International Journal of Scientific Research and Engineering Development-– Volume 4 Issue 6, Nov-Dec 2021

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Cancer Classification on Gene Expression Data Using Semi Supervised Learning Methods

P.Deepa[1], Dr.G.Tamilpavai. Phd [2]

[1], M.E Department of computer science Engineering , Government College of Engineering, Tirunelveli, India. Email: pdeepapauldurai97@gmail.com

[2] Associate Professor (CAS),Department of computer science Engineering , Government College of Engineering, Tirunelveli, Tirunelveli, India. Email:tamilpavai@gcetly.ac.in.

Abstract:

Accurate identification of cancer types is essential for cancer diagnoses and treatments. Now a day's accurate cancer classification is a challenging task. Normally, cancer tissue and normal tissue have different characteristics on their gene expression data. Gene expression data can be used as an efficient feature for cancer classification. It has a high dimension feature and limited data samples. Here, a new self-training subspace clustering (SSC) algorithm is introduced under low rank representation (LRR), called SSC-LRR for accurate cancer classification on gene expression data. In this proposed work LRR is used for converting the high dimension data into low dimension data. It is done by extracting the internal structure of gene expression data and also it uses a large volume of sample data. The efficiency of LRR and self-training has been examined by the step by step incorporation with baseline K means clustering algorithm. SSC method is used to generate the cancer classification predictions which yield better accuracy in comparison to traditional approaches.

Keywords: Cancer classification, Gene expression data, K-means clustering, Low rank representation, SSC-LRR.

I. INTRODUCTION

Bioinformatics is used to understand biological data by using methods and software tools. Understanding biological data may concern the fields of bioinformatics such as science, biology, Computer science, information engineering, mathematics and statistics. These computer based biological methods are used in genomics.Common uses of bioinformatics including the identification of candidate's genes

It also plays a major role in the analysis of gene expression and regulation. Bioinformatics tools used for compare, analyze and interpret genetic data and also understanding of evolutionary aspects of molecular biology. It is used for analyze and catalogues the biological pathways and networks that are an important part of systems biology. In structural biology, it aids in the simulation and modeling of DNA, RNA, proteins as well as bio molecular interactions. It is an interdisciplinary field is mainly involving molecular biology and computer science. From beginning to end the lives, healthy cells in human bodies divide and replace themselves in a controlled manner. When a cell is modified, it multiplies out of control. It creates cancer. Mass composed of a cluster of such abnormal cells is called a tumor. Not all tumors are cancerous but most cancers form tumors. Non cancerous tumors act not spread to other parts of the body and do not create new tumors. But cancerous tumors crowd out healthy cells and interfere with body functions, and draw nutrients from body tissues. Cancers continue to grow and spread by two ways. One is direct extension of the tumors. Another one is through a process called metastasis. In this progression, the

malignant cells travel through the lymphatic or blood - vessels eventually forming new tumors in other parts of the body. In worldwide cancer is a major and serious public health problem for human beings. In day by day, there is increasing death count in worldwide due to cancer. It is necessary to identify cancer types accurately for cancer diagnosis and prognosis. Accurate cancer classification has become more important in the field of cancer research. Traditional approaches to cancer classification concerning on interpretation of clinical and histopathological information of the patient in the clinical diagnoses and prognoses. It can lead to uncertain results even for the same cancer patient, because of the subjective interpretations and doctor's personal experience. . The word "Cancer" containing more than 100 diseases affecting nearly every part of the body, and all are potentially living aggressive diseases. The major types of cancer are lung, prostate, colorectal, breast, lymphoma, bladder, melanoma, utters, ovary, renal, pancreas. Leukemia cancer type gene and the most commonly diagnosed cancers begin in the skin, lungs, breasts, pancreas and other organs. Lymphomas are cancers of lymphocytes. Leukemia is cancer of the blood. It does not usually form solid tumors. Sarcomas cancer tissues appear in bone, muscle, fat, blood vessels, cartilage, or other soft or connective tissues of the body. These are relatively uncommon. Melanomas appear in the cells that create the pigment in skin Cancer. Cancer specialists who called oncologists encompass remarkable advances in cancer diagnosis, prevention, and treatment. In now a day's more people diagnosed with cancer are living longer. But still, some forms of the disease remain exasperatingly difficult to treat, but now a day the advanced Modern treatment can extensively improve quality of life and may make longer survival

II. RELATED WORK

Q. Liao et al. [1] In this paper, it proposes a Gauss-Seidel based non-negative matrix factorization (GSNMF) method to beat such imbalance deficiency between features and samples. Based on the projected data, GSNMF iteratively projects gene expression data onto the learned subspace and it is followed by adaptively updating the cluster centroids. While this data projection strategy considerably reduces the influence of imbalance between the number of samples and the number of genes. Ever since it uses solution of a linear system obtained by the Gauss-Seidel method and updates each factor matrix by, it converges rapidly without neither complex line search nor matrix inverse operators. It analyze the error bound and obtain a local minima can reduce the influence of imbalance between the number of genes and the number of samples for gene expression clustering. This method is insufficiently, because it is difficult to choose primary genes based only on few samples.

