DOI : https://doi.org/10.56452/7-793

ISSN: 0974-5823 Vol. 7 No. 1 January, 2022

International Journal of Mechanical Engineering

An Accurate Fuzzy System Based on Evaluation of Hybrid Ant-Bee Algorithm

Ashok Kumar Sahoo

Asst. Professor, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Abstract

A collection of membership criteria and guidelines used in data analysis make up a fuzzy expert system. Despite significant progress in understanding cancer molecular profiles, it is still difficult to effectively use microarray technology to standard clinical diagnoses. The dependability of the training data sets utilized to create the classifiers and the performance of the classifiers, particularly once the sample that has to be categorized failed to correspond to any of the available classes, are the two fundamental constraints of current procedures in the categorization of microarray data. Medical thermography has demonstrated its value in a number of implications in medicine, such as the identification of breast cancer wherever it is capable pinpoint the localized rise in temperature caused by cancer cells' intense metabolic activity. It has been demonstrated that it works better in these situations than other modalities, such as mammography, in detecting cancers in their initial phases or in thick tissue. Instead of being symbolic reasoning engines predominate like traditional competent systems, fuzzy expert systems more heavily weighted towards numerical processing. The system suggests the hybrid Ant Bee Algorithm (ABA) also evaluates it employing six gene expression data sets to solve the interpretability-accuracy tradeoff.

1. INTRODUCTION

With the hybridization process, microarray technology measures the amounts of almost every gene expressed in a biological sample, producing a large quantity of data in the process. It is reasonable to anticipate that understanding gained from microarray data will considerably benefit both basic biological research and clinical treatment. Classifying the samples is a vital part of microarray data analysis. To name a few, studies have considered cluster analysis of tumour along with healthy colon tissues, as well as categorization of acute leukaemia. The methodologies created in these studies include machine learning and discrimination methods for a comparative investigation.

The number of samples used in microarray investigations, or n, is relatively low in comparison to the number of genes, or p, which is often thousands. Standard statistical approaches for classification function poorly without performing a prior variable selection step since there are many more variables than data. Multicollinearity is one issue; as a result, estimation equations become singular and lack a stable, single solution. For instance, Fisher's linear discriminant function's pooled within-class sample covariance matrix is unique if n p + 2. Although though all genes may be employed in support vector machines, using all of the genes does not appear to make sense.

In fact, its application permits the presence of the noise linked to genes with weak or no genetic discriminating. When applied to an unclassified tumour, this hinders and impairs the functioning of the classification standards. Dimension reduction is required in this case to condense the high p-dimensional gene space. The authors of the majority of the publications previously mentioned have employed univariate techniques to reduce the number of genes. There are several methods that may be employed to solve the dimension reduction issue.

Chemometrics has seen the rise of similar data structures. In terms of statistics, PLS with PCR, it has been discovered that the partial least squares approach as well as principal component regression (PCR) are both helpful dimension reduction techniques. In regard to microarrays, the goal of the use of PCR, orthogonal tumour descriptors that reduce the dimension to a small number of gene components will be created. Nevertheless, the dimension reduction is carried out rather than taking the answer variable into account also perhaps effective. For prediction based on dimension reduction, PLS appears to be better suited than PCR. A linear combination of the p predictors (genes) with the response exhibits the highest sample covariance, which is why PLS components are actually chosen in this manner.

In children, non-Hodgkin lymphoma along with the Ewing family of cancers are together referred to as small, round blue cell tumours because of their resemblance in appearance on standard histology1. Nonetheless, a correct diagnosis of SRBCTs is crucial since there are a variety of treatment choices, therapeutic responses, and prognoses that rely on the diagnosis. These malignancies are difficult to detect by light microscopy, as their name suggests, and there is presently no one test that can accurately separate these tumours. With the identification of tumor-specific translocations in alveolar rhabdomyosarcoma, molecular methods like RT-PCR are increasingly employed for diagnostic confirmation (ARMS).

