1. $w_h=A_hq_h$, $c_h=w_h'M_hw_h$, $w_h=w_h/sqrt(c_h)$, and store $w_h$ into W as a column
2. $p_h=M_hw_h$, and store $p_h$ into P as a column
3. $q_h=A_h'w_h$, and store $q_h$ into Q as a column
4. $v_h=C_hp_h$, and $v_h=v_h/\|v_h\|$
5. $C_h+1=C_h - v_hv_h'$ and $M_h+1=Mh - p_hp_h'$
6. $A_h+1=C_hA_h$

The T of SIMPLS is computed as T=XW and B for the regression of Y on X is computed as B=WQ'[1].

**5.2.1 Total PLS Algorithm**

Total PLS algorithm achieves both PLS-based feature selection and feature extraction in a unified PLS framework, is called as Total PLS dimension reduction Algorithm. Steps for the Total PLS Algorithm are as follows.

Input: Train Xn×p, ClsYn×1, Dim

Output : XScore ‖ XScore is the score on PLS-based latent factor

(1) Initialization
-Encode class label ClsYn×1 and generate Class ClsYn×g
-Set nfac = unique(ClsY) ‖ unique(ClsY) indicates the number of category

(2) Feature Selection
-Obtain idx
PLSRFE(TrainX,ClasssY,nfac)
-Update TrainX, whose features only include top Dim features in idx

(3) Feature Extraction
-For j=1 to nfac do
-Calculate score matrix Tj=<TrainX,wj>
-Update Xscore so that Xscore = [Xscore,Tj]
Return Xscore[1]

**5.3. Hybrid Algorithm:**

In hybrid algorithm Total PLS Algorithm and MINE Algorithm is combined. We have applied the hybrid algorithm to improve the performance of the dimension reduction of the cancer microarray data. The Hybrid algorithm is applied on the microarray data to improve the performance of the dimension reduction method. Steps for the Total PLS algorithm as above and MINE algorithm are as follows.

**5.3.1 Maximal Information- Based Nonparametric Exploration (Mine) Method:**

The MINE method is used in statistics. In the MINE Method MIC (Maximal Information Coefficient) plays an important role. MIC is the larger part of the MINE technique. The idea of the MIC is that if the relationship is existing in between the two variable data.

The steps for the MIC are as:

1. Estimating the range of rows and columns i.e. x and y respectively for grid resolution plot two variable data.
2. Partitioning the data. In this step, Grid can be drawn on the disperse plot of the two variables that partitions the data for sum up that relationship of the data.
3. Placements the partitions. In this step the partitions of the data will be placed after grid resolution of the data.
4. Calculating the MIC (Maximal Information Coefficient). Here, to work out the MIC of a set of data which is two variable data, we determine all grids up to a maximal grid declaration, reliant on the sample dimension of the data.
5. Finding the highest mutual information to store each resolution. Reliant on the sample dimension, work out for every pair off of integers (x,y) the major probable common in sequence attainable by any x-by-y grid applied to the data. We name the attribute matrix M = (mx,y), where mx,y is in which the will be occurred the highest possible common information.
6. Normalization. We then normalize these common information values to make certain a fair comparison between grids of different dimensions and to get modified values among 0 and 1.
7. Storing normalized mutual information in the characteristic matrics M(x,y). We define the characteristic matrix M = (mx,y), where mx,y is the highest normalized common information attained by any x-by-y grid, and the statistic MIC to be the maximum value in M [3].

## 6. Results

In the experiments, we have examined the performance of Total PLS (partial least algorithm), Maximal Information-based Non parametric Exploration (MINE) Method. As a result, Hybrid algorithm is the combination of both algorithms (Total PLS and Maximal Information-based Non parametric Exploration (MINE) Method) that is Hybrid algorithm is drastically improved against those of Total PLS algorithm. We present a technique that improves performance of microarray data dimension reduction. We studied and implement dimension reduction algorithm for enhancing performance of dimension reduction.

Figure 3 shows the redundancy of the data by using Hybrid Algorithm which has the improved performance than previous algorithm. In this figure the non repeated and repeated data are shown.

