ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

Cancer Classification using Principal Component Analysis and Deep Neural Networks

Heena Sindhani, Dr. Kailash Shaw

M.E Student, Dept of Comp Engg, D Y Patil College of Engg, Akurdi, India

Professor, Dept of Comp Engg, D Y Patil College of Engg, Akurdi, India

ABSTRACT*:* Cancer is an uncontrolled division of abnormal cells within the body. Cancer develops when the body's normal control mechanism stops working. Identification of cancer by using some of the parameter plays very important role to reveal useful awareness of cancer disease. Classification of cancer determines appropriate treatment and helps determine the prognosis.In the system, new machine learning model for cancer detection were used which is known as Principal Component Analysis and Deep Neural Networks for identification and classification of cancer sub- types. Principal component analysis (PCA) is the dimensionality reduction technique of a data set consisting of many variables that may be correlated with each other, either heavily or lightly, while retaining the variation available in the dataset, up to the highest extent. The system collects cancer dataset from machine learning repository. In this paper, a systematic study to investigate PCAs feature extraction is presented. DNN with automatic feature extraction is firstly evaluated on a Breast Cancer dataset. Dataset holds features of cancer like cell size, cell shape etc. and these features are compared with trained input patients features and the matched cancer subtypes using various classifiers is predicted as a final output.

KEYWORDS*:* Cancer Classification, Principal Component Analysis, Deep Neural Networks,

I. INTRODUCTION

Cancer is a group of uncontrolled cell division that leads to abnormal tissue growth. Numerous a times, people often use the terms cancer and tumor synonymous. Yet all the know tumors are not cancerous. For disease pathogenesis, identification of tumor subtypes i.e. malignant and benign tumors is very useful. The first step is to cluster cancer patients to have a better view of virus causing the disease. Clustering is done according to their modalities such as gene expression and clinical information. Thus, provide a way to have better anticancer treatment for the cancer patients. In the recent development, it allows collecting the genomic data from the different platform for the tumor sample of the same set. Genomic data includes gene expression, DNA methylation and miRNA expression which is very useful for pathogenic mechanisms. The Breast Cancer Wisconsin (Diagnostic) Dataset that is used is freely available at UCI repository. The breast cancer dataset contains genetic factor expression and medical information, for example, survival time. Netherlands Cancer Institute provides this dataset. It consists of 569 instances and 32 attribute of patients report. Many feature selection algorithms have been employed to reduce gene expression data dimensions before further classification or clustering, such as principal component analysis (PCA) [1], [2], [3], [4], partial least squares (PLS) [5], independent component analysis (ICA) [6], [7], nonnegative matrix factorization (NMF) [8], [9], [10], and their variants [11], [12], [13], [14]. As a well-established dimension reduction technique, PCA projects data in an orthogonal subspace generated by the eigenvectors of the data covariance matrix. The maximum variance direction-based subspace spanning assures the minimum information loss in the feature selection. However, as a global feature selection algorithm, PCA can only capture these features contributing to the overall characteristics of data. It misses those features contributing to the local characteristics of data because each PC only contains some levels of global characteristics of data [15], [16]. This global feature selection mechanism not only leads to difficulty in interpreting each principal component (PC) intuitively but also prevents subtle data local latent structure discovery in the following

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

classification DNN (Deep Neural Networks-a neural network with multiple hidden layers) consists of layers of nodes (representing neurons) connected to nodes in adjacent layers using links that have weights (representing synapses) associated with them. Nodes in DNN can be active or inactive. Information is propagated in DNN by transmitting activation from one node to its neighbors. If a node is active, it can activate its neighbors only if they are connected by strong weights. Like its biological counterpart, DNN captures knowledge through distributed weights and activations of the nodes across the network. This is in stark contrast to the Symbolic paradigm. One of the important elements of a neural network is its ability to learn which is done by adjusting weights.

