Volume 3, Issue 11, November 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Predicting the Stock Trend by Artificial Intelligence Techniques based on Structural Risk Minimization

Abeer Badr El Din*
Department of Computer Science
Sadat Academy, Egypt

Basant Ali Sayed
MSc Student, Dept. of Computer Science
Sadat Academy, Egypt

Abstract: This paper presents an approach for minimizing the risk of prediction stock trend based on structural risk minimization (SRM) for artificial intelligence techniques: multi-layer precptron, gene expression programming, the new technique support vector machines introduced by Vapnik and Chervonenkies (VC). The proposed system presents the prediction the stock trend by learning process through classification. The paper shows a comparison among multi-layer precptron, gene expression programming, and support vector machines. The results indicate to the superior performance of the support vector machines (based on mathematical background) compared by multi-layer precptron, gene expression programming.

Key Words: Prediction, Multi-layer Preceptron, Support Vector Machines, Gene Expression Programming. Structural risk minimization

1- INTRODUCTION

It is known that the prediction process for the stock trend is an attractive issue for and more extensively worked by researchers from various sciences in particular the computer science, machine learning field. The main objective for many researchers is to build a well-established algorithm that covers more areas of applications for prediction of financial markets. There are many common algorithms that tackle this problem for instance; the neural network that has a widely used for many researchers is to build a well-established algorithm. But recently, there is a new trend for using a new approach which is called support vector machine (SVM) introduced by Vapnik and Chervonenkies (VC) 1990. This approach is based on strong mathematical background using kernel trick to transfer the sample from lower dimensions space to higher dimensions space for providing a simple way to classification. The researchers use support vector machine (SVM) to build an efficient learning algorithm that tackles the prediction process for the stock trend. The core idea of this system to achieve two conflicted objectives: maximizing the achieved profit result from stock option purchase, the second objective is realizing minimum level of risk.

No doubt that there are many of these literatures that based on the features selected for the given inputs for the considered machine learning. These algorithms are commonly achieved from the data that has the same market under concern. for instance the isolation process leaves out fundamental information transferred by other entities and realize the results of the prediction process is more vulnerable to local perturbations. But on the other hand there is a trend for breaking the boundaries by involving the external information by the modernist financial news or through the personal internet posts like Twitter. The trend of these approaches is defined as sentiment analysis. This trend is based on the attitudes of different key figures or an efficient analysis for the markets to explore the minds of general investor's .but there are some challenges for this trend. The most important of them is failing when some of the investors are biased, or have positive opinions that obey past good performance.

2- ADAPTIVE ACTIVATION FUNCTIONS

There is a method for improving the performance of the neural network based on enabling the functions of the activation to modify related to the training data characteristics obtained. Using the adaptive activations functions is considered one of the earlier methods which is developed by Zurada. In this case the slope of the given sigmoid activation function would be learned together accompanied the achieved weights. There is a slope parameter λ is obtained for every output and hidden unit. Additional development is achieved by Engelbrecht et al for the lambda-learning algorithm. This sigmoid function is defined as

$$f(net, \lambda, \gamma) = \frac{\gamma}{1 + e^{-\lambda net}}$$

Such λ is defined as the slope of the function and γ is considered the maximum range. There is an additional development is achieved by Engelbrecht et al for learning equations to can learn the maximum ranges of the used sigmoid functions, thereby doing automatic scaling process which is based on gamma-learning. This development has no need to scale target values to the range (0, 1). The result of modifying the slope and range of the sigmoid function which is called delta learning, the lambda and variations of gamma learning.

3- GENE EXPRESSION PROGRAMMING

Cândida Ferreira (Ferreira 1996) invented a new approach which is called Gene expression programming which introduce the solution to the expression –mutation problem, Ferreira present a new system for encoding the expressions that enable quick implementation for a wide range of mutation and cross-breeding methods that ensuring that the obtained expression is considered syntactically valid. The given approach is named the Gene Expression Programming (GEP). Experiments have proved that GEP is one hundred to sixty thousands times quicker than classical genetic algorithms. Gene Expression Programming is considered a method that imitate the biological evolution to make a computer system to model specific phenomenon. Gene expression programming usually used to create many various kinds of models containing techniques as decision trees, neural networks.

4- SUPPORT VECTOR MACHINES

The new trend in the artificial intelligence which is called the Support vector machine. Support vector machine can be defined as a computer algorithm that performs learning process by instance to allocate the labels to objects. For example, an SVM has the ability of learning to realize fraudulent as credit card problem by exploring more hundreds or thousands of samples has fraudulent and no fraudulent credit card activity reports. In the same the SVM can perform the learn process to optimizing the recognizing handwritten digits by exploring a wide range of collection for the scanned images of handwritten. SVMs has proved the successfully implementation to a develop the wide range of biological implantations.

