Assessing applicative performance of a hybrid machine learning model

Sung Ho Ha*, Jong Sik Jin, and Seong Hyeon Joo

***Abstract*—**Yield management in semiconductor manufacturing companies requires accurate yield prediction and continual control. However, because many factors are complexly involved in the production of semiconductors, manufacturers or engineers have a hard time managing yield precisely. Intelligent tools need to analyze multiple process variables concerned and to predict production yield effectively. This paper devises a hybrid method of incorporating machine learning techniques together to detect high and low yields in semiconductor manufacturing. The hybrid method has strong applicative advantages in these manufacturing situations, whereby control of a variety of process variables is interrelated. In the real applications, the hybrid method provides more accurate yield prediction than other methods that have been used. With this method, the company achieves a higher yield rate by preventing low-yield lots in advance.

***Keywords*—**Hybrid application, machine learning, case-based reasoning, feature weighting.

I. Introduction

IN the manufacturing of semiconductors, final products are fabricated through several hundreds of processes which are highly automated and dramatically interdependent. Most manufacturing processes in use today are complexly intertwined and become infinitesimal when using nanometer-scale technology.

For those manufacturers or engineers, yield is considered as a very important factor that has to be monitored and controlled. Yield is defined as the ratio of normal products to finished products. Yield management in the semiconductor industry is understood as a comprehensive analytical system which has the characteristic of complex systems. A complex system has many independent component variables that interact with each other in many complicated ways. Therefore it is considered to be difficult to predict and control.

Yield in semiconductor fabrication is strongly affected by several factors, including particles or contaminants on the wafer, substances in the manufacturing instruments, manufacturing process parameters, process engineers' attitudes, and the design of semiconductors.

Semiconductor companies can achieve a certain degree of yield by applying statistical process control and 6-sigma to a semiconductor. Yield enhancement employing statistical measurements, however, has difficulty in preventing low-yield lots effectively in advance. This is because manufacturing process variables which affect changes in the yield have a non-linear complex relationship with the yield. Due to this interactive effect among several variables, manufacturers can hardly pinpoint problems in time, when small changes in a relationship between process parameters can cause changes in the yield.

Other intelligent techniques are, thus, needed in order to detect the main process variables which seriously affect changes in the yield. We have developed a hybrid prediction system in the semiconductor industry as a complement to the existing statistical approach. This system is based on a hybrid application of machine learning techniques to effectively depict multiple process variables concerned with predicting the production yield in semiconductor manufacturing.

The hybrid prediction system adopts neural networks (NNs) and case-based reasoning (CBR) which can be directly applied to prediction purposes. However, CBR suffers from feature weighting; when it measures the distance between cases, some features should be weighted differently. Many feature-weighted variants of the *k*-Nearest Neighbor (*k*-NN) have been proposed to assign higher weights to more relevant features for case retrieval purposes [1][2]. Though those variants have been reported as improving their retrieval accuracy regarding some tasks, few have been used in conjunction with neural networks to predict yield performance in semiconductor manufacturing.

In order to weigh features, the hybrid system adopts four feature-weighting methods: Sensitivity, Activity, Saliency, and Relevance. Each method calculates the degree of each feature's importance by using the connection weights and activation patterns of the nodes in the trained neural network.

In order to validate this hybrid approach, the developed system was applied to an international semiconductor company, which has been ranked one of the top manufacturers in the world. After comparing this hybrid method with other methods that have been used, this paper shows the hybrid method provides more accurate yield prediction.

This paper is organized as follows: Section 2 reviews various approaches in providing yield management applied to semiconductor manufacturing. This section focuses on hybrid

Manuscript received May 7, 2007; Revised received February 27, 2008.
S. Ha is with Kyungpook National University, Daegu, 702-701, Korea. (phone: +82-53-950-5440; fax: +82-53-950-6247; e-mail: hsh@knu.ac.kr.

J. Jin is with Kyungpook National University, Daegu, 702-701, Korea. (e-mail: jjs@knu.ac.kr).

S. Joo is with Kyungpook National University, Daegu, 702-701, Korea. (e-mail: sideas@knu.ac.kr).

* Corresponding author

applications combining machine learning techniques. Section 3 describes the methodology of the hybrid yield prediction system in the semiconductor industry. Experimental results are presented in Section 4 to validate the system. Finally, this paper is concluded by briefly summarizing the study and the direction of future research.

