

Ovarian Cancer Diagnosis Using Artificial Neural Network.

Sukanya Pandey

Anna University Chennai, Department Of Biotechnology

Abstract:

Ovarian cancer is a quiet and dangerous enemy that affects women all over the world. It needs to early diagnosis. Radiology, cardiology, and cancer are expanding ANN research. Medical research uses ANNs. Thus, a Computer Aided Diagnosis (CAD) system using ANNs to categorise ovarian cancer based on biopsy pictures is being developed. Ovarian cancer is the fifth most common disease and seventh major cause of death among women. Many studies categorise ovarian cancer using ANN. Classification accuracy affects physicians. Better classification helps doctors choose treatments. Early and accurate diagnosis may reduce mortality. Comparing the proposed model to the other four categorization techniques. The recommended methodology classifies ovarian cancer 98.7 percentage more accurately than earlier algorithms.

Introduction:

Ovarian cancer is the kind of cancer that kills more women than any other form of cancer combined. It is the second most frequent type of gynaecologic cancer. Pathologists are prone to making mistakes while trying to differentiate between the four distinct kinds of ovarian cancer when they are studying and evaluating several cytology pictures. Image processing is a tool that may be used in the field of pathology to improve diagnostic accuracy while simultaneously reducing the amount of work required by pathologists. Using cytological pictures, Deep Convolutional Neural Networks (DCNN) based on Alex Net were used to automatically categorise the various kinds of ovarian cancer. Five convolutional layers, three max-pooling layers, and two full-reconnect layers make up the Deep Convolutional Neural Network (DCNN). After then, the model was trained using two distinct sets of input data: the first set consisted of the original picture data, while the second set consisted of improved image data that included image enhancement and image rotation. The test results, which were obtained by applying 10-fold cross-validation, demonstrate that the accuracy of the classification models was increased from 72.76 to 78.20% by utilising upgraded photos as training data. This is a significant increase from the previous level of 72.76. The developed scheme was useful for classifying ovarian cancers from cytological images. Techniques to develop a methodology for diagnosis of lung nodules. CADx system by combining two major methodologies, which are the fuzzy base systems and the evolutionary genetic algorithms. The accuracy of the system can be 97%. Used voxel-wise feature maps and SVM to investigate the added diagnostic value of 2 arterial spin labelling and diffusion tensor imaging to structural MRI for computeraided classification of Alzheimer's disease, front to temporal dementia, and controls. Chena et al. established an expert diagnosis system for cerebrovascular diseases and assessed accuracy of the

diagnosis system. The information presented here makes it quite clear that ML is used extensively in CADx.

Among them, we discovered that a subfield of machine learning known as deep learning has lately gained significant traction in the medical image processing domains. As an alternative to task-specific algorithms, deep learning is a member of a larger family of machine learning techniques that are centred on learning data representations. It all began in late 2012 with an event in which a deep learning strategy that was based on a convolutional neural network (CNN) earned an overwhelming win in the most well-known international computer vision competition. Deep learning, which includes deep belief nets (DBNs) and deep CNNs, utilises image pixel values directly as input data as opposed to image features calculated from segmented objects. As a result, manual feature calculation or object segmentation is no longer required, which results in a process that is both simpler and more time-effective. Since then, researchers from almost every subject, including medical imaging, have been actively contributing in the rapidly expanding field of deep learning. This includes the field of artificial neural networks. Deep Convolutional Neural Network (DCNN) activation characteristics were suggested by Xu et al. for the purpose of performing classification, segmentation, and visualisation in large-scale tissue histopathology pictures. Using DCNN, Teramoto et al. created an automatic classification system for lung tumours that were shown in microscopic pictures. The deep convolutional neural networks (DCNNs) were used by Gao et al. in order to offer an autonomous framework for the categorization of human epithelial-2 cell images. The purpose of practically every application that deals with image processing, on the other hand, is to derive information from the picture data. Filtering, converting, colouring, interactive analysis, or any of a variety of other approaches might be necessary in order to get the information.