X. Zhang et al [2] proposed this work introduce a semi-supervised projective non-negative matrix factorization method (Semi-PNMF) toward learn an effective classifier thus boosting sub sequent cancer classification performance. It uses both labeled and unlabeled samples. It incorporates statistical information and learns more representative subspaces and boost classification performance from the large volume of unlabeled samples in the learned subspace. In this paper, it developed a multiplicative update rule (MUR) to optimize Semi-PNMF and proved its convergence. In cancer data processing it provide a flexible framework for learning methods. But it does not explicitly guarantee and also degrades the clustering performance by only identify the meta patterns of various cancers for identifying different types of tumors

A. Alder et al. [3] uses a novel semi-supervised classification method 'self-training' based fuzzy K Nearest Neighbor algorithm which is improving the prediction accuracy of the cancer classification by utilizes the unlabeled samples along with the labeled samples. Proposed work performance is compared with its two other supervised counterparts namely, K-NN and fuzzy KNN classifiers and two non-fuzzy, non-NN based methods namely SVM and Naive Bayes classifier, but this proposed work

performance is higher than those methods But the unlabeled samples are relatively inexpensive and readily available. The sufficient numbers of labeled samples are very expensive and difficult to collect therefore the calculated accuracies of the classifiers trained with limited training samples are often very low.

A. alder et al [4] Introduce a novel local and global preserving semi - supervised dimensionality reduction based on random subspace algorithm, which utilizes random subspace semi- supervised dimensionality reduction, is proposed. In this algorithm initially designs multiple diverse graphs on different random subspace of datasets. Then dimensionality reduction is performed by fusing the graphs into a mixture graph. This mixture graph is constructed in lower dimensionality. It can relieve the issues on high dimensional graph construction on gene data samples. It holds difficult geometric distribution of datasets in the diversity of random subspaces. Public gene expression datasets experimental results show that the proposed work not only has performance on superior recognition for competitive methods. But also against a wide range of values of input parameters in robust. In this paper, it presents a novel local and global preserving semi -supervised dimensionality reduction. This method is not only for develop side efficiently, but also is robust to noise. But it is based on the local and global assumptions and uses Euclidean distance to define the neighborhood, so it cannot be used in real world.

X. Y. Chen [5] Proposed graph regularized subspace segmentation method (GRSS) for clustering gene expression data. Traditional NMF (non-negative matrix factorization) methods cannot deal with negative data and easily lead to local optimum because the iterative methods are adopted to solve the optimal problem. By solving a Sylvester equation provide global optimal solution of GRSS. GRSS is also a spectral clustering based subspace segmentation method. In this paper, author propose a novel graph regularized subspace segmentation method GRSS for gene expression data clustering. The main purpose of this paper is to solve those problems of NMF-based clustering methods and also Subspace segmentation is a powerful tool for clustering image the advantage of GRSS is mainly due to combining graph regularization with sub- space segmentation for modeling the intrinsic geometrical structure of the data space. But there is problem on this method that how to select parameters of GRSS. It makes the method the complex one.

III. PROPOSED WORK

This proposed system introduce a new self training subspace clustering algorithm under low-rank representation, called SSC -LRR, for cancer classification on gene expression data. Low-rank representation (LRR) algorithm is first applied to extract discriminative features from the high-dimensional gene expression data and provide low dimensional gene expression data profiles. Then self-training subspace clustering (SSC) method is used to predict the cancer classification predictions. In conclusion the experimental result shows the proposed method improve the classification for the given gene expression data profiles.