II. Literature review

Liao [3] surveyed expert systems development literature from 1995 to 2004. Based on his findings, the major applications implementing hybrid CBR have happened in the following areas: manufacturing design, fault diagnosis, knowledge modeling and management, and medical planning and application.

The hybrid CBR approach has been extensively adopted in manufacturing design and fault diagnosis. Hui and Jha [4] integrated NN, CBR, and rule-based reasoning to support customer service activities, such as decision support and machine fault diagnosis in a manufacturing environment. Liao [5] integrated a CBR method with a multi-layer perceptron for the automatic identification of failure mechanisms in the entire failure analysis process. Yang, Han, and Kim [6] integrated CBR with an ART-Kohonen NN to enhance fault diagnosis of electric motors. Tan, Lim, Platts, and Koay [7] integrated CBR and the fuzzy ARTMAP NN to support managers in making timely and optimal manufacturing technology investment decisions. Saridakis and Dentsoras [8] introduced a case-based design with a soft computing system to evaluate the parametric design of an oscillating conveyor.

The following research works in the knowledge modeling and management areas have been developed. Hui, Fong, and Jha [9] combined the CBR and NN approach to extract knowledge from the previous customer services and recall the appropriate service. Choy, Lee, and Lo [10] developed an intelligent supplier relationship management system using hybrid CBR and NN techniques to select and benchmark potential suppliers of Honeywell Consumer Products Limited in Hong Kong. Yu and Liu [11] proposed a hybridization of both symbolic and numeric reasoning techniques to achieve a higher accuracy and overcome the data scarcity problem in the construction project database. Chen and Hsu [12] solved potential lawsuit problems caused by change orders in construction projects. They utilized NNs to predict litigation likelihood, and utilized CBR to warn yields. Im and Park [13] developed a hybrid expert system of CBR and NN for a personalized counseling system for the cosmetic industry.

Hybrid CBR has also been used in the medical planning and application areas. Guiu, Ribé, Mansilla, and Fàbrega [14] introduced a case-based classifier system to solve the automatic diagnosis of Mammary Biopsy Images. Hsu and Ho [15] combined the CBR, NN, fuzzy theory, and induction theory together to facilitate multiple-disease diagnosis and the learning of new adaptation knowledge. Wyns, Boullart, Sette, Baeten, Hoffman, and Keyser [16] applied a modified Kohonen mapping combined with a CBR evaluation criterion to predict early arthritis, including rheumatoid arthritis and spondyloarthropathy. Ahn and Kim [17] combined the CBR with genetic algorithms to evaluate cytological features derived from a digital scan of breast fine needle aspirate (FNA) slides.

As well, hybrid CBRs have been used in the financial forecasting areas. Kim and Han [18] presented a case-indexing method of CBR which utilizes SOM for the prediction of corporate bond rating. Li, Sun, and Sun [19] introduced a feature-based similarity measure to deal with financial distress prediction (e.g., bankruptcy prediction) in China. Chang and Lai [20] integrated the SOM and CBR for sales forecasts of newly released books. Chang, Lai, and Lai [21] evolved a CBR system with genetic algorithm for wholesaler returning book forecasting. Chun and Park [22] devised a regression CBR for financial forecasting, which applies different weights to independent variables before finding similar cases. Kumar and Ravi [23] presented a comprehensive review of the works utilizing NN and CBR to solve the bankruptcy prediction problems faced by banks.

III. Hybrid case-based reasoning system

In order to improve the ability of predicting yield accurately, a hybrid prediction system was developed for the semiconductor industry. It is a hybrid method combining machine learning techniques, such as back-propagation network (BPN), CBR, and *k*-NN (see Fig. 1).

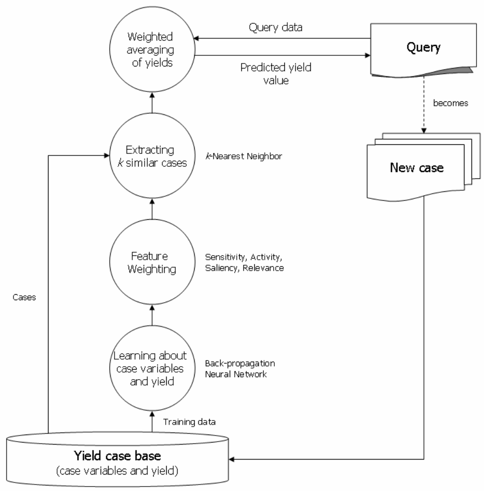

Fig. 1 The architecture of neural feature-weighted case retrieval

The hybrid system consists of four phases: Learning about the relationship between case variables and yield, feature weighting, extracting *k* similar cases, and weighted averaging of extracted yields. The first phase finds the relative importance of independent variables from the relationship between independent variables (i.e., manufacturing process variables) and a dependent variable (i.e., yield). When the training of a BPN is finished in the instance of the Yield case base, the connection weights of a trained neural network reveal the importance of the relationship between the process variables and yield.