2. METHODOLOGY

The gathering of images from the various databases is the initial step in the process. There is only one dataset that is kept in the needed work. The machine learning technology known as support vector machine is employed for training and testing purposes, and the photos from the database are used for both purposes. The photographs of ovarian cancer may be found in the files that are in the .jpg format. During the pre-processing step, the image is uploaded and the RGB gray-scale form is converted to the minimal image pixel size of the ovarian cancer picture. This is done so that the interference level in the original image, which is an MRI of ovarian cancer pictures, can be determined. The Otsu Method should then be used to locate the affected areas of the cancer picture. Following that, extract the features from the medical picture using the DWT method. It identifies two form coefficients, known as the lower and upper bounds respectively. K-SVM is used to train the cancer pictures in each stage's dataset, test the cancer detection, and increase the quality of the cancer image (MRI pictures). The classification and clustering techniques are applied using the datasets. Calculate the performance measures, such as the False Acceptance rate (FAR), the False Rejection rate (FRR), and the Accuracy rate, and compare them to methods that have been used in the past, specifically accuracy.

2.1 PROPOSED SYSTEM

The categorization of cancer is a crucial step in helping physicians decide on the best course of therapy, which may lower the death rate by saving lives. This research examined the accuracy of the categorization of ovarian cancer using a 15-neuron ANN model. There are five phases in the categorization model for cancer. The gathering of data is the initial stage. A stage in the experiment is data gathering. For the

categorization of cancer, it is crucial to gather information from malignant patents. The pre-processing of data is the next phase. The data samples are gathered as unprocessed data. The third stage is data partition, which comes after this step has been pre-processed for additional appropriateness for this suggested model. Training and testing datasets were segregated at this stage. The fourth stage in creating an accurate model to identify cancer is choosing the neuron types for an ANN. The Taguchi technique is used to choose the neuron in an ANN. This research performs well when the suggested strategy is used with the ANN model. Based on the solid performance, we discovered that 15 neurons make up the optimal classification model in the hidden layer. The final classification accuracy estimate is the last stage. This method used a dataset of ovarian cancer to obtain classification performance.