Unfortunately, molecular markers not necessarily offer a conclusive diagnosis since occasionally the conventional translocations cannot be detected because of the existence of variant translocations or technological problems. Several indicators may be analysed simultaneously using cDNA microarrays for gene-expression profiling, which has been used to divide tumours into subgroups. Although there are several statistical methods available for analysing gene-expression data, none of them have been thoroughly evaluated for their ability to identify effectively between tumours that fall into different diagnostic categories. One of the genetic research industry's fastest-growing technologies is DNA microarrays. These are tiny, stable platforms, such glass slides or membranes, where DNA sequences are anchored in a systematic pattern.

To examine and monitor the activity of genes, One plate might have tens of thousands of DNA probes attached on it. By observing variations in gene expression, researchers are able to better understand how cells react to various events (such as cancer, pest management, etc.) by utilising DNA microarrays. Although though microarrays are a potent source of biological data, it remains a difficult research challenge to use gene expression data to identify illnesses on a molecular level for clinical diagnosis.

It entails analysing gene expression levels from many studies, selecting significant genes (feature extraction), and then using precise and understandable categorization methods to reveal biological insight into the desired phenomena (classification). Machine learning techniques typically face a number of difficulties when categorising microarray data. The "small N, huge P" dilemma in statistical learning, where P is the number of variables (gene expressions), which is frequently significantly greater than N is the quantity of samples obtainable, is particularly relevant to microarray categorization. Six gene expression data sets are used in this study to test the system's hybrid Ant Bee Algorithm (ABA).

2. LITERATURE SURVEY

The classification algorithm for cDNA microarray data presented in this study, which is predicated upon graph theory, is able to address the majority of the drawbacks of existing classification techniques. In order to make predictions, the suggested decision rule attempts to mimic a human cognitive process by taking into account both the proximity scores' absolute value with their corresponding values.. It is significant to note that each proximity score that makes up PVs is calculated independently of the number of classes that are taken into account by the issue, making it a precise measure of how similar a sample is to a particular class. This contrasts with a number of cutting-edge classifiers where proximity measurements frequently rely on classes offered, as well as must be understood in respect to the average of all classes that were examined [1].

In order to categorise binary answer variables, this research suggests a novel approach that combines PLS with Ridge penalised logistic regression. The dimension-reduction step is integrated into the classification phase, and the approach includes a Ridge penalty step with a PLS step. For the leukaemia, colon, along with prostate data sets, the classification rule's predictive ability is demonstrated. As PLS handles continuous answers, it appears counterintuitive to directly apply PLS to GLM. The binary instance, however, is the most straightforward case that enables us to highlight whether or not such a process is effective and why. The classification of cancers

applying microarray gene expression data was proposed by the author as a statistical dimension-reduction strategy. Using this strategy, high-dimensionality of the gene expression space will be solved as well as the dimensionality curse. Also, a novel application of partial least squares to binary answer data was presented, which appears to have superior features to some of the already in use techniques. The approach may be expanded to include issues with several classes, and the selection of the parameter requires more consideration in the multi-class scenario than it does in the binary case [2].

This study created a technique for categorising tumours into distinct diagnostic groups based on their gene expression profiles (ANNs). Small, round blue-cell tumours (SRBCTs), which fall into four different diagnostic categories, were used as a model for training the ANNs. Further blinded samples were examined in order to assess the trained ANN models' capacity to identify SRBCTs. All samples were appropriately categorised by the ANNs, which also found the genes that were most important for the categorization. An additional technique for understanding tumour biology and the possibility for molecularly diagnosing cancer is the monitoring of global gene-expression levels using cDNA microarrays. A collection of 96 genes that are highly important to these tumours were discovered as a result of this quality filtration, which also yielded more reliable prediction models. By using more extensive arrays and greater training sample sets, this list may be extended [3].