A Deep Neural Network (DBN) is a pretending neural network which consist of binary variables. The network consists of one visible layer, vl ⋅ {0,1}n, where n is the number of visible variables and sequence of hidden layer, hl(1) ⋅ {0,1}m1, hl(2) ⋅ {0,1}m2,..., hl(t) ⋅ {0,1}mt, where m1,m2,...,mt are the number of hidden layer hl(1), hl(2),..., hl(t) respectively [1]. The hidden layer is constructed using conditional distributions of the Replicated Softmax Model as follows:

$$p(hl_j = 1|VL) = g(hbias_j + \sum_{i=1}^{n} W_{ij} vl_i) \quad \text{.....(1)}$$

where hlj is the hidden variable, V L is the visible variable vector, hbiasj is the bias of hidden variable, n is the number of visible variables, Wij is the weight matrix and vli is the visible variables, and g(·) is transform function. The reconstruction of the visible layer is given by using conditional distributions of the Replicated Softmax Model as follows:

$$p(vl_{ki} = 1|hl) = \frac{exp(vbias_i + \sum_{j=1}^{m} W_{ij} hl_j)}{\sum_{q=1}^{i} exp(vbias_q + \sum_{j=1}^{m} W_{jq} hl_j)} \quad \text{.....(2)}$$

where vlki is visible variable, hl is the hidden variable vector, vbiasi is the bias of visible variable, m is the number of hidden variable and Wij is the weight matrix. The DNN is learned using the equation (1) and (2). The computation of the weight matrix has high time complexity. The flow of this paper is as follows: Section II deals with Literature Review followed by Proposed System Overview in section III. Experiment Evaluation is been done in section IV and Conclusion is in section V..

II. LITERATURE SURVEY

An easy way to comply with IJRASET paper formatting requirements is to use this document as a template and simply type your text into it.

A. Principal Component Analysis:

Principal Component Analysis (PCA) is a linear dimensionality reduction method that generates linear combinations of original features that are capable of projecting original data on a reduced dimensional space [12]. It is a most popular data pre processing technique that reduces the feature space by capturing linear dependencies amongst various features. PCA calculates principal components (PCs) that are linear combinations of original attributes. The PCs are orthogonal to each other and capture maximum amount of variance in the data.Several variants of PCA exist that use methods other than covariance matrix such as correlation matrix for calculating eigenvalues. Advantage of using PCA is low computational cost and low noise sensitivity. However, PCA depends upon the scalability of data and relies on assumption that features covering maximum variance are the most important ones.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

B. Deep Neural Networks:

The conventional ML algorithms like Classification (LR, DT, Nave, RF, SVM and KNN) algorithms provide better accuracy in the diagnosis of cervical cancer. Machine learning is a quickly developed subarea of artificial intelligence. A branch of ML techniques based on the deep learning approach, such as Deep Neural Network (DNN) model, is producing good results across a variety of artificial intelligence tasks. The approach of deep learning is inspired by the ability of the human brain for complex tasks like image recognition and automatic translation. DNN is trained via layer-wise backpropagation to obtain tractable optimization Abid Sarwar, et al. [17], used Ensemble of Nested dichotomies (END), Nave Bayes, Rotation Forest algorithm, PART, Random forest, Bagging, Multiclass classifier, Decorate, Random Committee ,Decision Table, Random subset space, Filtered Classifier, Multiple back propagation artificial neural network (MBP ANN), J48 graft and Radial basis function network algorithms in their work using Weka tool. Finally they concluded that, the Ensemble of Nested Dichotomies (END) algorithm gives accurate results in the classification of cervical cancer. Support Vector Machine [18] is a binary classifier used to classify the data objects which are located in space. It is a supervised machine learning algorithm helps in both classification and regression. When the data objects are unable to distinguish in single dimension, with the help of SVM, we can project on multidimensional space for better classification.

III. PROPOSED METHODOLOGY

The main objective of the proposed system is to degrades the time complexity of computation by using the dot product calculation. And enhance the accuracy of the model. Input is given as a dataset which consists of two modalities such as gene expression and medical information, for example, survival time.

A. *Preprocessing*:

Most probably, real-world information contains numerous errors, incompleteness and inconsistency for that data preprocessing is needed. There are two steps in preprocessing:

1) The data set contains some arrangement of invalid esteems to dispose of those Imputer is utilized. It is an adaptable class that enables you to indicate the incentive to supplant (NaN) with mean, middle, or mode.

2) After removing null values, the next step is normalization of the dataset to get correct results, except class labels the remaining values should be converted (0-1).