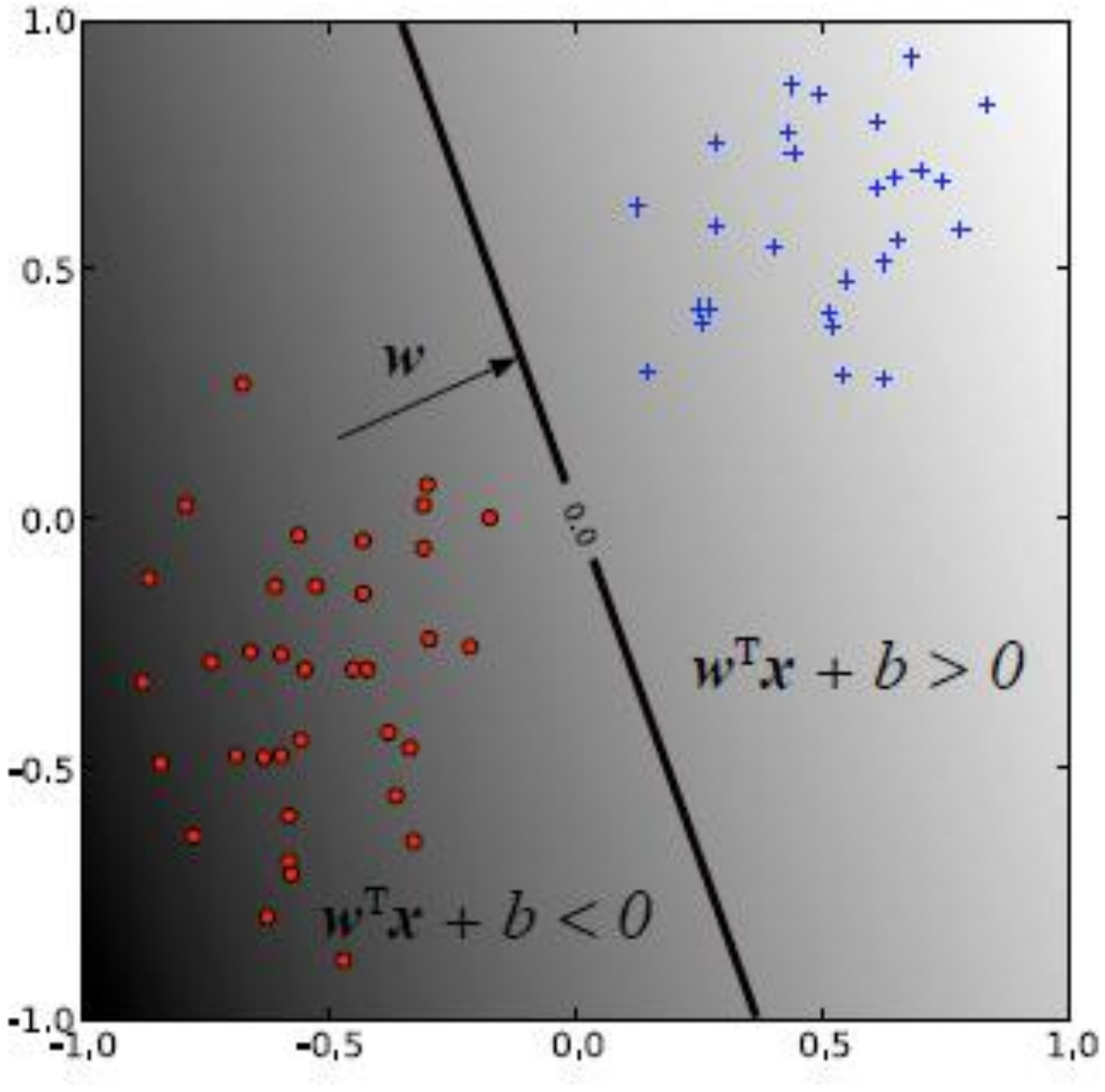
