

Image Classification using Hybrid Data Mining Algorithms – A Review

Thamilselvan P/Research Scholar
Department of Computer Science
Bishop Heber College, Tiruchirappalli,
Tamilnadu, India.
Thamilselvan1987@gmail.com

Dr. J. G. R. Sathiaselvan
Head, Department of computer Science
Bishop Heber College, Tiruchirappalli,
Tamilnadu, India.
jgrsathiaselvan@gmail.com

Abstract—Data mining is one of the most significant research area in computer science. It is a calculation process of finding and determining valuable information from huge data set. Image classification is an important technique to generate valuable information. The classification method provides the accurate result in their target class. This review compares the some predominant hybrid classification algorithms to find the classification accuracy for various data sets and their performance of techniques. It provides some important hybrid techniques that have been used for image classification. In this paper the hybrid data mining algorithms are studied like GA-SVM, EKM-EELM, AdaBoost-SVM, Decision Tree-Naive Bayes, and SVM-CART.

Index Terms— Hybrid Approach; Image Classification; Image Mining; Image datasets; Mining algorithms.

I. INTRODUCTION

A Data mining is a process of searching and extracting valuable information from a huge set of data. The data source contains data warehouses, web, information repository and databases flowed into system enthusiastically. The mining functionality is used to specify the patterns in data mining tasks. The fundamental data mining algorithms are emerging field in data science that includes automated algorithm to analyze the models and patterns for each kind of data. Data mining plays a vital role in database technology and visualization, artificial intelligence, pattern recognition, machine learning. The objective of building computer system is adaptive to environments and learning from researchers in many areas such as engineering, computer science, physics, neuroscience, mathematics. This research has come variety learning method that transforms scientific fields.

Mining is extracting knowledge from the huge amount of data has been anticipated by many research topics in database system. The machine learning methods are used in image technique [1] [2]. Data mining is used to collect actual models, abnormalities, roles, relations and patterns from the huge data. The most of the statistical models are used in feature extraction using data mining techniques some method based on fuzzy method [3] [4] and neural networks [5]. Image classification is one of the significant area in image mining it increases the demand for developing real world vision systems. Image

classification is one of the fundamental activities in image mining. In this survey, we have analyzed data mining method for image classification [6-10]. In general, image classification has two foremost stages. The first stage defines an effective representation of an image. It includes necessary information of the image for classification. The second stage classifies the new image without error. There are some impediments in image classification: dimensionality reduction and classifier. Image classification analysis the numerical values of various image features and grouping data into a class. The classification algorithms typically two types processing: training and testing. On training part usual images are isolated based on the unique description. On testing part images are partitioned based on the image features.

The number of combined classifiers is called as modular classifiers, aims to achieve high accurate classification. The performance of ensemble classifier is better than best the single classifier used in isolation. The simple technique to hybrid classifiers is high voting. The number of trained classifiers shares their inputs and outputs to yield to produce complete output. The experts can trained by different features of a given learning model trained by the same learning models. This study compares the performance of some hybrid data mining algorithms. The performance of data mining algorithms is calculated by based on the image classification accuracy.

II. LITERATURE SURVEY

Nabiha Azizi et al. [6] presented a new combined approach genetic algorithm and support vector machine to classify Mammogram images. Based on this approach aims to how to choose features from inaugural vector without reducing global classification performance. This work represents three fundamental physiognomies based on the image which are Co-occurrence matrix, HU moments and central moments. This author proposed a CAD system based on the feature selection in each stage which are preprocessing technique, classification and feature extraction. This proposed algorithm shows the 89% accuracy in image classification.

Feilong Cao et al. [7] suggested a new hybrid algorithm by using EKM and EELM. The EKM technique imitative from K-Means method and the EELM technique derivative from Extreme Learning Machine Algorithm. This work carried out

by using EKM and EELM algorithm for image classification based on a database. The EKM technique is used to classify dimensionality of reduced curvelet feature sets that is selected for each database arbitrarily. The ELM technique is a tremendously fast learning technique in training. This proposed method accomplishes better classification accuracy in human 2D face images. It illustrates 90% (approximately) accuracy in image classification.

Ghazanfar Raza et al. [8] presented a new hybrid algorithm by using naive bayes and support vector machine for drusen detection from fundus images. This proposed algorithm represents each region with number of features and then it's applied in this hybrid classifier as a naive bayes and SVM to classify the region as drusen and non-drusen. This algorithm estimated by STARE database using the accuracy, sensitivity and specificity. Different authors have presented different computer aided diagnostic system for different retinal syndromes [11-14]. This hybrid method provides good accuracy which means the fundus images are correctly classified. The work shows the accuracy, specificity and sensitivity are 0.98, 0.99 and 0.97 respectively. The system precisely segmented the images outstrips than the earlier proposed system. Finally, this proposed hybrid algorithm shows 98% accuracy in image classification. Support vector machine predicts most accurate results in classification, particularly in text classification.

Dewan Md. Farid et al. [9] anticipated a hybrid algorithm using decision tree and naive bayes algorithm. This method is mainly aimed to increase the accuracy of classification for multi class-class classification tasks. This proposed hybrid algorithm shows the better sensitivity, specificity, cross validation and classification accuracy on real benchmark data sets from UCI. The proposed system automatically extracts the value in training data sets. Moreover, it identifies the effective attributes from noisy complex training data sets. Data mining algorithm is used by computational intelligent researchers for solving classification and clustering problems [15-17]. The proposed hybrid algorithm shows the 90% accuracy in image classification.