B. *Principal Component Analysis:*

PCA can be most effectively used to reduce dimensionality for better visualization of dataset in lower dimensional space since it provides principal components which are linear combinations of uncorrelated attributes that best describes the variance among data. However, PCA does not provide a subset of real attributes and do not consider the correlation of attributes with class, hence, is not a suitable choice for feature selection. Feature evaluation and ranking algorithms such as filters and wrappers are capable of selecting suitable subset of features but they prove to be computationally inefficient and impractical for datasets having very high dimensionality. The output of PCA is a reduced set of uncorrelated features on which feature evaluation and ranking is applied. This will result in improvement of computation cost as compared to applying feature evaluation and ranking on all the features. The steps of PCA are as follows:

1) Standardize the data.

2) Calculate Covariance: Var[X1] = Cov[X1,X1] and Var[X2] = Cov[X2,X2].

3) Obtain Eigenvectors and Eigenvalues. $\det(\delta I - A) = 0$

$(\delta I - A)v = 0$

4) Form the Feature vector: Feature vector=(eig1, eig2)

5) Forming Principal Components: Newdata= Feature vector * Scaled data C. Deep Neural Network Deep Neural Networks as shown in figure2 are utilized to discover concealed learning from the complex datasets. Keras is a capable simple to-utilize Python library for creating and assessing deep learning models. It wraps the proficient numerical calculation libraries Theano and Tensor Flow [11] and enables you to characterize and prepare neural system models in a couple of short lines of code. In deep learning, epoch is a complete pass through given dataset. Errors can be minimized by adjusting the number of epochs using back propagation.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

C. *Deep Neural Networks:*

Deep Neural Networks as shown in figure2 are utilized to discover concealed learning from the complex datasets. Keras is a capable simple to-utilize Python library for creating and assessing deep learning models. It wraps the proficient numerical calculation libraries Theano and Tensor Flow [11] and enables you to characterize and prepare neural system models in a couple of short lines of code. In deep learning, epoch is a complete pass through given dataset. Errors can be minimized by adjusting the number of epochs using back propagation.

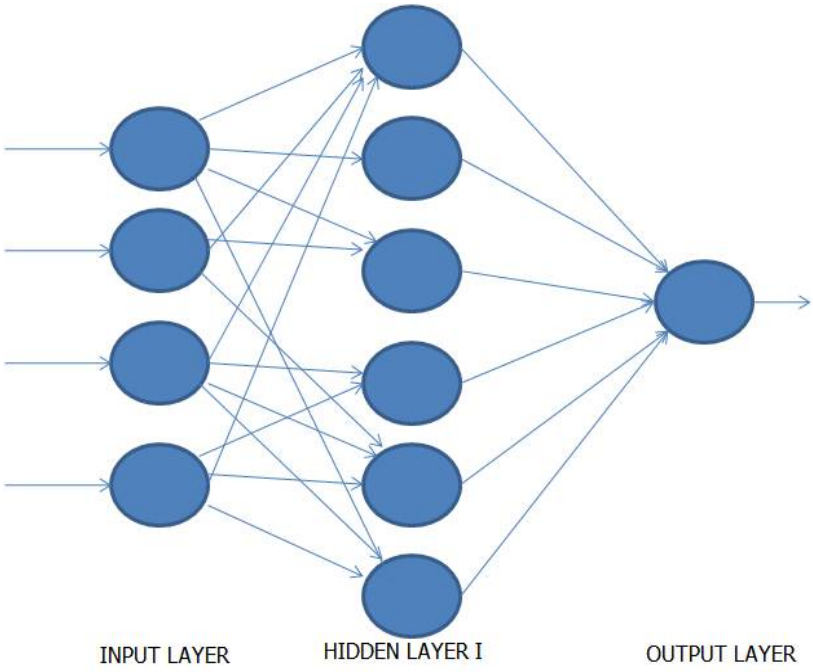

Fig 1: Deep Neural Network Model

DNNs can be trained with the standard back propagation algorithm [16], as follows, in the first N layers:

$$V = f(z^n) = f(W^n V^{n-1} + b^n), for 0 < n < N \quad \text{.....(3)}$$

And the sigmoid function is used as the activation function:

$$tanh(z) = e^z - e^{-z} / e^z + e^{-z} \quad \text{.....(4)}$$

where N is the number of layers, W is the weight, and Z is the output.