A. Dataset

In this project GCD dataset is used, which was created by Ramaswamy et al [11] and is available in online at http://portals.broadinstitute.org/cgi bin/cancer/datasets.cgi. It consists of the gene expression profiles of 218 tumor samples that representing 14 common human cancer types, and each sample contains 16,030 gene expression values. The 14 common cancer types are as they follow:

S.No	Cancer Type
1	Lung Cancer
2	prostate cancer
3	colorectal cancer
4	Breast cancer
5	Lymphoma cancer
6	Bladder cancer
7	Melanoma cancer
8	Utters cancer
9	CNS cancer
10	Ovary cancer
11	Mesothelioma cancer
12	Renal cancer
13	Leukemia cancer
14	Pancreas cancer

Table 3.1 Cancer Types

It is further divided into three subsets: a training subset of 144 samples, a testing subset of 54 samples, and a subset of 20 poorly differentiated samples (tumors). Since the poorly differentiated samples might induce a biased evaluation result, only the well differentiated samples are taken into consideration. The gene data represented as data matrix and the format of the matrix is $X=[x_1, x_2, x_{3.....}x_{n]\epsilon R}{}^{d*n}$. Here X denotes the value of the gene expression data.

In which each column is the d-dimensional feature vector of a sample (gene expression data), and n is the total number of samples. It contains clinical data samples include 14 cancer types of different patients from various hospitals, each element in the matrix represents the pixel value of the cancer patient clinical readings. The samples in the dataset arranged according to their class labels in ascending order. Figure 3.1 representing the block diagram of the proposed work.

B. Pre-Processing

The gene expression contains noise of very low values and the saturation effects of very high values. Hence it is very hard to access the gene expression data with the noisy values. So it is very essential to pre-process the data and remove the noise of very low values and the saturation effects of very high values. This is done by step by placing the gene expression. Data into a specific box constraint ranging from 20 to 16,000 units and then excluding those genes whose ratios across samples are fewer than 5 and absolute variations across samples are under 500, respectively. Here the absolute variance value is finding by using the following steps:

1. Find the mean value of the data
2. Subtract the mean value by each row of the data
3. Square the subtracted result
4. Add the results together and get the absolute variance value.

The mean value is calculated using the following formula:

$$\text{Mean} = \sum_{i=1}^{n} \frac{X^i}{N}$$

Here,

n- Number of samples.

x- Value of the gene data.

N-Total length of the data

The original GCM data set has a dimension of 16,030 rows and 144 columns. After the Pre Process the dimension is reduced to 13083 rows and 144 columns. This is called R-GCM representing Reduced-GCM dataset.

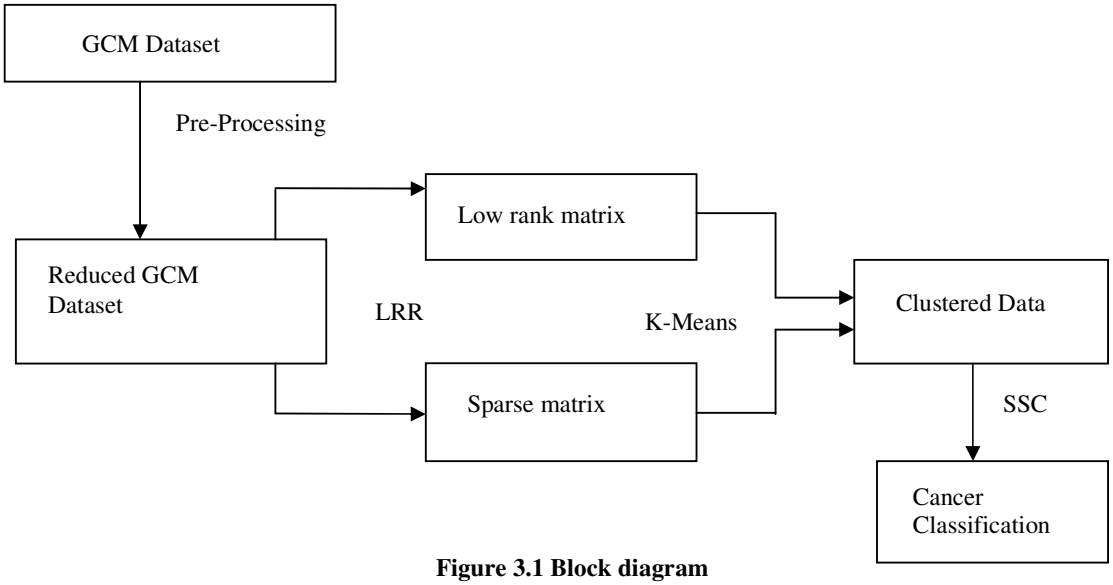

Figure 3.1 Block diagram

C. Low –Rank Representation Algorithm

Low Rank representation algorithm method seeks the lowest rank representation among all the candidates that can represent the data points as linear combinations of the bases in a given dictionary. Then nuclear norm minimization function (1) is used to minimize the rank of the representation matrix. Let X be an original data matrix, among which each column is the d dimensional feature vector of a sample (gene expression data in this study), and n is the total number of samples. Then LRR is performed by the following way:

LRR Procedure:

Step 1: Perform LRR on original data matrix X. First, apply LRR to the original data matrix X, and the decomposed low-rank matrix (Z)and sparse matrix (E)are obtained.

$$\text{Min}_{Z,E\|Z\|+} \ \ell \ \|E\|_{2,1;} \text{—} (1)$$

ℓ-Control parameter of LRR.

X=XZ+E

Re-arrange Z and E as Z = [Z_l, Zu] and E =[E_l, E_u] , respectively;

Current Item Num<-0; // Counter of clustering iterations.

Step2: The matrix Z can be divided into a labeled matrix and unlabeled Matrix. Similar to E Matrix can be divided into a labeled matrix and unlabeled Matrix.

D. K- Means Clustering

K-means clustering algorithm is performed on Z and E, respectively. The key problem for performing K-means is how to initialize the central points of clusters. Taking Z = [Z_l, Zu] as an example, the initial point of cluster (class) I can be determined by

$$\mathbf{p}^{(i)} = \frac{\sum_{j=1}^{n_l^{(i)}} \mathbf{z}_{l,j}^{(i)}}{n_l^{(i)}}$$

Based on the initial central points of clusters obtained, perform the standard K-means algorithm on matrix Z until each of the unlabeled samples is clustered into one of the C clusters. According to the clustering results on Z, the labels of those unlabeled samples are predicted. The predicted labels of Z_u, together with the labels of Z_l form the label vector of Z, denoted as Z_l. This procedure can be formulated as follows:

Z_l = K-means(Z,distZ)

Similarly, it can obtain the label vector of E , denoted as lE, by using the same procedure of obtaining lZ,

$$E_l = \text{K-means}(E, \text{dist}E)$$

where dist E denotes the distance metric used for clustering E . The K-means algorithm outlined above can be easily extended, with any other appropriate distance scales, to facilitate different application scenarios of data clustering problems.

E. Self- Training Procedure

The self-training procedure is introduced that having the following procedure for classification. That indicating selects unlabeled samples as labeled ones for next round clustering. After obtaining the clustering results, i.e., Z_l and E_l, unlabeled samples to be selected is decided and used as labeled samples for the next round clustering. An unlabeled sample, say sample Zu, will be selected as the labeled data for the next round clustering if and only if

$$Z_l = le$$

Where Zl is the predicted label of the unlabeled sample according to the clustering results of Z_l, and i le is the predicted label of the unlabeled sample according to the clustering results of E_l. All the unlabeled samples satisfying constitute the set of chosen samples, denoted as S, for next round clustering.

The algorithm is made with respect to S. If S =φ or current iteration number is greater than the predefined iteration number, the procedure is stopped and Z_l is returned as the final clustering results. Otherwise, Z and E are updated as follows: For each selected unlabeled sample i in S , then update Z_l , and u Z by moving u Z_l from Z_u to Z_l. Similarly, we update E_l and E_u by moving u from E_u to E_l. The updated Z_l and Zu will be merged as new Z, and the updated E_l and E_u will be merged as new E, for next round clustering. After this update step, the procedure goes to K means clustering for next round clustering.

F.Performance Metric

In this proposed work the performance is evaluated by Recall (REC), False Positive Rate (FPR), and Matthews Correlation Coefficient (MCC), Overall Accuracy (Q) and Generalized Correlation (GC) metrics. The performance evaluation is performed by using the following formulas:

Recall:

$$\mathbf{REC(i) = TP(i)/TP(i) + FN(i))}$$

False positive rate:

$$\mathbf{FPR = FP(i)/(TN(i) + FP(i))}$$

Matthews's correlation coefficient:

$$\frac{\mathbf{(TP(i).TN(i) - FP(i).FN(i)))}}{\sqrt{\mathbf{(TP(i)+FP(i)).(TP(i)+FN(i)).(TN(i)+FP(i)).(TN(i)+FN(i))}}}$$

Overall accuracy:

$$\mathbf{Q} = \frac{\sum_{i-1}^{c} \mathbf{TP(i)}}{\mathbf{N}}$$

Generalized Correlation:

$$\mathbf{GC} = \sqrt{\frac{\sum_{i-1}^{C}\sum_{i-1}^{C}\frac{m_{ij}-e_{ij}}{e_{ij}}}{\mathbf{N(c-1)}}}$$

Using this performance metrics Performance evaluation is performed for the proposed work**.**

IV. RESULTS

A. Dataset

In this project benchmark GCD dataset is used, which was created by Ramaswamy et al [11]. It is publically available downloaded from http://portals.broadinstitute.org/cgi/bin/cancer/datasets.cgi.