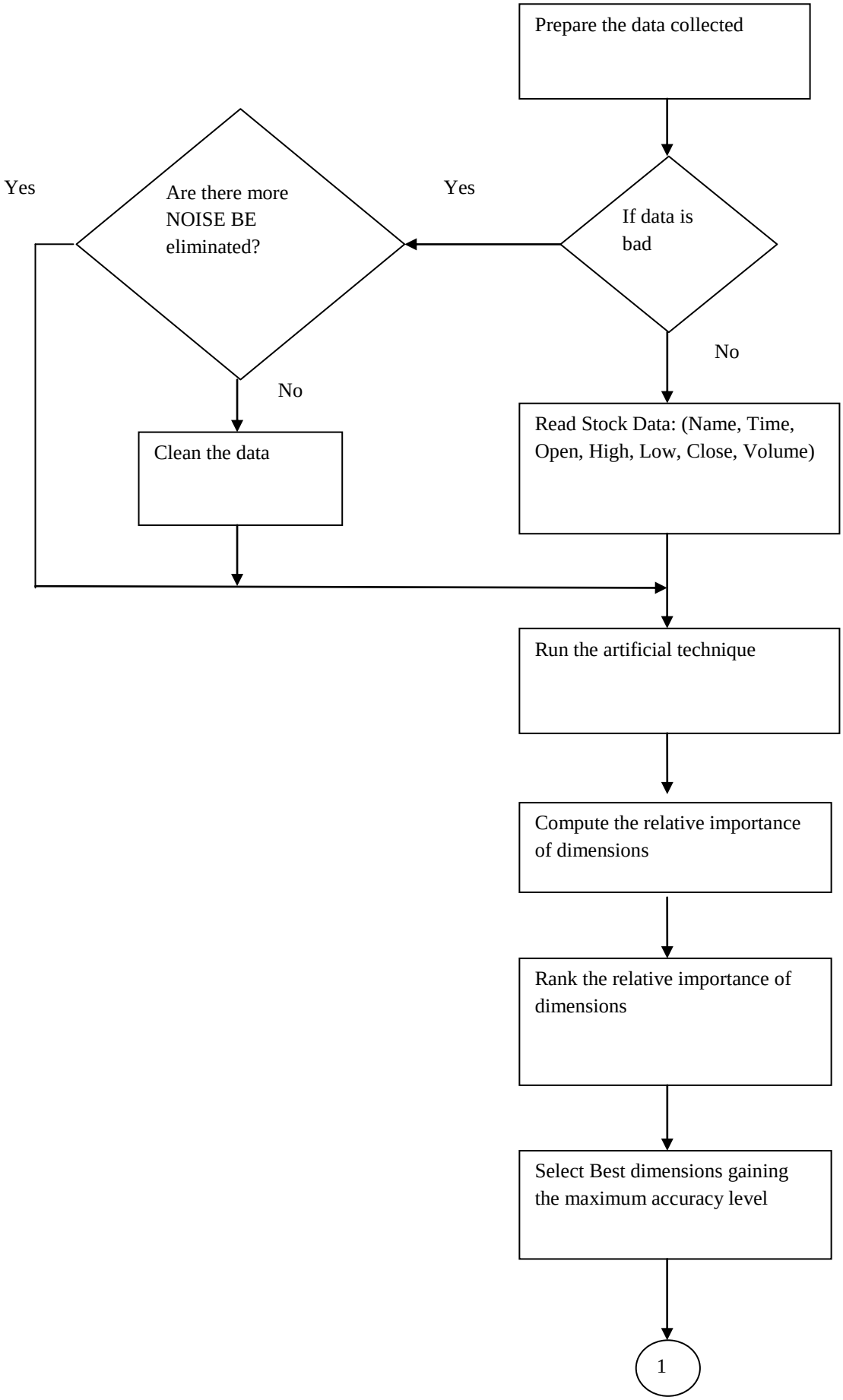