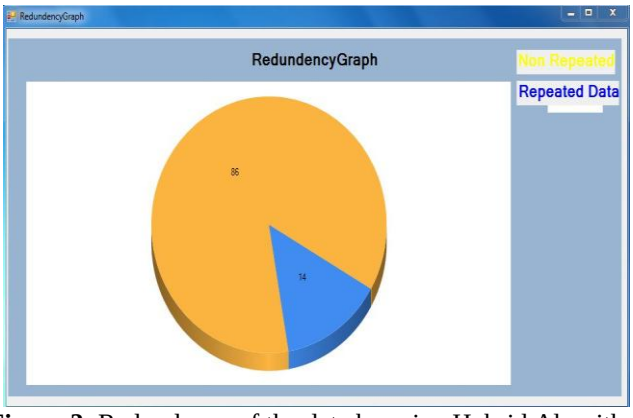


**Figure 3:** Redundancy of the data by using Hybrid Algorithm

Figure 4 shows the recognition accuracy by using hybrid algorithm is more than total PLS. It indicates improvement in the result. As shown in Figure 5, the results from TotalPLS and Hybrid algorithm shown in terms of the standard deviation which indicates the difference in the performance of both method. Hybrid algorithm can pick up the prediction accuracy and is more constant.

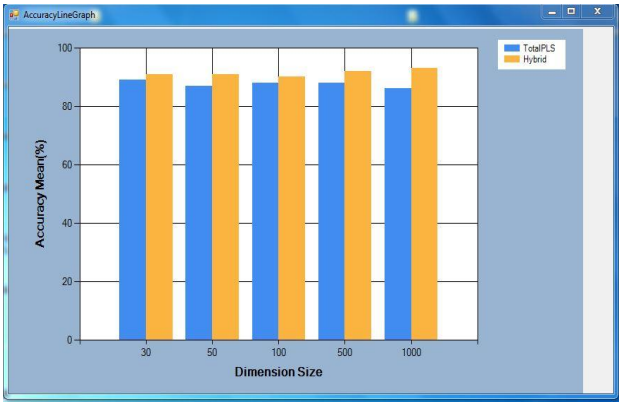


**Figure 4:** Improved Recognition rate mean accuracy by Hybrid algorithm

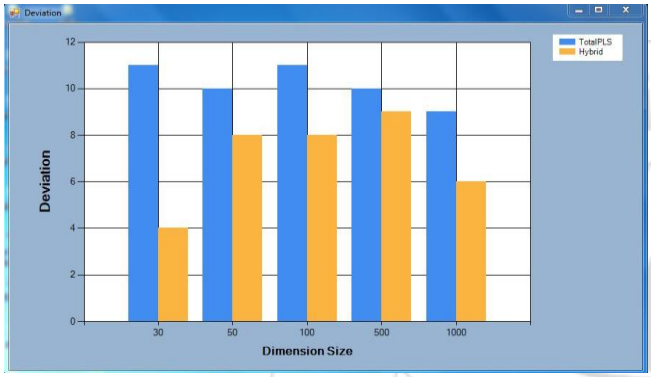


**Figure 5:** Perfomance analysis of Standard deviation based on Total PLS and Hybrid algorithm

The relationship between the number of chosen feature and recognition accuracy on test sets are recognized in Figure 6. The recognition rate from Hybrid method is much better than the TotalPLS methods.

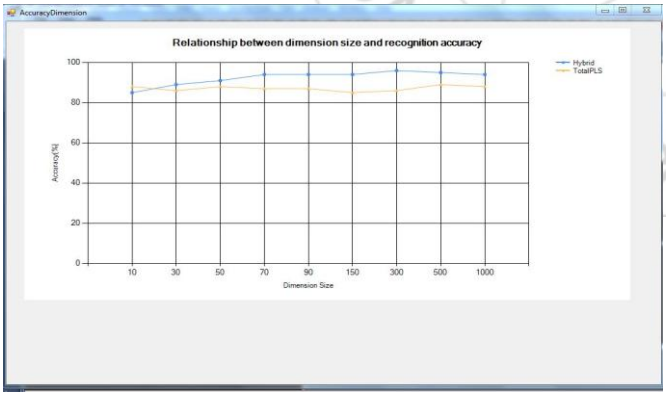


**Figure 6:** Performance analysis of both algorithm which shows relationship between dimension size and recognition accuracy

Figure 7 shows the efficiency of the Hybrid algorithm is more than the Total PLS algorithm that means the computing speed of the hybrid algorithm is good.

