ABSTRACT

Cervical cancer is one of the most dangerous cancerous which ranks four among the cancers affected worldwide. The major cause of the disease is the ignorance about this type of cancer. Human Papillomavirus (HPV) is the major cause of cancer. The risk factors for cervical cancer are sex at an early age, weak immune system, usage of drugs, consumption of birth control pills etc. The test carried out for the detection of HPV is a pap smear test. At present, the pap images are analyzed manually for the detection of cervical cancer. The manual method of detection is erroneous as the abnormal cells are very similar to the normal cells in appearance and as a result, it is subject to false positive and false negative cases.

This research work involves automated detection of cervical cancer from pap images. For the automated detection of cervical cancer machine learning algorithms, neural networks and transfer learning algorithms are used. Geometrical Features (19 Nos.) and texture features (4 Nos.) are extracted to classify the cell into normal and abnormal cells. If the twenty three features are considered for the whole process it will consume time. In order to overcome this limitation, the four most significant features are identified by applying Principal Component Analysis (PCA). These significant features are given as the input to the machine learning algorithms and neural network. Machine learning algorithms like KNN, Fine Gaussian SVM, Ensemble Bagged trees and Linear Discriminant are used for the classification of the input images into normal and abnormal. The main limitation of machine learning algorithm is that the features are extracted manually. The accuracy of the system will depend on the features extracted.

The detection of cervical cancer is also implemented by neural networks. The neural network depicts the working principle of the human brain. The neural network has input layers, hidden layer and output layer. The performance of the neural network with 10 neurons and 20 neurons are analyzed by computing the accuracy, precision, sensitivity, specificity and F1-score. Neural network learns more from the class which have more members and learns less from the class which has fewer members.

In order to overcome the limitation of machine learning and neural networks, deep learning algorithm may be suggested. However, a deep learning algorithm requires a large number of labeled images. The labeled images in the medical field are very rare and this adds to the limitation of deep learning algorithm. Transfer learning algorithm overcomes the limitation of deep learning algorithm. The knowledge gained during the training phase of classification on a large database is applied on the small database which is our intended work. The performance of different transfer learning algorithms like Resnet50, Googlenet, Squeezenet, Alexnet and Mobilenet are analyzed for different hyper parameters. The hyperparameters like epoch and learning rates are varied and the performance of the system is analyzed. The number of images taken for training is also varied from 10%-90% for each epoch and learning rate. The highest accuracy of 98.23% is obtained for Googlenet for epoch 6, learning rate 0.003 with 80% of the images considered for the training.