Recent research has demonstrated via means of microarray gene expression data for phenotypic categorization of several disorders. Yet, there are far more characteristics (genes) than cases (tissue samples). A limited sample of genes can be chosen as features for classification using feature selection approaches to get around the curse of dimensionality problem. The author created a hybrid strategy that integrates analysis of gene clusters with ranking of genes to accomplish this aim. As compared to methods that employ top-ranked genes directly for classification, our methodology is capable of picking a small number of marker genes while providing the same or superior leave-one-out cross-validation accuracy. Compared to traditional methods that directly use all of the top-50 or top-100 rated genes, it can increase classification accuracy. The use of Gene Ontology to direct this selection procedures is one potential strategy that the authors are now looking into [4].

This article describes a technique for automatically classifying text documents that uses symbolic rule induction and decision trees. The system employs a quick algorithm for decision tree induction and a novel technique for transforming a decision tree into a smaller a rule set that is technically equivalent to the source tree notwithstanding this. It assigns a confidence measure—the estimated in-class chance that a document complies with the rule—to each rule. The KitCat approach allows one to focus on one category at a time by considering categorization in relation to each category as a distinct binary-classification issue. This results in a different rule set for each category. It contains a approach for turning a decision tree into a symbolic rule set by breaking down a complex rule set into its simplest equivalent based on an analysis of the decision tree's structure, as well as a justification for the soundness of this method [5].

3. PROPOSED SYSTEM

To create a precise fuzzy system with easy-to-understand, condensed rules The suggested approach uses each data set separately, a hybrid Ant Bee Algorithm (ABA). The technique makes advantage of mutual information to identify informative genes. Because it has strong empirical achievements, is resilient, scalable, and nonlinear. Six gene expression data sets are used to assess the proposed method.

According to their appearance and protein expression, histology and immunohistochemistry are presently used to diagnose tumours. Yet, by means of standard histology, weakly differentiated tumours might be challenging to identify. However, a tumor's histological appearance cannot show the underlying genetic anomalies or biological mechanisms that support the malignant process. An ant colony and a list of permitted ranges (PRs) connected to each potential discrete value of the design variable are maintained via ant colony optimization. Each ant may select an acceptable range that denotes the path.

DOI : https://doi.org/10.56452/7-793

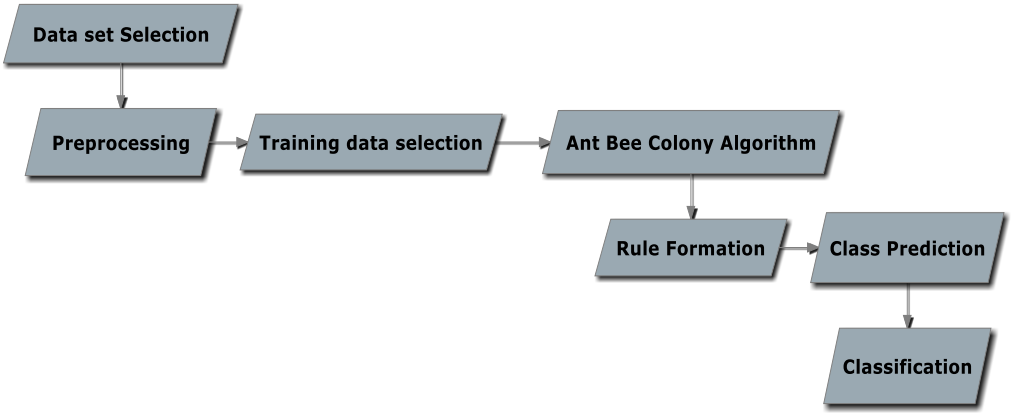

Fig 1: System Architecture

Once every ant in the colony selects a path, the candidate value for the ants is the potential discrete value associated with that path. The goal function is then evaluated using a mixture of the possible values from each ant. The more closely a point on the ROC curve resembles an individual point in the ROC space, the more accurate the test is thought to be. The test's accuracy decreases with increasing distance from the diagonal on the ROC curve.

