in the prediction of the final outcome and is also independent of the presence or absence of other attributes. The multiclass classifier works on the principal of divide and conquer technique and transforms the complex problem into simple two class problem which can be solved using one verses one path is traced in each if the decision trees constructed and this is repeated for all the training tuples. The final classification class for an instance is considered by taking the votes from all the decision trees.

The system of nested dichotomies recursively transforms the multiclass categorization problem into a relatively easy two class categorization problem which can easily be represented using binary trees. The resultant classification is obtained by aggregating the results obtained by the individual binary trees. In random committee, an assembly of individual multiple tree classifiers is made, each of is trained and tested with a common dataset but using a different randomness generating seed value. The DECORATE algorithm iteratively adds a new learning classifier to the current ensemble at each iteration. The data used for training is build by adding some artificial data to the original data. A new classifier is added only if on addition of new classifier, the total error is reduced. Rotation forest works by creating an ensemble or classifiers each trained using a different subset of attributes from the main dataset. J48-Graft creates a special decision tree which is grafted by a technique that adds fresh nodes on the places of either the old leaf nodes or in between the leaf nodes such that the accuracy in prediction is improved. Naïve Bayes classifier is based on the theory of conditional independence, which assumes that each attribute has its independent roll in the prediction of the final outcome and is also independent of the presence or absence of other attributes. The multiclass classifier works on the principal of divide and conquer technique and transforms the complex problem into simple two class problem which can be solved using one verses one path is traced in each if the decision trees constructed and this is repeated for all the training tuples. The final classification class for an instance is considered by taking the votes from all the decision trees.

Multiple tree classifiers is made, each of is trained and tested with a common dataset but using a different randomness generating seed value. The DECORATE algorithm iteratively adds a new learning classifier to the current ensemble at each iteration. The data used for training is build by adding some artificial data to the original data. A new classifier is added only if on addition of new classifier, the total error is reduced. Rotation forest works by creating an ensemble or classifiers each trained using a different subset of attributes from the main dataset. J48-Graft creates a special decision tree which is grafted by a technique that adds fresh nodes on the places of either the old leaf nodes or in between the leaf nodes such that the accuracy in prediction is improved. Naïve Bayes classifier is based on the theory of conditional independence, which assumes that each attribute has its independent roll in the prediction of the final outcome and is also independent of the presence or absence of other attributes. The multiclass classifier works on the principal of divide and conquer technique and transforms the complex problem into simple two class problem which can be solved using one verses one path is traced in each if the decision trees constructed and this is repeated for all the training tuples. The final classification class for an instance is considered by taking the votes from all the decision trees.

The system of nested dichotomies recursively transforms the multiclass categorization problem into a relatively easy two class categorization problem which can easily be represented using binary trees. The resultant classification is obtained by aggregating the results obtained by the individual binary trees. In random committee, an assembly of individual multiple tree classifiers is made, each of is trained and tested with a common dataset but using a different randomness generating seed value. The DECORATE algorithm iteratively adds a new learning classifier to the current ensemble at each iteration. The data used for training is build by adding some artificial data to the original data. A new classifier is added only if on addition of new classifier, the total error is reduced. Rotation forest works by creating an ensemble or classifiers each trained using a different subset of attributes from the main dataset. J48-Graft creates a special decision tree which is grafted by a technique that adds fresh nodes on the places of either the old leaf nodes or in between the leaf nodes such that the accuracy in prediction is improved. Naïve Bayes classifier is based on the theory of conditional independence, which assumes that each attribute has its independent roll in the prediction of the final outcome and is also independent of the presence or absence of other attributes. The multiclass classifier works on the principal of divide and conquer technique and transforms the complex problem into simple two class problem which can be solved using one verses one path is traced in each if the decision trees constructed and this is repeated for all the training tuples. The final classification class for an instance is considered by taking the votes from all the decision trees.

The system of nested dichotomies recursively transforms the multiclass categorization problem into a relatively easy two class categorization problem which can easily be represented using binary trees. The resultant classification is obtained by aggregating the results obtained by the individual binary trees. In random committee, an assembly of individual multiple tree classifiers is made, each of is trained and tested with a common dataset but using a different randomness generating seed value. The DECORATE algorithm iteratively adds a new learning classifier to the current ensemble at each iteration. The data used for training is build by adding some artificial data to the original data. A new classifier is added only if on addition of new classifier, the total error is reduced. Rotation forest works by creating an ensemble or classifiers each trained using a different subset of attributes from the main dataset. J48-Graft creates a special decision tree which is grafted by a technique that adds fresh nodes on the places of either the old leaf nodes or in between the leaf nodes such that the accuracy in prediction is improved. Naïve Bayes classifier is based on the theory of conditional independence, which assumes that each attribute has its independent roll in the prediction of the final outcome and is also independent of the presence or absence of other attributes. The multiclass classifier works on the principal of divide and conquer technique and transforms the complex problem into simple two class problem which can be solved using one verses one path is traced in each if the decision trees constructed and this is repeated for all the training tuples. The final classification class for an instance is considered by taking the votes from all the decision trees.