D. *System Architecture:*

To classify the cancer data, Principal Component Analysis along with Deep Neural Networks is used. The proposed system architecture is shown in the Figure 3. In Figure 3 two modalities i.e. gene expression and clinical information are given as the input. Deep Neural Network (DNN) is used for each modality. DNN is composed of multiple layers in which the first layer is the visible layer and other are the hidden layer. For gene expression, DNN-1 is set up with one visible layer and three hidden layers. There are 30 visible variable in the visible layer, 800 hidden variables in the first hidden layer, 80 hidden variables in the second hidden layer and 2 hidden variable in the third hidden layer. Similarly, for clinical data, DNN-2 is set up with one visible layer and one hidden layer. There is 32 visible variable in the visible layer and one hidden variable in the hidden layer. The joint representation shows the top hidden layer with two hidden variable which is the class label.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

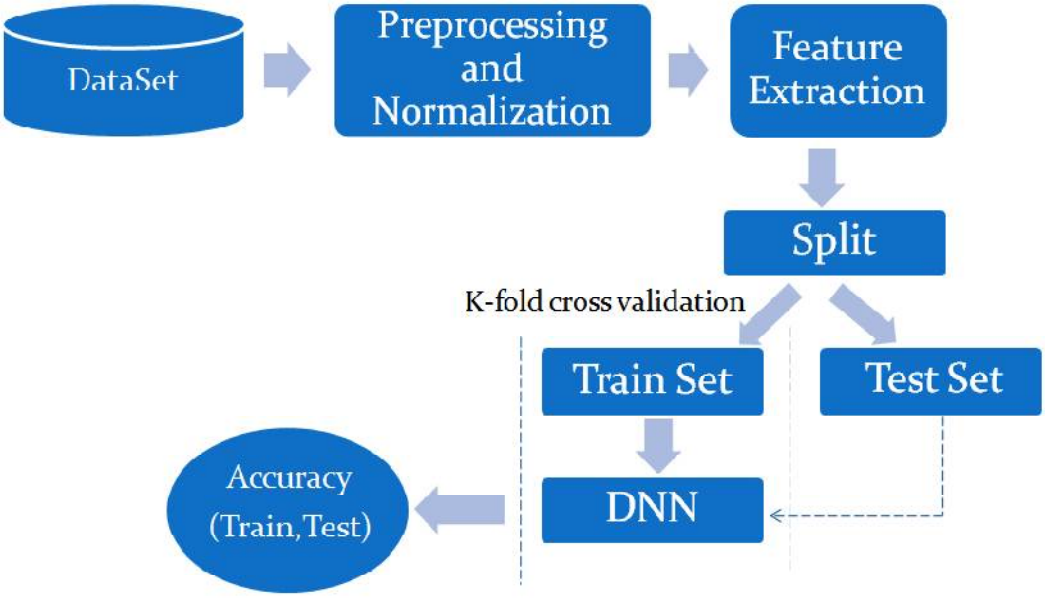

Fig 2: System Architecture

The conditional density distribution is calculated as follows:

$$P(hl_j = 1|vl) = sig(\sum_{i=1}^{m} W_{ij} vl_i + hbias_j) \quad \text{.....(5)}$$

where hlj is the hidden layer variable, vl is the visible variable vector, m is the number of visible layers, Wij is the weight matrix, vli is the visible variable and hbiasj is the bias of hidden layer. In comparison to equation (1), we utilize g(_) as sig(_) transform function. The conditional density distribution uses the sigmoid function, sig(y); which given as follows:

$$sig(y) = \frac{1}{1+e^{-y}} \quad \text{.....(6)}$$

The learning algorithm is known as Contrastive Divergence (CD) is used to infer the variables in the layer of DNN. CD is
applied using Gibbs sampling for kth times. For which visible and hidden layer are reconstructed for kth times. The vg is
reconstructed by conditional density distribution as follows:

$$P(vl_i = 1|hl) = sig(\sum_{j=1}^{n} W_{ij} hl_j + vbias_i) \quad \text{.....(7)}$$

where hli is the visible layer variable, hl is the hidden variable vector, n is the number of hidden layers, Wij is the weight matrix, hlj is the hidden variable and vbiasj is the bias of visible layer. The sigmoid function, sig(y), is given in equation (5). Equation (6) compared with equation (2), results in fewer computational cost and result in faster execution.