Fig 1: A linear classifier. The decision boundary (points x such that $w^T x+b=0$) divides the plane into two sets depending on the sign of $w^T x+b$.

5- PROPOSED ALGORITHM

Signal Module Proposed System Flow Chart



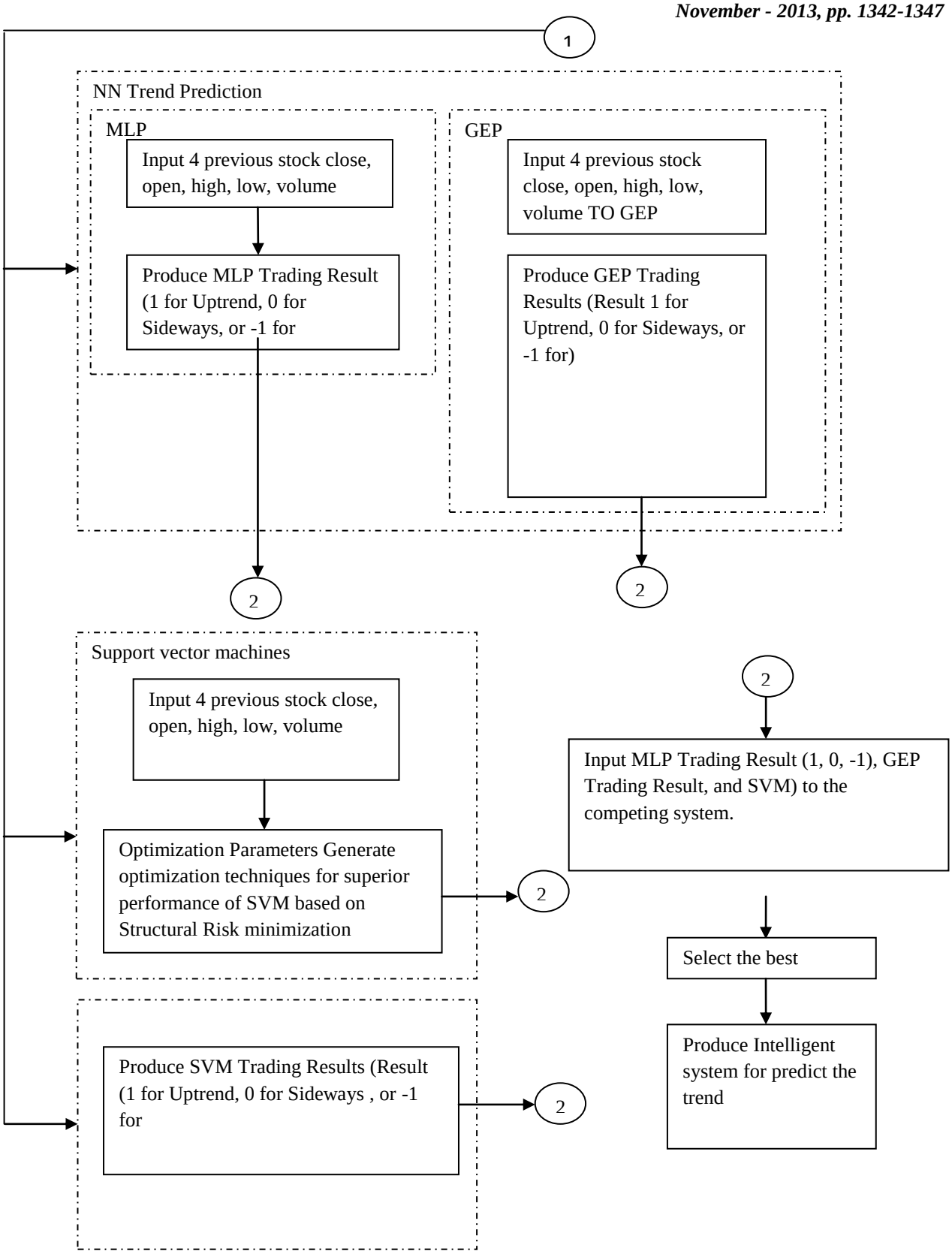


Fig 2: Signal Module Proposed System Flow Chart

6- EXPERIMENTS & RESULTS

This section presents two sections: the first section shows how to optimize the parameters of the artificial intelligence techniques.

6.1-SIMULATION PROCEDURE

The simulation is based on from each data set, 50% training and 30% validating 20% testing were produced: 20 times repeated 10-fold cross validation are used.

6-2. PERFORMANCE MEASURES

to compare paper results with the best known measure in the literature, the classification error is used as performance measure.

6.3- OPTIMIZATION THE PARAMETERS

This section presents the mechanism of optimization parameters of the artificial intelligence (Multi-Layer Preceptron) as example. The experiment is performed using 4-fold cross-validation.

TABLE 1: MODEL SIZE SUMMARY REPORT

Hidden layer 1 neurons	% Misclassifications
2	37.09273
3	32.20551
4	32.83208
5	32.08020
6	32.08020
7	32.58145
8	33.33333
9	31.95489
10	34.58647
11	33.83459
12	32.08020
13	31.82957 <-- Optimal size
14	32.08020
15	33.45865
16	32.95739
17	32.95739
18	33.45865
19	33.45865
20	33.70927

From the previous table it is clear that the network will be built using 13 neurons for hidden layer 1. The proposed structure has minimum classification error. The main idea of improving the accuracy of classification is to avoid the overfitting problem by considering the principal of Vapnik –Chervonenkis which is called **S**tructural **R**isk **M**inimization.