To obtain a set of feature weights from the trained network, four feature-weighting methods are utilized: Sensitivity, Activity, Saliency, and Relevance [2][24][25]. Each of these methods calculates the degree of each feature's importance by using the connection weights and activation patterns of the nodes in the trained neural network. The feature-weighting algorithms are briefly described as follows:

- 'Sensitivity' weighting method

An input node *i*'s sensitivity (Sen_i) is calculated by removing the input node from the trained neural network. The sensitivity of an input node is the difference in error between the removal of the feature and when it is left in place. Sen_i is calculated by the following equation:

$$Sen_i = \max\left(\frac{E(0) - E(w^f)}{E(w^f)},\ 0\right) \tag{1}$$

where $E(0)$ indicates the amount of error after removing an input node *i* and $E(w^f)$ means the error value when the node is left untouched. The error value is based on the following equation:

$$E = \sum_{CB} \left| y - o_y \right| \tag{2}$$

where *CB* is a case base which contains case variables (features) and corresponding yield and y indicates the actual yield value and o_y indicates the yield value observed by the BPN.

- 'Activity' weighting method

Node activity is measured by the variance of the level of activation for the training data. When the activity value of a node varies significantly according to its input value, the activity of the node is high.

Activity of an input node i (Act_i) is calculated by the following equations:

$$Act_i = \sum_{j=1}^{H} w_{ji}^2 Act_j^h \tag{3}$$

$$Act_j^h = \sum_{k=1}^{O} w_{kj}^2 \cdot \text{var}\left(s\left(\sum_{i=1}^{N} w_{ji} x_i \right)\right) \tag{4}$$

where x_i represents input nodes ($i = 1, \dots, N$), Act_j^h is the activity of a hidden node *j* ($j = 1, \dots, H$); *k* signifies output nodes ($k = 1, \dots, O$); w_{kj} represents a connection weight from a hidden node *j* to an output node *k*, w_{ji} is a weight connected from an input node *i* to a hidden node *j*, $\text{var}()$ is the variance function, and $s()$ is the activation function.

- 'Saliency' weighting method

Saliency is measured by estimating the second derivative of the error with respect to weight. Saliency is used to prune neural networks iteratively: that is, to train to reasonable error levels, compute saliencies, delete low saliency weights, and resume training.

Saliency of an input node i (Sal_i) is calculated by the following function:

$$Sal_i = \sum_{j=1}^{H} \sum_{k=1}^{O} w_{kj}^2 w_{ji}^2 \tag{5}$$

- 'Relevance' weighting method

The variance of weight in a node is a good predictor of the node's Relevance. This relevance is a good predictor of the expected increase in error when the node's largest weight is deleted.

Relevance of an input node i (Rel_i) is calculated by the following equations:

$$Rel_i = w_{ji} \cdot Rel_j^h \tag{6}$$

$$Rel_j^h = \sum_{k=1}^{O} w_{kj}^2 \cdot \text{var}\left(w_{ji}\right) \tag{7}$$

where Rel_j^h is the relevance of a hidden node *j*.

After weighting features based on the trained neural network and four weighting methods, a *k*-NN algorithm finds the most similar *k* cases from the case base. When a new query comes in, the normalized Euclidean distance, $\Delta(\mathbf{q}, \mathbf{x})$ between the query and the case is calculated as follows:

$$\Delta(\mathbf{q}, \mathbf{x}) = \sqrt{\sum_{i=1}^{N} w_i d\left(q_i, x_i\right)^2} \tag{8}$$

$$d\left(q_i, x_i\right) = \begin{cases} \left|q_i - x_i\right| & \text{if attribute } i \text{ is numeric;} \\ 0 & \text{if attribute } i \text{ is symbolic and } q_i = x_i; \\ 1 & \text{otherwise.} \end{cases} \tag{9}$$

where $\mathbf{q}$ is the query and $\mathbf{x}$ is a case which is stored in the case base, q_i and x_i are the i^{th} input feature values of $\mathbf{q}$ and $\mathbf{x}$. In this case, w_i is one of the Sensitivity, Activity, Saliency, and Relevance weights, which is assigned to the i^{th} feature. $d\left(q_i, x_i\right)$ is the difference between the two values q_i and x_i.