The DECORATE algorithm iteratively adds a new learning classifier to the current ensemble at each iteration. The data used for training is build by adding some artificial data to the original data. A new classifier is added only if on addition of new classifier, the total error is reduced. Rotation forest works by creating an ensemble or classifiers each trained using a different subset of attributes from the main dataset. J48-Graft creates a special decision tree which is grafted by a technique that adds fresh nodes on the places of either the old leaf nodes or in between the leaf nodes such that the accuracy in prediction is improved. Naïve Bayes classifier is based on the theory of conditional independence, which assumes that each attribute has its independent roll in the prediction of the final outcome and is also independent of the presence or absence of other attributes. The multiclass classifier works on the principal of divide and conquer technique and transforms the complex problem into simple two class problem which can be solved using one verses one path is traced in each if the decision trees constructed and this is repeated for all the training tuples. The final classification class for an instance is considered by taking the votes from all the decision trees.

RESULTS

In general, all the algorithms performed better for 2-class problem as compare to the 7-class problem. For 2-class problem algorithms showed efficiency in the range of 93% to 95%, while as for 7-class problem, efficiency in the range of 69% to 78%. The reason for this is attributed to the increase in the complexity of the classification problem for classifying the cell instances into corresponding class among 7 classes as compared to classifying into just 2 classes. Among all the algorithms used, multiple backpropagation neural networks presented an efficiency of about 78.0%, whereas the efficiency of all other algorithms was in the range of 69% to 76%. Detailed results regarding the accuracy obtained by each intelligent methodology on both the 2-class and the 7-class problem are presented in Table 1.

For developing the semi-automated cervical cancer screening tool we used multiple back propagation algorithm, it being the best predictor among all the algorithms under consideration. The architecture used was standard backpropagation-based multilayer perceptron (MLP). In this neural network the neurons are organize in layer based graph with directed edges where each layer is fully connected to the next layer in feed-forward manner. The randomly initialized weights are updated and adjusted using the error values between network outputs and actual outputs. Backpropagation network training is a systematic way of training the network using extended gradient-descent technique based on delta-learning rule in order to determine and adjust the weights in the direction of minimizing error function. For processing the inputs at the artificial neurons, sigmoid function was used as the activation function on the network nodes both at the hidden layer and output layer. For realization of desired ANN, we have developed a self designed decision support tool in Matlab. The network was trained and tested for different number of hidden neurons and the weights corresponding to the best trained network were saved. The network exhibited the best results for 21 neurons in the hidden layer so the numbers
Table 1: Results of 10 folds cross validation for different algorithms

<table>
<thead>
<tr>
<th>S No.</th>
<th>Algorithms</th>
<th>Binary classification problem</th>
<th></th>
<th>Multiclass classification problem</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Correctly classified (%)</td>
<td>Sensitivity</td>
<td>Correctly classified (%)</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>1.</td>
<td>Bagging</td>
<td>95.2536%</td>
<td>0.953</td>
<td>76.4629%</td>
<td>0.765</td>
</tr>
<tr>
<td>2.</td>
<td>PART</td>
<td>94.4083%</td>
<td>0.944</td>
<td>73.5587%</td>
<td>0.736</td>
</tr>
<tr>
<td>3.</td>
<td>Decision Table</td>
<td>94.5384%</td>
<td>0.945</td>
<td>74.0355%</td>
<td>0.74</td>
</tr>
<tr>
<td>4.</td>
<td>Decorate</td>
<td>95.0585%</td>
<td>0.951</td>
<td>76.4629%</td>
<td>0.769</td>
</tr>
<tr>
<td>5.</td>
<td>END</td>
<td>93.9965%</td>
<td>0.940</td>
<td>75.8778%</td>
<td>0.759</td>
</tr>
<tr>
<td>6.</td>
<td>Filtered Classifier</td>
<td>94.7334%</td>
<td>0.947</td>
<td>74.5340%</td>
<td>0.745</td>
</tr>
<tr>
<td>7.</td>
<td>J48 Graft</td>
<td>94.4083%</td>
<td>0.944</td>
<td>73.6238%</td>
<td>0.736</td>
</tr>
<tr>
<td>8.</td>
<td>Multi-Class Classifier</td>
<td>94.9502%</td>
<td>0.950</td>
<td>76.2029%</td>
<td>0.762</td>
</tr>
<tr>
<td>9.</td>
<td>Rotation Forest</td>
<td>95.4270%</td>
<td>0.954</td>
<td>75.1409%</td>
<td>0.751</td>
</tr>
<tr>
<td>10.</td>
<td>Naïve Bayes</td>
<td>93.6064%</td>
<td>0.936</td>
<td>69.7226%</td>
<td>0.697</td>
</tr>
<tr>
<td>11.</td>
<td>Radial Basis Function</td>
<td>94.1482%</td>
<td>0.941</td>
<td>74.7941%</td>
<td>0.748</td>
</tr>
<tr>
<td>12.</td>
<td>Random Committee</td>
<td>95.3836%</td>
<td>0.954</td>
<td>76.2029%</td>
<td>0.762</td>
</tr>
<tr>
<td>13.</td>
<td>Random Forest</td>
<td>95.2336%</td>
<td>0.953</td>
<td>75.3576%</td>
<td>0.754</td>
</tr>
<tr>
<td>14.</td>
<td>Random Sub Space</td>
<td>95.2336%</td>
<td>0.953</td>
<td>76.9181%</td>
<td>0.766</td>
</tr>
<tr>
<td>15.</td>
<td>Multiple Backpropagation ANN</td>
<td>95.6220%</td>
<td>0.956</td>
<td>78.0668%</td>
<td>0.781</td>
</tr>
</tbody>
</table>