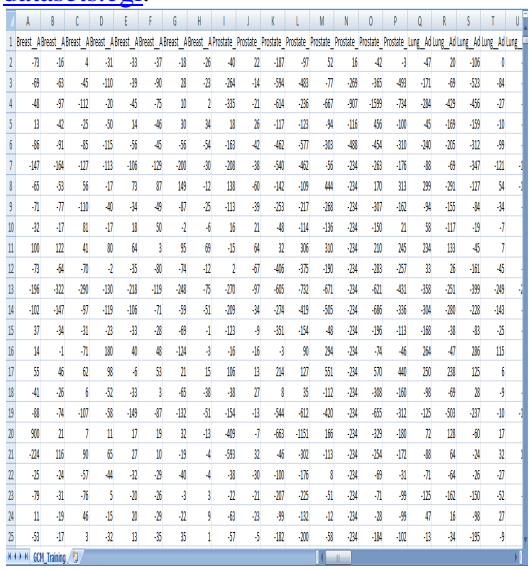

Figure 4.1 Dataset Example Samples

B. Pre-Processing

The GCM data set has a dimension of 16,030 rows and 144 columns. After the pre-processing the dimension is reduced to 13083 rows and 144 columns. This is called R-GCM representing Reduced-GCM dataset

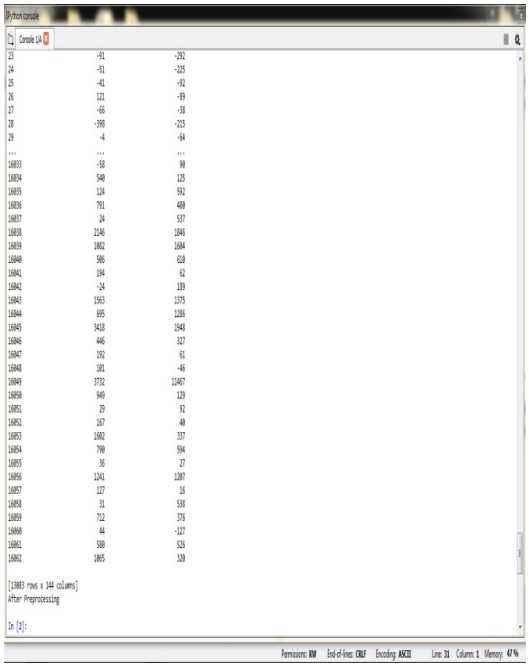

Figure 4.2 Pre Processing Data

C. Low Rank Representation Algorithm

In this module the low rank representation algorithm is applied to the pre processed input and it provide the following two matrices.

1. Low-Rank Matrix

Here it is spilt into the following two types of matrixes.