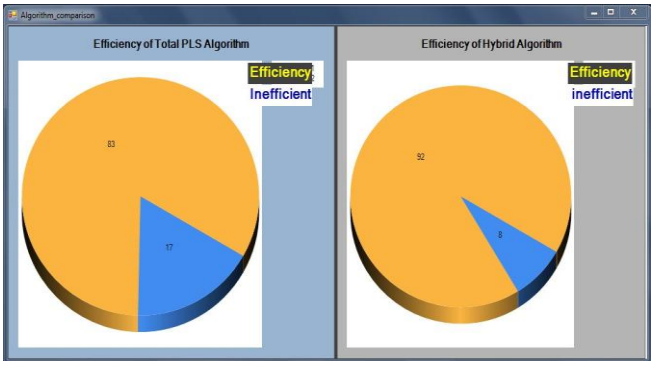


**Figure 7:** Comparision in Efficiency of the Total PLS and Hybrid algorithm

## 7. Conclusion

In this paper, we studied and implement dimension reduction algorithm for enhancing performance of the data dimension reduction. In proposed method used Total PLS and MINE algorithm which are combined as in the form of Hybrid algorithm. Using proposed method, we can minimize the time, redundancy of data analysis. We need an important goal of data mining having reasonable and effective access to useful knowledge. The aim of the study was to enhance performs data dimension reduction for the microarray data analysis. This paper introduced the two algorithms. The proposed algorithm that is Hybrid algorithm gives the improved performance in terms of efficiency, recognition accuracy, redundancy of the data as compared to the previous algorithm.

## References

[1] Wenjie You, Z Yang, M Yuan, and Guoli Ji "Total PLS: Local Dimension Reduction for Multicategory Microarray Data," IEEE Trans. on human-machine systems, vol. 44, no. 1, february 2014.
[2] D. Araujo and A .D. Neto and A. Martins and J. Melo, "Comparative Study on Dimension Reduction Techniques for Cluster Analysis of Microarray Data," Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, July 31 – August 5, 2011.
[3] I. Gheyas and L. Smith, "Feature subset selection in large dimensionality domains," Pattern Recognit., vol. 43, no. 1, pp. 5–13, 2010.
[4] J. Hua, W. D. Tembe, and E. R. Doughertya, "Performance of feature selection methods in the classification of high-dimension data," Pattern Recognit., vol. 42, no. 3, pp. 409– 424, 2009.
[5] A. Anaissi, P.J. Kenned, M. Goyal "A Framework for High Dimensional Data Reduction in the Microarray Domain," 9781-42446439-5/10/$26.00 ©2010 IEEE.
[6] A. K. Jain, R. P.W. Duin, and J. Mao, "Statistical pattern recognition: A review," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 1, pp. 4–37, Jan. 2000.
[7] J. Fan and Y. Fan, "High-dimensional classification using features annealed independence rules," Ann. Statist., vol. 36, no. 6, pp. 2605– 2637, 2008.
[8] W. H. Yang, D. Q. Dai, and H. Yan, "Feature extraction and uncorrelated discriminant analysis for high-

dimensional data," IEEE Trans. Knowl. Data Eng., vol. 20, no. 5, pp. 601–614, May 2008.

[9] J. J. Dai, L. Lieu, and D. Rocke, "Dimension reduction for classification with gene expression microarray data," Statist. Appl. Genet. Mol. Biol., vol. 5, no. 1, pp. 1–19, 2006

[10] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," Mach. Learn., vol. 46, no. 1–3, pp. 389–422, 2002.

[11] J. Yan, B. Zhang, N. Liu, S. Yan, Q. Cheng, W. Fan, Q.Yang, W. Xi, and Z. Chen, "Effective and efficient dimensionality reduction for largescale and streaming data preprocessing," IEEE Trans. Knowl. Data Eng., vol. 18, no. 3, pp. 320–333, Mar. 2006.

[12] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," IEEE Trans. Neural New., vol. 13, no. 2, pp. 415–425, Mar. 2002

[13] S.Deegalla, H. Bostrom, Fusion of Dimensionality Reduction Methods: A Case Study in Microarray Classification" 12th International Conference on Information Fusion Seattle, WA, USA, July 6-9, 2009.

[14] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G.McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, "Detecting novel associations in large data sets," Science, vol. 334, pp. 1518–1524, 2011.

[15] Yu Wanga, Igor V. Tetkoa, Mark A. Hallb, Eibe Frankb, "Gene Selection from microarray data for cancer classification-a machine learning approach", in Proc. Computational Biology and Chemistry, 29 (2005) 37–46.

[16] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science, vol. 290, no. 5500, pp.2323–2326, December 2000.

[17] I. Fodor, "A Survey of Dimension Reduction Techniques" Lawrence Livermore National Lab., CA (US) UCRL- ID-148494, 2002.