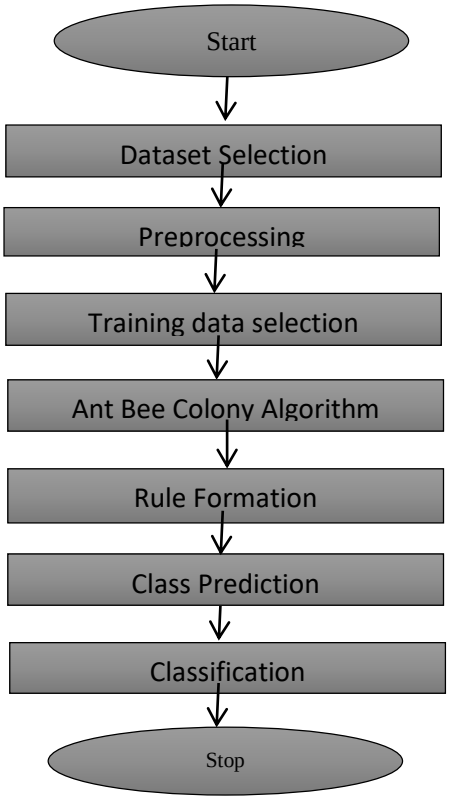

Fig 2: Flow Diagram

Four distinct strategies were created to evaluate how well the proposed ABA performed in comparison. First up is the Binary Coded Genetic Algorithm, which employs bitwise mutation, tournament selection, and two point crossover as well as fundamental genetic operators as solution variables. The next algorithm is Real Coded Genetic Algorithm, which employs tournament selection, BLX-a crossover, and Non-uniform mutation for genetic operation and expresses solution variables as floating point values.

DOI : https://doi.org/10.56452/7-793

The suggested ABA is seen to accomplish each rule's gene is filtered depending on expression values using linguistic selection, which renders the rule set simpler, more interpretable, and more compact than GA. Also, the suggested ABA's ABC algorithm effectively optimises the membership function to be compared to GA and PSO, has a higher classification accuracy and is less complicated. The following are only a few of the suggested approach's many benefits: precision.

- For each data collection, it produces concise rules.
- The study of the disease's nature and the genetic pathways causing it is aided by the use of a restricted selection of genes.
- It can be scaled.

4. RESULTS

In order to make decisions regarding data, fuzzy expert systems require a set of membership functions and rules. Microarray technology continues to present difficulties despite significant advancements in the discovery of cancer molecular profiles. The dependability of the training data sets utilized to create the classifiers as well as the performance of the classifiers are the two key constraints of current methods for classifying data from microarrays. Thermography in medicine has proven effective in a number of medical contexts, including the identification of breast cancer. The system suggests a hybrid Ant Bee Algorithm to handle the interpretability-accuracy tradeoff, also it is tested by employing six gene expression data sets. Comparatively simpler than GA and PSO, the ACO and ABC algorithms increase classification accuracy.