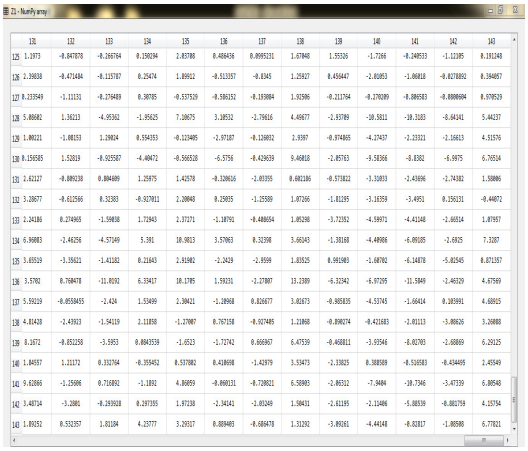

Figure 4.3 Labeled Low Rank Matrix

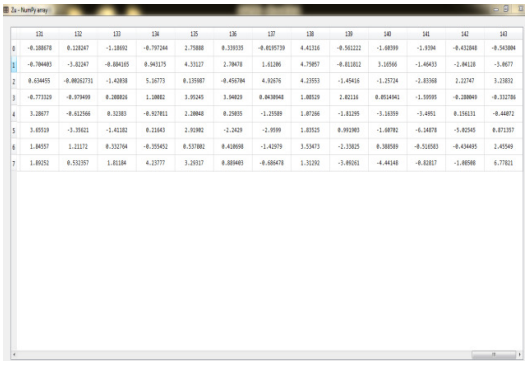

Figure 4.4 Unlabeled low rank Matrix

2.Sparse Matrix

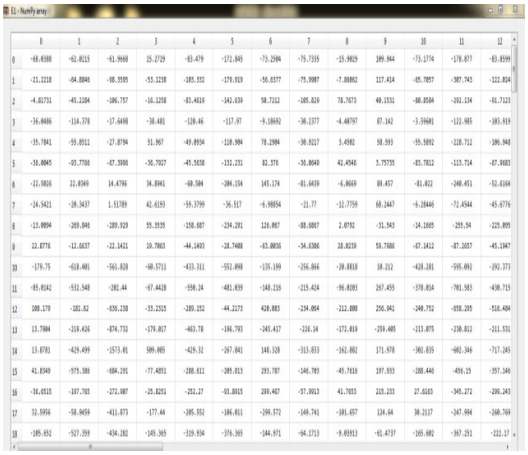

Figure 4.5 Labeled sparse matrix.

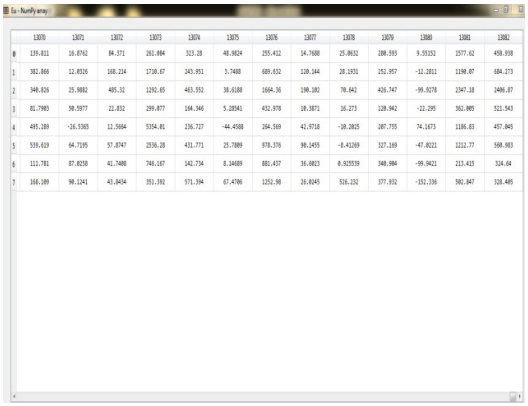

Figure 4.6 Unlabeled sparse matrix.

D. K- Means Clustering

Here the K means clustering algorithm is applied to the low rank matrix and sparse matrix and the output is centroids points of each cluster which indicating 14 cancer types

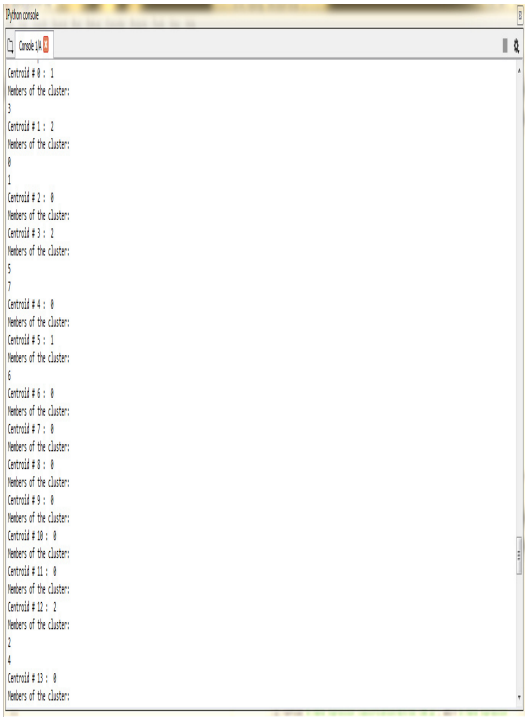

Figure 4.6 Clustered Data

E. Self Training Procedure

In this module, it introduces a self-training procedure and the procedure is stopped with updated low -rank matrix and sparse matrix and returns the final predicated clustering results

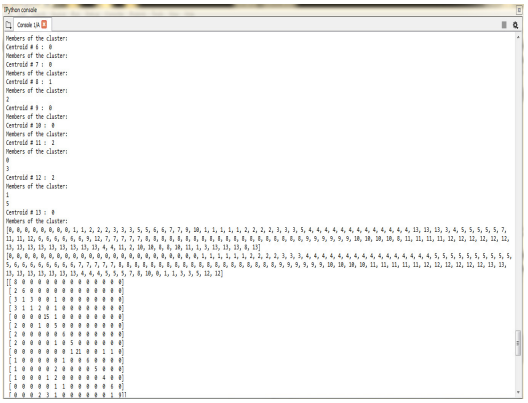

Figure4.7 Predicated Cancer classification

E. Performance Metric

Performance evolution of the proposed work.