DISCUSSION

Cervical cancer being one of the most common women cancers in the world is a major cause of female mortality. In developed countries there is a proper system of annual regularly screening all the females that are susceptible to cervical cancer, unfortunately in low income countries and countries having a large boom of population like India, such mass level screening programs are very difficult to implement owing to various economic and social constraints. India is contributing to more than 25% of the world’s burden of cervical cancer which accounts for death of eight patients every hour. India being a developing nation with world’s second largest share of population, implementation of regular screening for cervical cancer like that of pulse polio is next to impossible. In such a scenario computer based intelligent screening techniques can prove to be very useful in reduction of burden of cervical cancer on a nation in particular and on the whole world in general. Such tools also can help to identify the cases that are more prone to this disease and as such can aid in timely identification of cases having premalignant transformations that may later on develop into cancers.

Authors also compared the prognostic ability of the cervical cancer screener proposed in this work with the machine learning based diagnostic systems developed by other researchers. Olivier Regnier-Coudert et al. in 2012 have done a comparison of 3 different machine learning models for diagnosis of Prostate cancer. The models used by them were artificial neural networks, bayesian networks and logistic regression which expressed the true positive rates of 0.656, 0.679 and 0.675 respectively. Compared with the prognostic results of these automated system build for prostate cancer, the novel cervical cancer screener shows much better sensitivity of 0.781, which is very close to the best possible classification potential by any classifier. Sokouti et al., 2014, have proposed also proposed a model (Levenberg–Marquardt feedforward MLP neural network-LMFFNN) for classification of cervical cell images. The system proposed by them is trained with data obtained from 100 patients and classifies the cervical cells in normal and cancerous. The system proposed in this work not only classifies the cells into normal and abnormal, but also categorizes them according to Bethesda system of classification which is universally accepted system of reporting cervical smears. Chiu et al., also studied the application of artificial neural networks in predicting the skeletal metastasis in patients suffering from prostate cancer. They analyzed whole body bone scintigraphies of patients with prostate cancer who underwent the technethium-99m methylene diphosphate (tc-99m MDP). For evaluating the performance of ANN, Receiver Operating Characteristics (ROC) analysis was used. The area under the ROC curve (0.88±0.07) revealed excellent discriminatory power (p<0.001) with the best simultaneous sensitivity (87.5%) and specificity (83.3%). Authors concluded that that an ANN, which is based on limited clinical parameters, to be a promising method in forecasting of the skeletal metastasis. The system proposed in this work is comparable with their automatic diagnostic system. Artificial intelligence based diagnostic systems have several advantages over conventional methods, as these programs have a potential to greatly reduce the number of diagnostic possibilities by using the patho-physiological reasoning. As such they can create a logical model of a specific patient’s medical condition. Such models can also model the complexities generated by overlapping symptoms of multiple diseases.
Although only a few such programs are currently available, the evidence suggests that the continued development of artificial intelligence techniques will eventually give the computer a major role as an expert consultant to the physician.

**CONCLUSION**

The results demonstrate that the artificial intelligence based techniques can effectively be used for developing computer assisted intelligent tools for mass level screening of cervical cancer.

**REFERENCES**

2. Vinod B Shidham et al. p16 INK4a immunocytochemistry on cell blocks as an adjunct to cervical cytology: Potential


Source of Support: Nil; Conflict of Interest: None
Submitted: 25-02-2017; Accepted: 02-05-2017; Published: 06-06-2017