When learning is completed, the values of delW is the weight matrix, delV bias is the bias vector of the visible layer and delHbias is the bias vector of the hidden layer are updated by using following equation:

$$delW = delW + p(HL_i = 1|vl^{(0)}) \cdot vl_j^{(0)} - p(HL_i = 1|vl^{(r)}) \cdot vl_j^{r} \quad \text{.....(8)}$$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

$$delHbias = delHbias + p(HL_i = 1|vl^{(0)}) - p(HL_i = 1|vl^{(k)}) \quad \text{.....(9)}$$

$$delVbias = delVbias + vl_j^0 - vl_j^k \quad \text{.....(10)}$$

In the equation (7), the computation is done using the dot product between the constructed hidden variables from the first Gibbs sampling and first visible vector. Second dot product in the same equation is done between the constructed hidden variables from the last Gibbs sampling and last visible vector. As compared to the equation (3), computational time complexity is degraded in equation (7) by using the dot product calculation. Then the constructed hidden layer, h1g, is set as the visible layer of the DNN-2 and hidden layer, h2g is constructed using the visible layer of DNN-2. Similar to the construction of hidden layer, h1g. The same process will be done for clinical data. In the joint representation, fuse the top hidden layer of DNN-1 and DNN-2 into one hidden layer and set this hidden layer as the visible layer of DNN. The last hidden layer constructs the data into the different cluster.

E. *Software Requirement Specification:*

- Hardware Configuration

– Processor : Minimum 2.6 GHz and above
– RAM : 2GB
– Hard Disk : 2GB
– Key Board : Standard Windows Keyboard
– Mouse : Two or Three Button Mouse
– Monitor : Super Video Graphics Array (SVGA)

- Software Configuration

– Operating System : Windows 7/8
– Programming Language : JAVA JDK 1.6 (or above)
– IDE : Eclipse or Netbeans for Java

IV. RESULTS AND DISCUSSIONS

A. *Dataset:*

Complete dataset is collected from Breast Cancer Wisconsin (Diagnostic) Dataset which is freely available. The breast cancer dataset contains genetic factor expression and medical information, for example survival time. Netherlands Cancer Institute provides this dataset. It consists of 569 instances and 9 attribute of patients report. These data present in dataset may be used for predicting tumor of patient and there survival time. The Lung Cancer Dataset contains gene expression. Machine learning repository provides this dataset. It consists of 32 instances and 57 attribute. Out of which class label isdiscarded, resulted in 56 attributes.

Dataset	No. of instances	No. of attributes	Size	Year	Publication
Breast Cancer Wisconsin	569	9	19k	1992	Netherlands Cancer Institute
Lung Cancer Dataset (LCD)	32	56	6k	1992	Machine Learning Repository

Table 1: Dataset

B. *Results:*

In the table 2, it show that two algorithms i.e. Stacked autoencode and Principal Component Analysis in combination with Deep Neural Networks are given. Compare the result of this algorithm and show that there is expected increase in accuracy of Principal Component Analysis in combination with Deep Neural Networks as compare with Stacked

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

autoencoders. The dataset is divided into Training set and Testing test by using the K-fold cross validation method. And hence the accuracy will be calculated separately for Training set and Testing set by calculating the Mean Square Estimation(MSE).

Sr. No.	Algorithm	Input	Expected output	Accuracy
1	Stacked Auto-encoder	Attribute value of patient	Show cancer subtype	95%
2	PCA+DNN	Attribute value of patient	Show cancer subtype	96%

Table 2: Expected Result

The accuracy will be measured by the following indicators:

- Mean Square Error: If n predictions is given by vector ^ Y , and observed values of data being predicted is given by vector Y , then mean square error (MSE) of given sample is calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad \text{.....(11)}$$

A) Training set: The dataset is divided into training set and testing set to measure the accuracy of the proposed model. The
data is divided into 70% of training set data. On this data the MSE is calculated and the accuracy as depicted as the MSE
curve which shows the graph as follows:

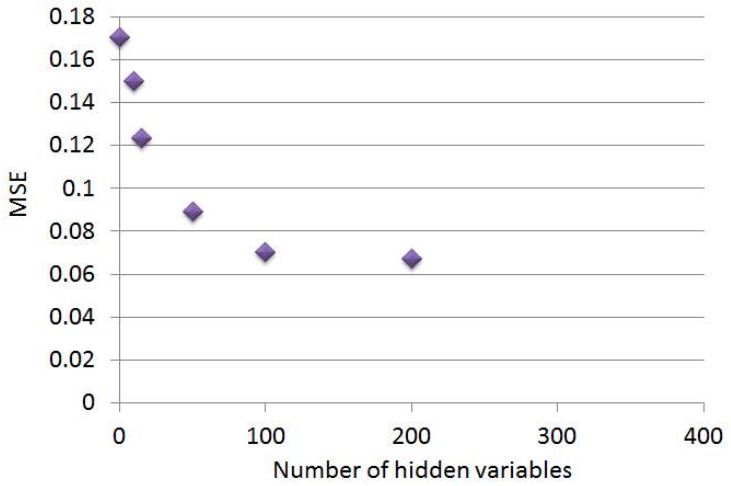

Fig.3. Plot graph between number of hidden layer vs the mean squared error of Breast cancer data.

B) Testing Set: The dataset is divided into training set and testing set to measure the accuracy of the proposed model. The
data is divided into 30% of testing set data. On this data the MSE is calculated and the accuracy as depicted as the MSE curve which shows the graph as follows:

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

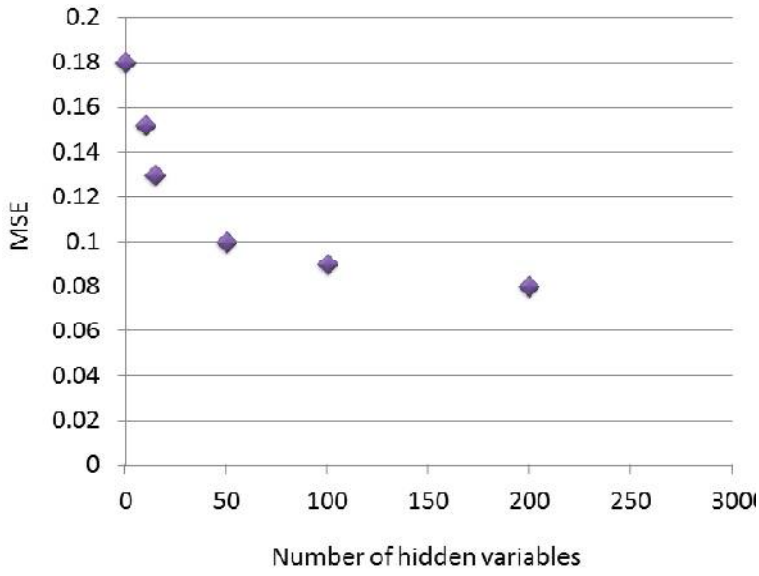

Fig.4 Plot graph between number of hidden layer vs the mean squared error of Lung cancer data.