Finally, in order to calculate the predicted value of yield, the hybrid prediction system calculates the weighted average of yields from these k similar cases. At this time, the normalized distances to the query are used as the weights. The predicted value of yield is calculated as follows:

$$\text{Predicted yield} = \frac{\sum_{l=1}^{k} \Delta(\mathbf{q},\ \mathbf{x})_l^{-1} y_l}{k} \quad (10)$$

where y_l is the l^{th} production yield which is discovered from the feature-weighted case retrieval.

IV. Application and evaluation

In order to verify the effectiveness of the hybrid method devised in this paper, it was applied to the production data collected from the manufacturing lots of a Korean semiconductor company: 230 high-yield lots, 230 low-yield lots, and 16 process variables. By definition, a high-yield lot delivers more than 90% yield from a lot and a low-yield lot conveys less than 60% yield from a lot. This definition is still in use by the engineers of the company. Sixteen process variables were allowed to be collected due to the data security constraint enforced by the company. Among the real lot data collected, 276 lots (60%) were utilized as training data and 184 lots (40%) were utilized as testing data.

A BPN was constructed in order to learn the relationship between case variables and yield. Akaike Information Criterion (AIC) is used to determine the optimal topology of the neural network. AIC, which is as famous as Schwartz Bayesian Criterion, picks up the optimal number of hidden nodes through a heuristic search [26]. When the number of input nodes is set to 16, the number of hidden layers is one, and the number of output nodes is one, AIC determined 25 hidden nodes as an optimal structure of the BPN.

To evaluate the predictive performance of CBR with the Sensitivity weighting scheme (BPN+CBR_Sen), the BPN had 16 input nodes, 25 hidden nodes, and one output node in its structure. 276 training data flowed through this BPN in order. When training was done, the Sensitivity weights of the 16 input nodes (i.e., process variables of manufacturing) were determined. Using these weights, a k-NN algorithm acquired the k nearest cases from a set of cases in the Yield case base. We reiterated the same experiment ten times and calculated the variance of predictive performance of the BPN+CBR_Sen.

Following the experimental procedure done for the BPN+CBR_Sen, the CBR with the Activity weighting scheme (BPN+CBR_Act), the CBR with the Relevance weighting scheme (BPN+CBR_Rel), and the CBR with the Saliency weighting scheme (BPN+CBR_Sal) were constructed, trained, and evaluated against the testing data. Also, these four hybrid methods were compared with a CBR-alone method in order to show performance comparison.

Table 1 shows the performances (e.g., averages and standard deviations) of all feature-weighting methods, according to varying k. The second to fifth columns show mean errors and their standard deviations of the feature-weighting CBRs, computed from experiments conducted ten times for each k. The ‘CBR-alone’ column shows CBR errors without feature weighting, that is, pure k-NN algorithm. Since we performed one experiment with the CBR-alone method for each k, the CBR-alone column does not include standard deviations.

According to the table, the BPN+CBR_Sen shows the lowest error rate when k is set to five; the BPN+CBR_Act has the lowest error rate when k is set to 11; the BPN+CBR_Sal shows the lowest error rate when k is set to five; and the BPN+CBR_Rel shows the lowest error rate when k is set to nine. Beyond those points of k in each weighting method, the error rates slightly increased.

Fig. 2 illustrates the average prediction accuracy of all feature-weighting methods, according to varying k.

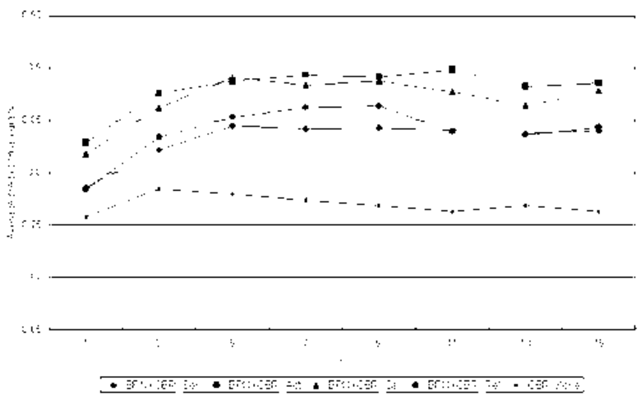

Fig. 2 Average prediction accuracy of each weighting scheme

All the four weighting methods excelled the CBR-alone method in every experiment. Furthermore, in most experiments, the BPN+CBR_Act showed the highest prediction accuracy, followed by BPN+CBR_Sal, BPN+CBR_Rel, and BPN+CBR_Sen.