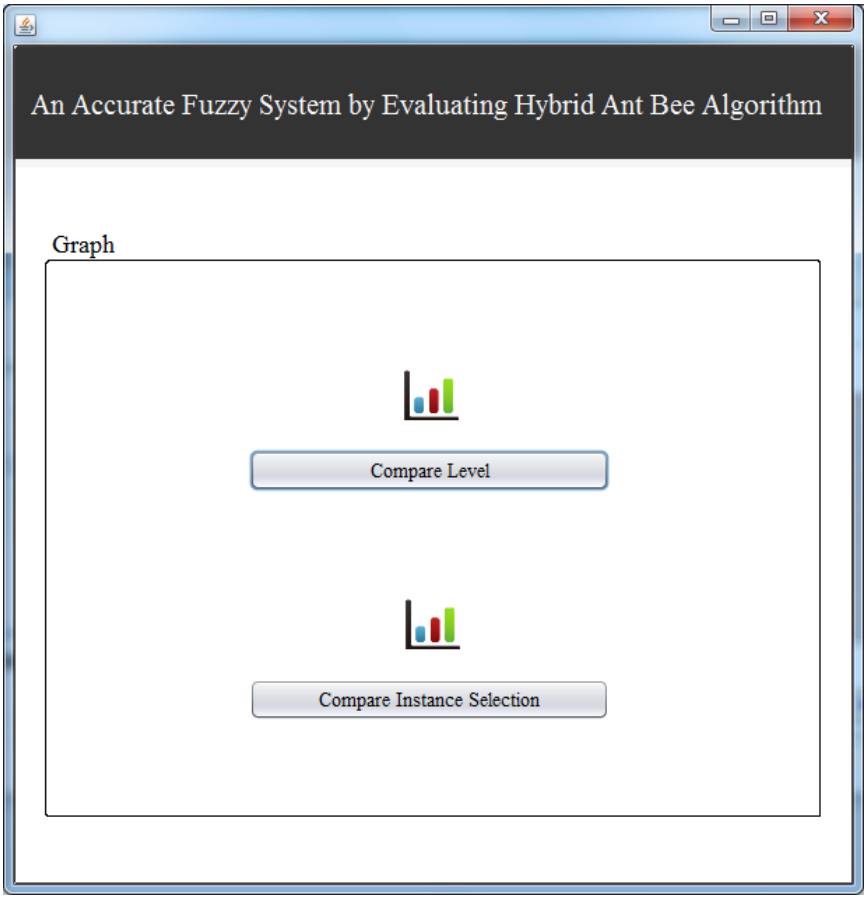

Fig 3: Graph

DOI : https://doi.org/10.56452/7-793

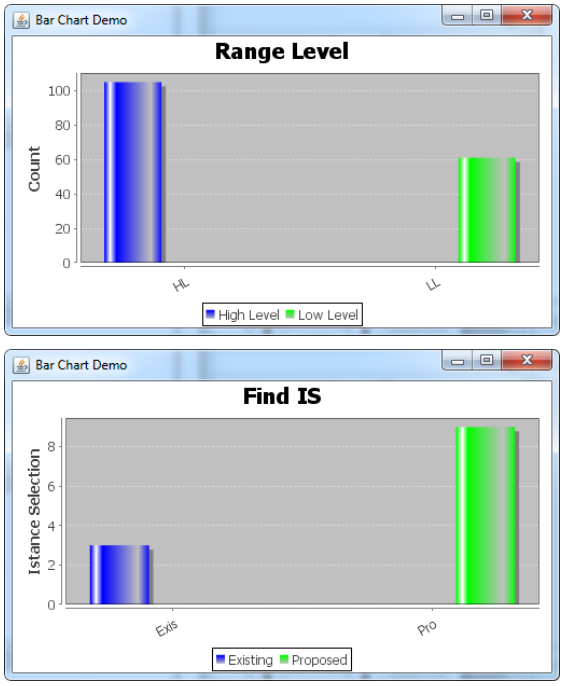

Fig 4: Performance Analysis

5. CONCLUSION

In this project, the system's proposal for the development of a fuzzy expert system is the Ant Bee Algorithm. Despite significant progress in understanding cancer molecular profiles, it is still difficult to effectively use microarray technology to standard clinical diagnoses. According to their appearance and protein expression, histology and immunohistochemistry are presently used to diagnose tumors. Yet, by means of standard histology, weakly differentiated tumours might be challenging to identify. However, a tumor's histological appearance cannot show the underlying genetic anomalies or biological mechanisms that support the malignant process.

An ant colony and a list of permitted ranges (PRs) connected to each potential discrete value of the design variable are maintained via ant colony optimization. Each ant may select an acceptable range that denotes the path. The dependability of the training data sets employed to create the classifiers along with the performance of the classifiers, particularly when the sample to be categorised failed to correspond to any of the offered classes, are the two fundamental constraints of current procedures in the categorization of microarray data. Medical thermography has demonstrated its value in a number of implications in medicine, such as the identification of breast cancer wherever it is capable pinpoint the localized rise in temperature caused by cancer cells' intense metabolic activity. The approach has been presented to handle the sample classification accuracy interpretability tradeoff. Eventually, the outcome shown that the suggested ABA technique produced a small, precise, and understandable fuzzy expert system.