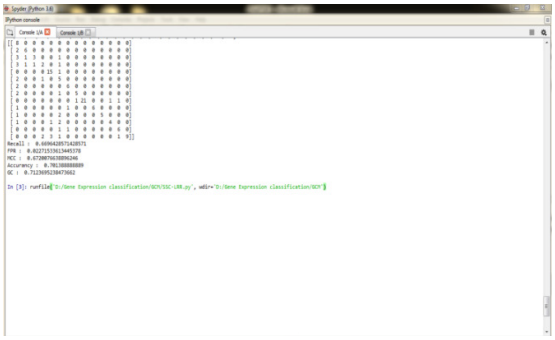

Figure 4.8 Performance Metric

In this proposed work the performance metrics are measured by calculating the following measures:

Recall	0.66964
False Positive Rate	0.02271
Matthews correlation coefficient	0.67200
Overall Accuracy	0.70158
Generalized Correlation	0.712369

Table 4.1 confusion matrix.

V. CONCLUSION & FUTURE ENHANCEMENT

Traditional cancer classification approaches face lots of difficulties such as high dimensionality, small sample size, and enrichment of unlabeled samples. To overcome these difficulties, a semi-supervised self-training subspace clustering algorithm

based on low rank representation, called SSC-LRR is proposed. LRR is introduced to relieve the difference between High-dimensionality and small sample size data features, by the Extraction of the intrinsic structure of gene expression data, which are then encoded into low-dimensional discriminative Features. The efficiency of LRR and self-training has been examined by step by step incorporation of baseline K-means clustering algorithm. Then a self –training procedure on evolution returns the final predicated clustering results. On evaluation, SSC-LRR achieved an overall accuracy 70.15% and a General correlation 0.712, which is 18.9% and 24.4% higher than that of the state -of-art methods. Despite the encouraging results of SSC-LRR, there is still considerable room for further improvement. First, more efficient methods are needed for identifying the intrinsic structure of gene expression data to extract more discriminative features for cancer classification. K-means clustering algorithm was explored in this study for implementing the SSC-LRR in further enhancement considering various clustering and algorithms.

REFERENCES

1. Q. Liao, N. Guan, and Q. Zhangg, "Gauss-Seidel based non-negative matrix factorization for gene expression clustering," in 2016 IEEE International Conference on Acoustics, Speech andSignal Processing (ICASSP), 2016, pp. 2364-2368.
2. X. Zhang, N. Guan, Z. Jia, X. Qiu, and Z. Luo, "Semi-Supervised Projective Non-Negative Matrix Factorization for Cancer Classification," PLoS One, vol. 10, p. e0138814, Sep 22 2015
3. A. Halder and S. Misra, "Semi-supervised fuzzy K-NN for cancer classification from microarray gene expression data," 2014 FirstInternational Conference on Automation, Control, Energy &Systems (ACES-14), pp. 266-270, 2014.
4. A. Halder and S. Misra, "Semi-supervised fuzzy K-NN for cancer classification from microarray gene expression data," 2014 FirstInternational Conference on Automation, Control, Energy &Systems (ACES -14), pp. 266-270, 2014.
5. X. Y. Chen and C. R. Jian, "Gene expression data clustering based on graph regularized subspace segmentation," Neurocomputing, vol. 143, pp. 44-50, Nov 2 2014.
6. Y. Tan, L. Shi, W. Tong, and C. Wang, "Multi-class cancer classification by total principal component regression (TPCR) using microarray gene expression data," Nucleic acids research, vol. 33, pp. 56-65, 2005.
7. Y. Piao, M. Piao, K. Park, and K. H. Ryu, "An ensemble correlation based gene selection algorithm for cancer classification with gene expression data," Bioinformatics, vol. 28, pp. 3306-3315, Dec 2012.
8. G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in Proceedings of the 27th internationalconference on machine learning (ICML-10), 2010, pp. 663-670.
9. R. Díaz-Uriarte and S. A. De Andres, "Gene selection and classification of microarray data using random forest," BMCbioinformatics, vol. 7, p. 1, 2006.
10. Y. Liu, "Active learning with support vector machine applied to gene expression data for cancer classification," Journal of chemical information and computer sciences, vol. 44, pp. 1936-1941, 2004
11. S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M.Angelo, et al., "Multiclass cancer diagnosis using tumor gene expression signatures," Proceedings of the National Academy ofSciences of the United States of America, vol. 98, pp. 15149-15154,Dec 18 2001.
12. K. Prokopiou, E. Kavallieratou, and E. Stamatatos, "An Image Processing Self Training System for Ruling Line Removal Algorithms," 2013 18th International Conference on Digital Signal Processing (DSP), pp. 1-6, 2013.
13. X. R. Zhao, N. Evans, and J. L. Dugelay, "Semi-Supervised Face Recognition with Lda Self-Training," 2011 18th IEEE International Conference on Image Processing, pp. 3041-3044, 2011.