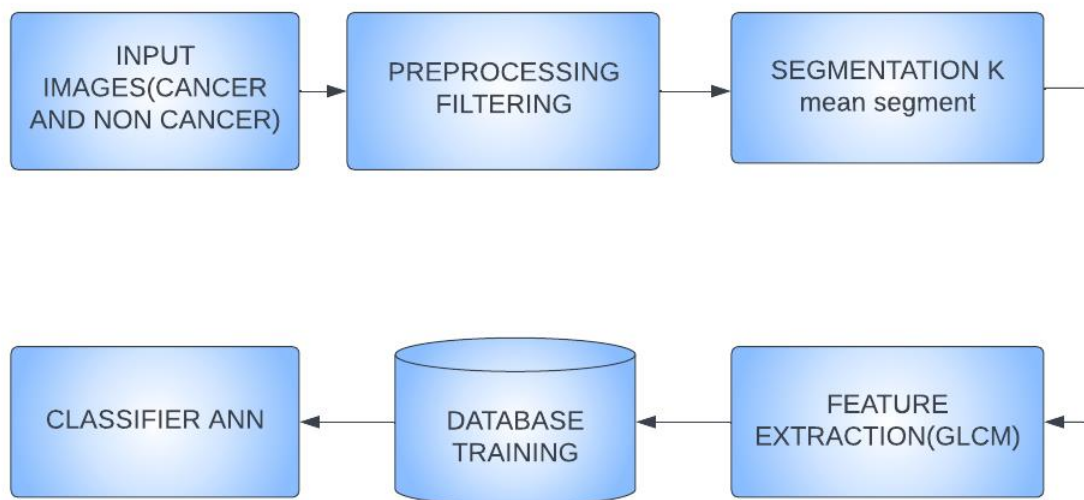


Fig 1: Proposed Model

Due to noise, illumination fluctuations, climatic circumstances, picture resolution, undesired backdrop, etc., image acquisition images may not be acceptable for identification and classification. The method assigns each data point to a cluster when the number of clusters is pre-determined. FCM assesses the probability, not the absolute membership, of a data point to a cluster. The precision of the clustering needed in practise determines the tolerance measures. FCM is quick because the amount of iterations needed for a clustering exercise matches the necessary precision because absolute membership is not determined. Colour space lets us define, visualise, and generate colours. Colour feature extraction techniques vary. Extracting colour features: a. Histogram intersection: Histogram Intersection (HI) evaluates global colour features. In Histogram intersection technique, the number of bins affects efficiency. X and Y are colour histograms with k bins each. The image's complicated representation by several bins increases computing complexity.

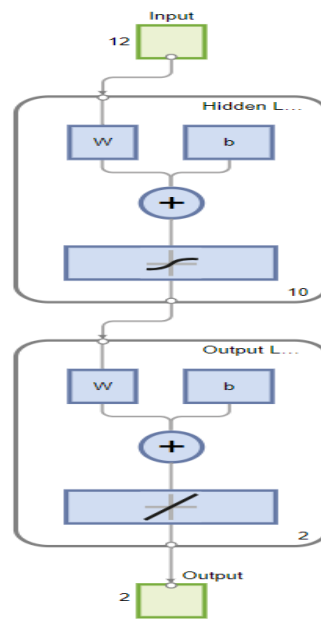


Fig 2 : Network Model

Data categorization and clustering employ neural networks (NN). It aims to create a learning machine that mimics brain activity. Examples teach NN. NN can classify data and find new patterns if given enough instances. Basic NN has input, output, and hidden layers. Input and hidden layer nodes are linked. Hidden and output layer nodes are linked. Those links reflect node weights. ANN consider classification one of the most active research and application topics. The main drawback of ANN is finding the best training, learning, and transfer function for classifying data sets with increasing features and classified sets. The influence of different functions on ANN classifiers and their accuracy for diverse datasets are explored. The network produces unseen data after training. Multilayer FFNNs transfer signals just from input to output. Paired data teaches the network input-output mapping. After neuron weighting, the network classifies new data. Classification uses net propagation of input unit signals to activate output units. Each input unit's activation value represents a net-external trait. Input units tell hidden units their activation values. Each hidden unit calculates and delivers its activation value to output units. Simple activation functions calculate each receiving unit's activation. The function sums all sending units' contributions—the weight of the connection between sending and receiving units multiplied by their activation value. The activation total is often set to 0 or 1 until a threshold level is reached.

3. RESULT & DISCUSSION:

There are 25 images of microscopic ovary biopsy that are categorized into cancer and non-cancer categories and different stages of cancer, which can be classified in further steps using network designing.