Fig 3 : model size and error rate chart

The optimal structure of gene expression as the following:
Generations required to train model = 1,488,
Complexity of model before simplification = 86,
Complexity of model after simplification = 33 ,
Generations required for simplification = 113
Number of evaluations of the fitness function = 278,60 ,
and Number of execution threads used = 4

The optimal structure of SVM grid and pattern searches found optimal values for parameters: Search criterion: Minimize total error , Number of points evaluated during search = 136 , Minimum error found by search = 0.308657, Parameter values: Epsilon = 0.001, C = 6359.04005 , Gamma = 4.51600535 , and Number of support vectors used by the model = 510

6.4- COMPARATIVE RESULTS

This section introduces comparative experiments performed on three artificial intelligence techniques

Training

TABLE 2: COMPARATIVE RESULTS FOR TRAINING

	MLP			SVM			GEP		
	Down	Stable	Up	Down	Stable	Up	down	stable	up
dimension 5	69.76%	96.24%	71.52%	76.66%	96.24%	78.17%	62.48%	96.36%	65.87%
dimension 4	70.26%	96.24%	72.02%	73.53%	96.24%	75.03%	74.78%	96.24%	76.04%
dimension 3	72.27%	96.24%	74.28%	74.65%	96.24%	76.16%	74.78%	96.24%	76.04%
dimension 2	72.27%	96.24%	74.28%	73.90%	96.24%	75.16%	52.82%	96.36%	56.46%

Table 2 show that Comparative results for training group based on the accuracy level for dimension 5 ,4, 3,and 2 be achieved by MLP,SVM, and gene expression programing .the table display the three categories Down, Stable, and UP for the training process.

Validation

TABLE 3: COMPARATIVE RESULTS FOR VALIDATION

	MLP			SVM			GEP		
	Down	Stable	Up	Down	stable	Up	down	Stable	up
dimension 5	70.43%	96.24%	72.18%	70.43%	96.24%	71.93%	64.62%	95.73%	67.63%
dimension 4	70.30%	96.24%	72.06%	71.93%	96.24%	73.18%	74.40%	96.24%	75.66%
dimension 3	71.18%	96.24%	72.93%	73.31%	96.24%	74.56%	72.65%	96.24%	74.15%
dimension 2	72.68%	96.24%	74.19%	73.18%	96.24%	74.44%	52.70%	96.24%	55.71%

Table 3 shows Comparative results for validation group based on the accuracy level for dimension 5 ,4, 3,and 2 be achieved by MLP,SVM, and gene expression programing .the table display the three categories Down, Stable, and UP for the test process.

7- CONCLUSION

The paper presents results obtained by three compared techniques :Multi-layer perceptron ,Support vector machine, and Gene expression programming . It is found statistically that superior performance of support vector machines compared by gene expression programming, multi-layer perceptron. This superior performance related to strong mathematical foundations of support vector machine invented by Vapnik 1990.This led to good classification performance by support vector machines.

REFERENCE

[1] B. Scholkopf, K. Tsuda, and J.P. Vert, editors. *Kernel Methods in Computational Biology. MIT Press series on Computational Molecular Biology*. MIT Press, 2004.
[2] W.S. Noble, "What is a support vector machine?" Nature Biotechnology, 24(12):1564{1567, 2006.
[3] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge UP, Cambridge, UK, 2004.
[4] B. Sch□olkopf and A. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002. .
[5] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin," A practical guide to support vector classi_cation ," Technical report, Department of Computer Science, National Taiwan University, 2003.
[6] S. Sonnenburg, M.L. Braun, C.S. Ong, L. Bengio, S. Bottou, G. Holmes, Y. LeCun, K M uller, F Pereira, C.E. Rasmussen, G. Ratsch, B. Sch olkopf, A. Smola, P. Vincent, J. Weston, and R.C. Williamson, "The need for open source software in machine learning", Journal of Machine Learning Research, 8:2443{2466, 2007.
[7] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge UP, Cambridge, UK, 2000.
[8] T. Hastie, R. Tibshirani, and J.H. Friedman , "The Elements of Statistical Learning", Springer, 2001.
[9] C.M. Bishop,"Pattern Recognition and Machine Learning", Springer, 2007.
[10] C. Cortes and V. Vapnik, *Support vector networks*, Machine Learning, 20:273{297, 1995.
[11] O. Chapelle. Training a support vector machine in the primal. In L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, editors, Large Scale Kernel Machines. MIT Press, Cambridge, MA., 2007.
[12] F. Provost," Learning with imbalanced data sets 101 ," In AAAI 2000 workshop on imbalanced data sets, 2000.