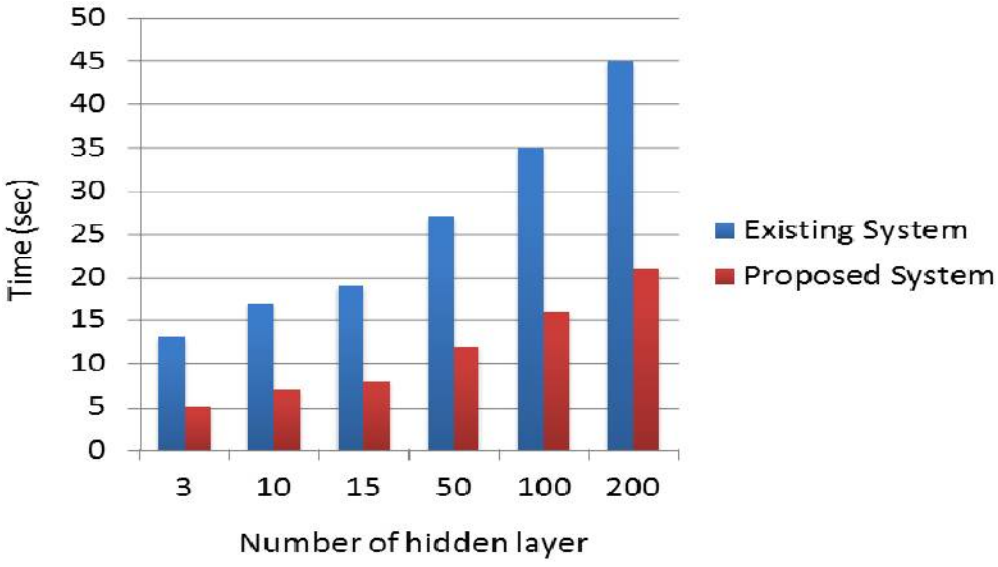

Fig 5. Comparison of Existing and Proposed system

Figure 5 show the comparison between the existing system and proposed system. The x-axis gives the number of hidden layer and y-axis gives the time in seconds. As the bar chart show that by using the enhance RBM the speed of the proposed system is increased.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

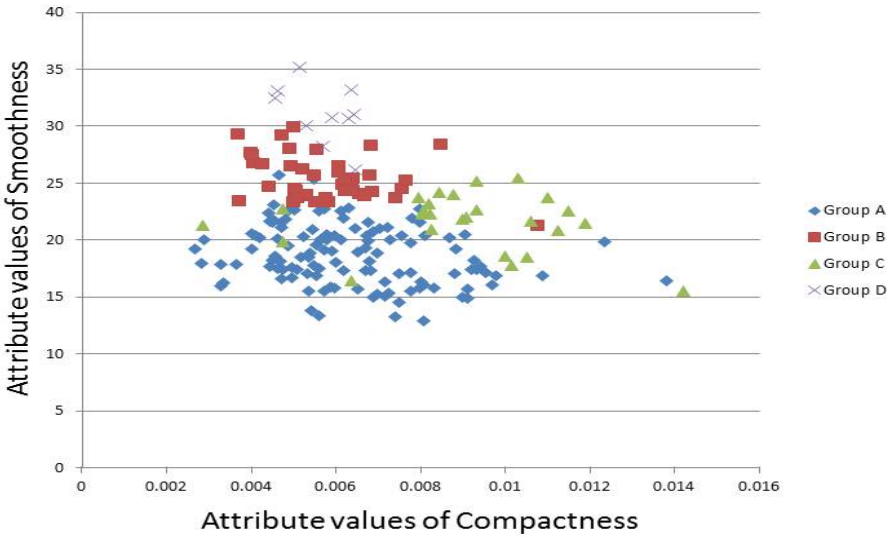

Fig. 6: Cluster dataset into four group according to the clinical data.

Figure 6 illustrates the cluster identified by our proposed system. It successfully identified four clusters out of which 118 the number of samples is identified in cluster 1, 42 number of samples is identified in cluster 2. Similarly, 27 samples in cluster 3 and 10 samples in cluster 4. The scatter point is plotted using the attribute value of patients i.e. on x-axis values of compactness attributes are used and on y-axis values of smoothness attributes are used.

V. CONLCUSION

For disease pathogenesis, identification of tumor subtypes i.e. malignant and benign tumors is very useful. The first step is to cluster cancer data to have a better view of virus causing the disease. Clustering is done according to their modalities such as gene expression and clinical information. Thus, provide a way to have better anticancer treatment for the cancer patients. The removal of irrelevant and insignificant features from large datasets leads to improvement in data mining tasks such as classification. For dimensionality reduction, PCA is most widely used method but it does not provide a subset of real features. On the other hand, filters and wrappers are capable to find reduced subset of features but are computationally expensive. Proposed method provides a way in which PCA can be used for subset selection using feature evaluation and ranking such that accuracy is maintained i.e. it does not fall below the accuracy achieved using all features. Computational cost could be reduced significantly since feature evaluation and ranking is applied only to those features that are selected after PCA step.

VI. ACKNOWLEDGEMENT

I would like to thank the researcher as well as publishers for making their resources available and the teachers for their guidance. I would also thank the college authorities for providing the required infrastructure and support. Finally, i would like to extend a heartfelt gratitude to all friends and family members.