14. X. Zhu, "Semi-Supervised Learning Literature Survey," Computer Science, vol. 37, pp. 63-77, 2008.
15. Y. Y. Xu, F. Yang, Y. Zhang, and H. B. Shen, "An image-based multi-label human protein subcellular localization predictor (iLocator) reveals protein mislocalizations in cancer tissues," Bioinformatics, vol. 29, pp. 2032-40, 2013.
16. X. Zhu, H. I. Suk, L. Wang, S. W. Lee, and D. Shen, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," Medical Image Analysis, vol. 75, pp. 570-577, 2015.
17. Z. S. Wei, K. Han, J. Y. Yang, H. B. Shen, and D. J. Yu, "Protein-protein interaction sites prediction by ensembling SVMand sample-weighted random forests," Neurocomputing, vol. 193,pp. 201-212, 2016.
18. S. D. Konduri, K. S. Srivenugopal, N. Yanamandra, D. H. Dinh, W. C. Olivero, M. Gujrati, et al., "Promoter methylation and silencing of the tissue factor pathway inhibitor-2 (TFPI-2), in human glioma cells," Oncogene, vol. 22, pp. 4509-16, 2003.
19. S. Kurscheid, P. Bady, D. Sciuscio, I. Samarzija, T. Shay, I. Vassallo, et al., "Chromosome 7 gain and DNA hypermethylation at the HOXA10 locus are associated with expression of a stem cell related HOX-signature in glioblastoma," Genome Biology, vol. 16, pp. 1-15, 2015.
20. Y. Zhu, S. Ren, T. Jing, X. Cai, Y. Liu, F. Wang, et al., "Clinical utility of a novel urine-based gene fusion TTTY15-USP9Y in predicting prostate biopsy outcome," Urologic Oncology, vol. 33, pp. 384.e9 384.e20, 2015.
21. H. Zhang, C. Zhu, Y. Zhao, M. Li, L. Wu, X. Yang, et al., "Long non-coding RNA expression profiles of hepatitis C virus related dysplasia and hepatocellular carcinoma," Oncotarget, vol. 6, pp. 43770 43778, 2015.
22. M. Condomines, D. Hose, T. Rème, G. Requirand, M. Hundemer, M. Schoenhals, et al., "Gene expression profiling and real-time PCR analyses identify novel potential cancer testis antigens in multiple myeloma," Journal of Immunology, vol. 183, pp. 832-40, 2009.
23. M. T. Dorak, F. S. Oguz, N. Yalman, A. S. Diler, S. Kalayoglu, S. Anak, et al., "A male-specific increase in the HLA-DRB4 (DR53) frequency in high-risk and relapsed childhood ALL," Leukemia Research, vol. 26, pp. 651-656, 2002.
24. X. Zhu and Z. Ghahramani, "Learning from Labeled and Unlabeled Data with Label Propagation," 2003.
25. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer Statistics, 2016," Ca-a Cancer Journal for Clinicians, vol. 66, pp. 7-30, Jan-Feb 2016.

AUTHORS PROFILE

P.Deepa , She completed her B.E in Computer Science and Engineering in Government College of Engineering, Tirunelveli, Tamil Nadu, India. She completed her P.G in Computer Science and Engineering in Government College of Engineering, Tirunelveli, Tamil Nadu, India. Her research interest consists of medical image processing and bio-informatics.

Dr.G.Tamilpavai, she completed her B.E in Computer Science and Engineering from Thiagarajar College of Engineering, Madurai, Tamil Nadu, India. She did her P.G in Government College of Engineering, Tirunelveli, Tamil Nadu, India. She Completed her Ph.D. at Anna University, Chennai, Tamil Nadu, India. Her area of interest includes medical image processing, remote sensing, bio-informatics and operating systems. She is working as Associate Professor (CAS) and Head in Department of Computer Science and Engineering at Government College of Engineering, Tirunelveli. She has 20 years of teaching experience. She is recognized guide in Anna University, Chennai, Tamil Nadu, India. She has 16 publications in international journals especially in biomedical image processing and bio informatics. She has published many papers in National and International conferences. She has life membership in ISTE, IE and BMESI.