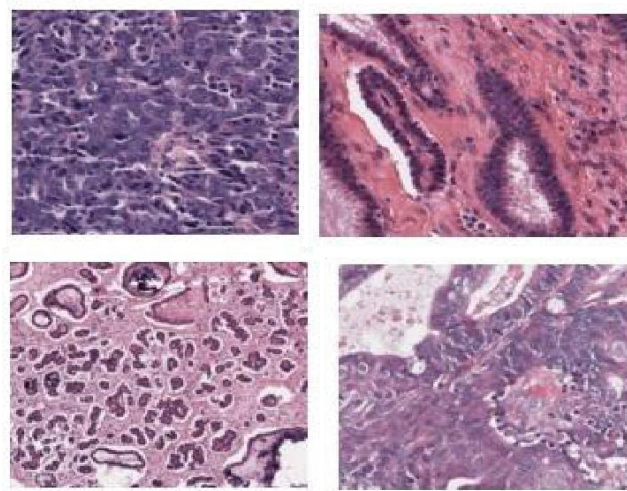


Fig 3 : Input Images

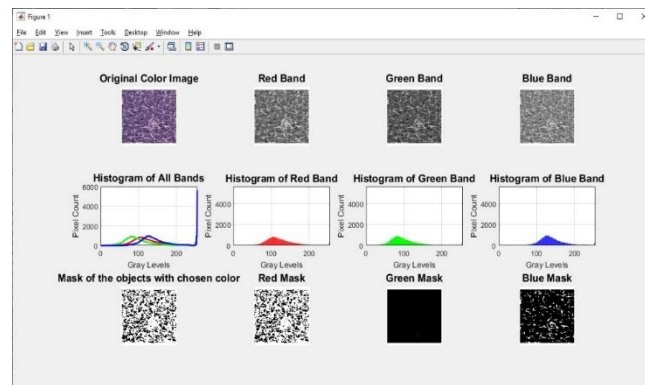


Fig 4 : Filter Images

The colour picture has been converted to a grayscale version since the colour information does not assist us in identifying significant edges or other characteristics. A linear filter is what's known as a Gaussian filter. It is used to minimise noise, and on its own, it will make edges less distinct and lower contrast. The size of the items rises as a result of dilation, which also helps to fill in holes and spaces and link sections that were previously separated by spaces that were less than the size of the structural element. As using grayscale photos, the brightness of the image will also rise as it is enlarged Erosion is a tool that is used for the purpose of deleting unimportant size information from a binary picture. It reduces the size of the picture and causes the image to become thinner; however, in this case, the image is being subtracted so that the temporal derivative of intensity may be estimated at each point in a succession of photographs. It is used to make the picture seem more ominous. It is necessary to do both the inverse subtraction and the subtraction of the image in order to get the original pixel of the picture that is included inside the input image.

The sole damaged area is clustered together by segmentation, which also clusters the affected region of the tumour cell. So that we may concentrate solely on the damaged region, there is no need to search the biopsy picture for all of the features of the unimportant surrounding area. It helps to cluster the sole afflicted area, which is the cluster impacted zone of the tumour cell, and segmentation does this.

After pre-processing the images and locating the large particle in the image, a mask was created with filled holes. This allowed for erosion to be applied in order to get rid of unwanted regions, and a boundary was

given in the defected area, which allowed for the affected area to be determined. The affected area was found out by determining the big particle in the biopsy image. This was done after pre-processing the images.

After all of that image processing and training the data, the input was given in the form of an image, which was decoded and converted into text formed after that neural network was applied to generate Simulink model, to initialise weights and biases, calculation of performance, evaluation of network output, and then finally train the networks, and the image was found to be cancerous.

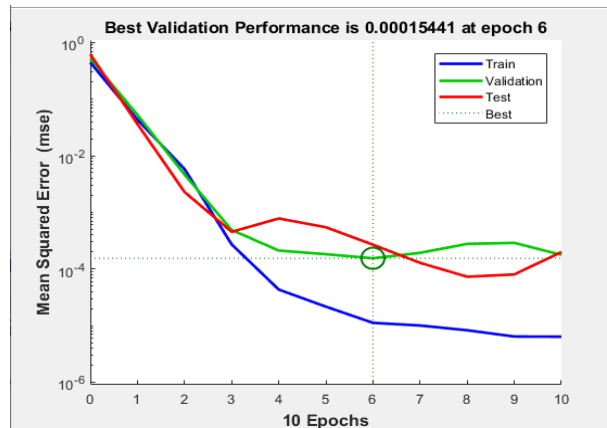


Fig 5: Train, Test & Validation

It was a graph showing the validation performance in terms of number of epochs and mean squared error. The training was terminated when the validation error rose for six iterations, which took place at iterations (or epoch) 10, as shown by the green circle in the graph. It was shown by the use of the graph that the final mean-square error is 1.01, which is a negligible value. The test set error and the validations set error had characteristics that were quite close to one another. In the figure, it was represented by both green and red lines. It was not surprising to find that the mean square error (MSE) decreased with an increase in the number of epochs used to train, validate, and test the model, despite the fact that there was no significant difference in the slope. The epoch 6 performance had the best validation results in terms of MSE, which was 0.00015441.