REFERENCES

[1] N. Pochet, F. De Smet, J.A.K. Suykens, and B.L.R. De Moor, "Systematic Benchmarking of Microarray Data Classification: Assessing the Role of Non-Linearity and Dimensionality Reduction", Bioinformatics, vol. 20, no. 17, pp. 3185-3195, 2004.
[2] I.T. Jolliffe, Principal Component Analysis, second ed. Springer, 2002.
[3] S. Bicciato et al., "PCA Disjoint Models for Multiclass Cancer Analysis Using Gene Expression Data", Bioinformatics, vol. 19, pp. 571-578, 2003.
[4] R. Lilien and H. Farid, "Probabilistic Disease Classification of Expression-Dependent Proteomic Data from Mass Spectrometry of Human Serum", J. Computational Biology, vol. 10, no. 6, pp. 925- 946, 2003

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

[5] D.V. Nguyen and D.M. Rocke, ”Multi-Class Cancer Classification via Partial Least Squares with Gene Expression Profiles”, Bioinformatics, vol. 18, pp. 1216-1226, 2002.
[6] D. Huang and D. Zheng, ”Independent Component AnalysisBased Penalized Discriminant Method for Tumor Classification Using Gene Expression Data”, Bioinformatics, vol. 22, pp. 1855- 1865, 2006.
[7] W. Kong, C. Vanderburg, G. Gunshin, J. Rogers, and X. Huang, ”A Review of Independent Component Analysis Application to Microarray Gene Expression Data”, Biotechniques, vol. 45, no. 5, pp. 501-520, 2008.
[8] J. Brunet, P. Tamayo, T. Golub, and J. Mesirov, ”Molecular Pattern Discovery Using Matrix Factorization”, Proc. Natl Academy of Sciences USA, vol. 101, no. 12, pp. 4164-4169, 2004.
[9] K. Devarajan, ”Nonnegative Matrix Factorization: An Analytical and Interpretive Tool in Computational Biology”, PLoS Computational Biology, vol. 4, no. 7, 2008, doi:10.1371/journal.pcbi.1000029.
[10] H. Kim and H. Park, ”Sparse Non-Negative Matrix Factorizations via Alternating Non-Negativity-Constrained Least Squares for Microarray Data Analysis”, Bioinformatics, vol. 23, no. 12, pp. 1495- 1502, 2007.
[11] S. Li, W. Xin, J. Hong, and S. Qian, ”Learning Spatially Localized, Parts-Based Representation”, Proc. Conf. Computer Vision and Pattern Recognition (CVPR 01), vol. 1, pp. i207-i212, 2001.
[12] Y. Gao and G. Church, ”Improving Molecular Cancer Class Discovery through Sparse Nonnegative Matrix Factorization”, Bioinformatics, vol. 21, no. 21, pp. 3970-3975, 2005.
[13] A. Pascual-Montano, J.M. Carazo, K. Kochi, D. Lehmann, and R.D. Pascual-Marqui, ”Nonsmooth Nonnegative Matrix Factorization”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 3, pp. 403-415, Mar. 2006.
[14] C. Ding, X. He, and H. Simon, ”On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering” Proc. SIAM Intl Conf. Data Mining, pp. 606-610, 2005.
[15] J. Weston et al., ”Feature Selection for SVMs” Advances in Neural Information Processing Systems (NIPS), vol. 13, pp. 668-674, MIT Press, 2000
[16] D. Lee and H. Seung, ”Learning the Parts of Objects by NonNegative Matrix Factorization” Nature, vol. 401, pp. 1788-1791, 1999.
[17] Hayder K. Fatlawi, ”Enhanced Classification Model for Cervical Cancer Dataset based on Cost Sensitive Classifier” in International Journal of Computer Techniques- Volume 4 Issue 4, July August 2017
[18] Wen Wu, Hao Zhou, ”Data-driven Diagnosis of Cervical Cancer with Support Vector Machines-Based Approaches”, IEEE ACCESS, volume 7, pp 25189- 25195, Oct 2017
[19] Suja, S., George, N., George, A. (2018). ”Classification of Grades of Astrocytoma Images from MRI Using Deep Neural Network.” 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). doi:10.1109/icoei.2018.8553882