As k increases to 11, the difference in the prediction accuracy gets bigger between the CBR-alone and BPN+CBR_Act weighting methods. There are, however, small differences in the prediction accuracy among the four feature-weighting methods.

In general, it is difficult to decide which weighting method is the best. We suggest that one should test the four methods at the initial development phase and then implement the one with the lowest prediction error in the production phase. In this case, adopting the BPN+CBR_Act weighting method is an acceptable solution to predict the yield rate in semiconductor manufacturing.

TABLE 1. STATISTICS OF CASE-BASED REASONING WITH FOUR WEIGHTING SCHEMES

k	BPN+CBR_Sen	BPN+CBR_Act	BPN+CBR_Sal	BPN+CBR_Rel	CBR-alone
1	0.2142±0.0253	0.1712±0.0267	0.1822±0.0322	0.2162±0.0301	0.2424
3	0.1781±0.0233	0.1241±0.0214	0.1381±0.0287	0.1661±0.0277	0.2152
5	0.1551±0.0232	0.1123±0.0222	0.1094±0.0293	0.1467±0.0275	0.2207
7	0.1581±0.0242	0.1063±0.0287	0.1162±0.0218	0.1378±0.0232	0.2261
9	0.1571±0.0231	0.1081±0.0239	0.1121±0.0245	0.1361±0.0283	0.2315
11	0.1593±0.0256	0.1015±0.0194	0.1226±0.0299	0.1607±0.0213	0.2370
13	0.1622±0.0255	0.1174±0.0203	0.1357±0.0223	0.1639±0.0288	0.2315
15	0.1598±0.0214	0.1141±0.0233	0.1214±0.0264	0.1562±0.0238	0.2370

V. CONCLUSION

Yield management in the semiconductor industry is understood as a very important management practice that has to be monitored and controlled completely. Because manufacturing process variables have a non-linear complex relationship with the yield, manufacturers need an intelligent approach to pinpoint the relationship between process parameters in time.

In this paper, we devised and applied a hybrid method combining BPN and CBR, to predict the yield of the target semiconductor manufacturing company. In the hybrid prediction system, the BPN was used to assign relative weights to manufacturing process features of each case in the yield case base.

As the literature review in Section 2 revealed, there was no previous similar research for predicting the yield rate of the semiconductor company utilizing the neural feature-weighting CBR. The hybrid system showed that the CBR with the ‘Activity’ weighting method had a better prediction rate, outperforming the CBR-alone and all other weighting methods. The hybrid CBR also showed better performance than the existing statistical approach.

However, in order to achieve a more accurate prediction rate, the hybrid system needs more process variables and data from the target company. Even though the existing 16 variables used in this paper were determined by the manufacturing engineers, it is difficult to achieve a more accurate prediction rate by only using these variables and data. This will be the next challenging research.

REFERENCES

[1] H. Ahn, K. Kim, and I. Han, “Global optimization of feature weights and the number of neighbors that combine in a case-based reasoning system,” *Expert Systems*, vol. 23, pp. 290-301, 2006.

[2] D. Wettschereck, D. W. Aha, and T. Mohri, “A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms,” *Artificial Intelligence Review*, vol. 11, pp. 273-314, 1997.

[3] S.-H. Liao, “Expert system methodologies and applications—a decade review from 1995 to 2004,” *Expert Systems with Applications*, vol. 28, no. 1, pp. 93-103, 2005.

[4] S. C. Hui and G. Jha, “Data mining for customer service support,” *Information & Management*, vol. 38, no. 1, pp. 1-13, 2000.

[5] T. W. Liao, “An investigation of a hybrid CBR method for failure mechanisms identification,” *Engineering Applications of Artificial Intelligence*, vol. 17, no. 1, pp. 123-134, 2004.

[6] B.-S. Yang, T. Han, and Y.-S. Kim, “Integration of ART-Kohonen neural network and case-based reasoning for intelligent fault diagnosis,” *Expert Systems with Applications*, vol. 26, no. 3, pp. 387-395, 2004.