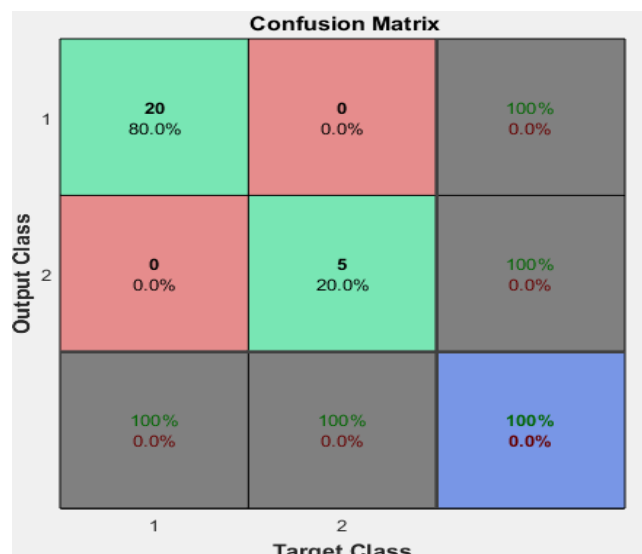


Fig 6: Confusion Matrix

The percentage of correct classifications is also shown. Twenty biopsies have been appropriately identified as having benign results. This accounts for 80.0% of all 26 biopsies that were performed. In a similar vein, five instances have been appropriately identified as having a malignant nature. This accounts for twenty percent of all biopsies performed.

The percentage of erroneously identified biopsies, which amounts to 0.0% of the total of 26 biopsies included in the data, is 0 for malignant and 0 for benign. In a similar vein, twenty of the benign biopsies are mistakenly labelled as cancer, and this accounts for one hundred percent of the data.

Twenty forecasts have been made, and each one has been spot on one hundred percent of the time. With regard to the five forecasts of malignant behaviour, each one is accurate one hundred percent of the time. It is properly predicted that none of the twenty benign instances are cancerous, resulting in a perfect prediction rate of 100%. There are 20 instances of malignancy, all of which are appropriately categorised as malignant, while none of them are correctly classified as benign.

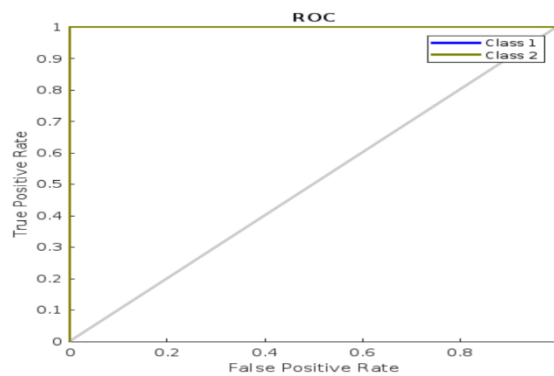


Fig 7: ROC Curve

That is, a separate cutoff value is represented by each individual point on the ROC curve. The points are linked together in order to create the curve. Values for the cutoff that provide low rates of false-positive results also tend to produce low rates of true-positive results. Both the true-positive rate and the false-positive rate tend to rise in tandem with each other. The more accurate a diagnostic test is, the closer it comes to having a true positive rate of 1, which is the same as a rate of 100%. The ROC curve of a diagnostic test that is almost perfect would be practically vertical from 0 to 1 and then horizontal from 1 to 0. This would indicate that the test is nearly flawless. Due to the fact that it represents the ROC curve of a diagnostic test that randomly identifies the condition, the diagonal line functions as a reference line.