REFERENCE

[1] T.L. Bergemann and L.P. Zhao, “Signal Quality Measurements for cDNA Microarray Data,” IEEE/ACM Trans. Computational Biology and Bioinformatics, vol. 7, no. 2, pp. 299-308, Mar./Apr. 2010.

[2] A. Benso, S.D. Carlo, and G. Politano, “A cDNA Microarray Gene Expression Data Classifier for Clinical Diagnosis Based on Graph Theory,” IEEE/ACM Trans. Computational Biology and Bioinformatics, vol. 8, no. 3, pp. 577-591, May/June 2011.

[3] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri,

C.D. Bloomfield, and E.S. Lander, “Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring,” Science, vol. 286, pp. 531-537, 1999.

[4] L. Li, “Gene Selection for Sample Classification Based on Gene Expression Data: Study of Sensitivity to Choice of Parameters of the ga/knn Method,” Bioinformatics, vol. 17, pp. 1131-1142, 2001.

[5] S. Dudoit, J. Fridlyand, and T.P. Speed, “Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data,” J. Am. Statistics Assoc., vol. 97, no. 457, pp. 77-87, 2000.

[6] G. Fort and S.L. Lacroix, “Classification Using Partial Least Squares with Penalized Logistic Regression,” Bioinformatics, vol. 21, no. 7, pp. 1104-1111, 2005.

[7] L. Fan, K.L. Poh, and P. Zhou, “A Sequential Feature Extraction Approach for Na€ıve Bayes Classification of Microarray Data,” Expert Systems with Applications, vol. 36, no. 6, pp. 9919-9923, 2009.

[8] J. Khan, J.S. Wei, M. Ringner, L.H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C.R. Antonescu, C. Peterson, and P.S. Meltzer, “Classification and Diagnostic Prediction of Cancers Using Gene Expression Profiling and Artificial Neural Networks,” Nature Medicine, vol. 7, pp. 673-679, 2001.

[9] T.S. Furey, N. Cristianini, N. Duffy, D.W. Bednarski, M. Schummer, and D. Haussler, “Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data,” Bioinformatics, vol. 16, pp. 906-914, 2000.

[10] A.C. Tan and D. Gilbert, “Ensemble Machine Learning on Gene Expression Data for Cancer Classification,” Applied Bioinformatics, vol. 2, pp. 75-83, 2003.

[11] D.E. Johnson, F.J. Oles, T. Zhang, and T. Goetz, “A Decision-TreeBased Symbolic Rule Induction System for Text Categorization,” IBM Systems J., vol. 41, no. 3, pp. 1-10, 2002.

[12] J.S.R. Jang, C.T. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computting. Prentice Hall, 1997.

[13] A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, and D. Geman, “Simple Decision Rules for Classifying Human Cancers from Gene Expression Profiles,” Bioinformatics, vol. 21, pp. 3896-3904, 2005.

[14] Y. Yoon, S. Bien, and S. Park, “Microarray Data Classifier Consisting of k-Top-Scoring Rank-Comparison Decision Rules with a Variable Number of Genes,” IEEE Trans. Systems, Man, and Cybernetics-Part C: Applications and Rev., vol. 40, no. 2, pp. 216-226, Mar. 2010.

[15] P. Woolf and Y. Wang, “A Fuzzy Logic Approach to Analyzing Gene Expression Data,” Physiological Genomics, vol. 3, pp. 9-15, 2000.

[16] S. Vinterbo, “Small, Fuzzy and Interpretable Gene Expression Based Classifiers,” Bioinformatics, vol. 21, no. 9, pp. 1964-1970, 2005.

[17] G. Schaefer, “Thermography Based Breast Cancer Analysis Using Statistical Features and Fuzzy Classification,” Pattern Recognition, vol. 42, no. 6, pp. 1133-1137, 2009.