[7] K. H. Tan, C. P. Lim, K. Platts, and H. S. Koay, “An intelligent decision support system for manufacturing technology investments,” *International Journal of Production Economics*, vol. 104, no. 1, pp. 179-190, 2006.

[8] K. M. Saridakis and A. J. Dentsoras, “Case-DeSC: A system for case-based design with soft computing techniques,” *Expert Systems with Applications*, vol. 32, no. 2, pp. 641-657, 2007.

[9] S. C. Hui, A. C. M. Fong, and G. Jha, “A web-based intelligent fault diagnosis system for customer service support,” *Engineering Applications of Artificial Intelligence*, vol. 14, no. 4, pp. 537-548, 2001.

[10] K. L. Choy, W. B. Lee, and V. Lo, “Design of an intelligent supplier relationship management system: a hybrid case based neural network approach,” *Expert Systems with Applications*, vol. 24, no. 2, pp. 225-237, 2003.

[11] W.-D. Yu and Y.-C. Liu, “Hybridization of CBR and numeric soft computing techniques for mining of scarce construction databases,” *Automation in Construction*, vol. 15, no. 1, pp. 33-46, 2006.

[12] J.-H. Chen and S. C. Hsu, “Hybrid ANN-CBR model for disputed change orders in construction projects,” *Automation in Construction*, vol. 17, no. 1, pp. 56-64, 2007.

[13] K. H. Im and S. C. Park, “Case-based reasoning and neural network based expert system for personalization,” *Expert Systems with Applications*, vol. 32, no. 1, pp. 77-85, 2007.

[14] J. M. G. Guiu, E. G. Ribé, E. B. Mansilla, and X. L. Fàbrega, “Automatic diagnosis with genetic algorithms and case-based reasoning,” *Artificial Intelligence in Engineering*, vol. 13, no. 4, pp. 367-372, 1999.

[15] C.-C. Hsu and C.-S. Ho, “A new hybrid case-based architecture for medical diagnosis,” *Information Sciences*, vol. 166, no. 1-4, pp. 231-247, 2004.

[16] B. Wyns, L. Boullart, S. Sette, D. Baeten, I. Hoffman, and F. D. Keyser, “Prediction of arthritis using a modified Kohonen mapping and case based reasoning,” *Engineering Applications of Artificial Intelligence*, vol. 17, no. 2, pp. 205-211, 2004.

[17] H. Ahn and K.-J. Kim, “Global optimization of case-based reasoning for breast cytology diagnosis,” *Expert Systems with Applications*, Available online 17 Nov. 2007.

[18] K.-S. Kim and I. Han, “The cluster-indexing method for case-based reasoning using self-organizing maps and learning vector quantization for bond rating cases,” *Expert Systems with Applications*, vol. 21, no. 3, pp. 147-156, 2001.

[19] H. Li, J. Sun, and B.-L. Sun, “Financial distress prediction based on OR-CBR in the principle of k-nearest neighbors,” *Expert Systems with Applications*, Available online Oct. 2007.

[20] P.-C. Chang and C.-Y. Lai, “A hybrid system combining self-organizing maps with case-based reasoning in wholesaler’s new-release book forecasting,” *Expert Systems with Applications*, vol. 29, no. 1, pp. 183-192, 2005.

[21] P.-C. Chang, C.-Y. Lai, and K. R. Lai, “A hybrid system by evolving case-based reasoning with genetic algorithm in wholesaler’s returning book forecasting,” *Decision Support Systems*, vol. 42, no. 3, pp. 1715-1729, 2006.
[22] S.-H. Chun and Y.-J. Park, “A new hybrid data mining technique using a regression case based reasoning: Application to financial forecasting,” *Expert Systems with Applications*, vol. 31, no. 2, pp. 329-336, 2006.
[23] P. R. Kumar and V. Ravi, “Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review,” *European Journal of Operational Research*, vol. 180, no. 1, pp. 1-28, 2007
[24] H. Liu and H. Motoda, *Feature selection for Knowledge discovery and data mining*, Kluwer, Norwell, MA, 1998.
[25] L. Zhang, F. Coenen, and P. Leng, “Formalising optimal feature weight setting in case based diagnosis as linear programming problems,” Knowledge-Based Systems, vol. 15, pp. 391-398, 2002.
[26] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: A practical-theoretic approach*, 2nd ed., Springer-Verlag, 2002.