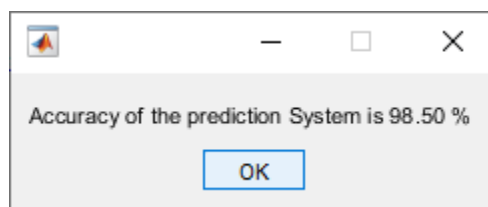


Fig 8: Accuracy Prediction

The accuracy of the prediction method for ovarian cancer was attained at a level of 98.5%, making it significantly more accurate than the study that had been done before on a topic comparable. The accuracy of the prediction is the most essential aspect in determining whether or not the label is right, and the accuracy varies from model to model as well as from algorithm to algorithm.

References

1. W.Gonzalez , “Digital Image Processing”, 2nd ed. Prentice Hall, Year of Publication 2008.
2. Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, Second Edition.
3. Rafael C. Gonzalez, Richard E. Woods, Steve L. Eddins, Digital Image Processing Using MATLAB, 2003.
4. Rania Hussien Al-Ashwal, Eko Supriyanto, et.al., “Digital Processing for Computed Tomography Images: Brain Tumor Extraction and Histogram Analysis”, Mathematics and Computers in Contemporary Science, 2013
5. J. Selvakumar, A. Lakshmi and T. Arivoli, “Brain Tumor Segmentation and its area calculation in Brain MR images using K-means clustering and fuzzy C-mean algorithm”, International Conference on Advances in Engineering, Science and Management, 2012
6. R. Rajeswari, P. Anadhakumar, “Image segmentation and identification of brain tumor using FFT techniques of MRI images”, ACEEE International Journal on Communication, Vol. 02, No. 02, July 2011
7. Mustaqeem, Anam, Ali Javed, and Tehseen Fatima, “An efficient brain tumor detection algorithm using watershed and thresholding based segmentation”, International Journal 4, 2012
8. P.Dhanalakshmi, T.Kanimozhi, “Automated Segmentation of Brain Tumor using K-Means Clustering and its area calculation”, IJAEEE, 2013.
9. Q.Hu, G. Quian, A. Aziz, W.L.Nowinski, “Segmentation of Brain from Computed Tomography head images,” Engineering in Medicine and Biology 27th Annual Conference, 2005.
10. Natrajan P. , Krishnan N. , Natasha Sandeep kenkre and et.al ,”Tumor Detection using Threshold operation in MRI Brain Images,” IEEE International Conference on Computational Intelligence and Computing Research, 2012.
11. P. Natrajan, Debsmita Ghosh, kenkre Natasha Sandeep, Sabiha Jilani, ”Detection of Tumor in Mammogram Images using extended Local Minima Threshold,” International Journal of Engineering and Technology, Vol. 5, No . 3, jun-jul 2013.
12. X. Zang, J. Yang, D. Weng, Y. Liu and Y. Wang, “A novel anatomical Structure segmentation method of CT head images,” International Conferences on complex medical Engineering, 2010.
13. A. Padma and R. Sukanesh, “Automatic Classification and segmentation of brain tumor in CT images using optimal dominant gray level run length texture features,” International journal of Advanced Computer Science and Applications, 2011.
14. R.C. Patil and Dr. A.S. Bhalachandra, “Brain tumour extraction from MRI images using MATLAB,” International Journal of Electronics, Communication & Soft Computing Science and Engineering, vol. 2
15. Ahmed Osmanovic, Adnan Hodzic, Layla Abdel-Ilah, Jasmin Kevric. Ovary Cancer Detection using Decision Tree Classifiers based on Historical Data of Ovary Cancer Patients; 2017.
16. Kusworo Adi, Catur Edi Widodo , Aris Puji Widodo , Rahmat Gernowo , Adi Pamungkas and Rizky Ayomi Syifa. Detection Lung Cancer Using Gray Level Co-Occurrence Matrix (GLCM) and Back Propagation Neural Network Classification; 2018.
17. Kalpana U. Rathod and Prof. Y. D. Kapse. MRI Brain Tumor Detection Using Artificial Neural Network; 2016.
18. American Cancer Society. Cancer facts and figures. US: American Cancer Society; 2007.