Lufimpu-Luviya Yannick et al. [10] developed a hybrid algorithm by using Support Vector Machine and Classification and Regression Tree (CART) to identify the age band of a 2D image face. The first method is used to find the age of face images and the second method is used to solve the classification problem. This result illustrates that second method overcomes the first method. The second method produces some risky confusions. This proposed system provides two advantages: a. It shrinkage the computational time and b. It provides the same results in the entire face. It is more difficult to predict the age band and provides lower performance in that process. This work is carried out peculiar the gender rendering to the marketing business constraints. This proposed hybrid algorithm provides 84% accuracy in image classification.

Alireza Taravat et al. [18] anticipated hybrid algorithm for to increase the development of sky images. In this experiment multilayer perceptron neural network and support vector

machine algorithm are hybrid. This two classification algorithm (MLPNN and SVM) have been compared. The proposed algorithm is applied in dataset containing more than 250 images from various test sites. The obtained accuracy shows the better detection result. The results show MLPNN algorithm is automated algorithm for image classification.

A. Kannan et al. [19] suggested a hybrid algorithm for classifying the MRI images. Here, the authors designed hybrid method using KNN and SVM concepts for MRI image classification. K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are the most prevalent data mining method in image classification. The proposed system shows the 86.7% accuracy in image classification. The accuracy of the result is basically increased and the error rate has been progressively reduced.

Bor-Chen Kuo et al. [20] presented a hybrid algorithm using Radial Basis Function (RBF) and Support Vector Machine to improve the classification accuracy in images. This algorithm is analyzed in real time datasets to measure the classification performance. In this approach two things can be achieved coefficients being calculated: small subset of ranking and features. The proposed method demonstrates 90% accuracy in image classification. From this experiment classification accuracy is increased considerably from the subset.

M.R. Homaeinezhad et al. [21] grants a hybrid algorithm using support vector machine and k-means algorithm. The proposed hybrid algorithm mainly aimed to increase the robustness of classification accuracy in QRS images. In this proposed method the ECG signal detected and delineate using robust wavelet based algorithm. This method shows higher performance than other method. The proposed classification method shows error rate variation to another algorithm for giving common data sets. To calculate the performance of the proposed hybrid algorithm, the succeeded results were associated with several related studies. The proposed method shows better accuracy in QRS images.

Bichen Zheng et al. [22] enhanced a hybrid method by using k-means and support vector machine. This algorithm is mainly aimed to improve the classification accuracy tested on a breast cancer image data set. The proposed hybrid approach tested only on a breast cancer image dataset, it also savings the training phase time. It is a potential way for reducing the dimension of the training set. The proposed method shows the better classification accuracy in breast cancer images.

Mohammad Reza Zare et al. [23] cultivated a hybrid algorithm by using generative and discriminative approach. The proposed method is precisely predict the target value in each case of data. This is mainly designed to expand the classification performance medical images. This model is used to classify the any type of X-ray images. In training process, the classification method finds the relationships among the predictor value. Different classification methods used for discovering the relationships. The used dataset 11 000 X-ray medical images from 116 groups. This proposed hybrid method shows 92.5% accuracy in classification in medical X-ray images.

III. HYBRID APPROACH

Hybrid algorithms are alternatives for increasing realistic, complex problems, uncertainty, involving ambiguity and high dimensional data. Figure 1 shows the uneven representation of the computational areas covered by the hybrid approach. This hybrid approach represents the combinations of computational techniques to solve a problem. The hybrid intelligent system describes the software system in parallel. It contains four techniques such as hybrid optimization method, combined classifier, Nature inspired system and uncertainty management. The hybrid intelligent system is a combinational methods of artificial intelligence.

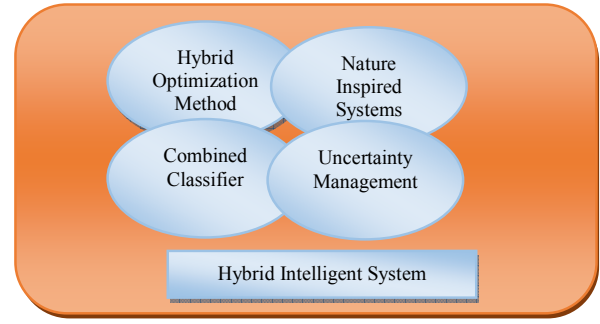


Fig. 1. Domains of Hybrid Intelligent System

TABLE I. HYBRID ALGORITHMS TAKEN FOR THIS REVIEW

Table 1	Description of Hybrid methods						
S. No	<i>Proposed Hybrid Approach</i>	<i>Acronym</i>	<i>Year</i>	<i>Developed By</i>	<i>Implementation Tool</i>	<i>Limitation</i>	<i>Purpose of Development</i>
1	Genetic Algorithm and Support Vector Machine	GA-SVM	2014	Nabiha Azizi et al.	WEKA	Shows the high error rate	To reduce the dimensionality and optimize the classification process.
2	Extreme K-Means Algorithm and Effective Extreme Learning Machine	EKM-EELM	2013	Feilong Cao et al.	MAT LAB R2010	Slow Training Process	To improve the classification accuracy.
3	Naive Bayes and Support Vector Machine	NB-SVM	2013	Ghazanfar Raza et al.	MAT LAB	Several key parameters needed to achieve the best classification results	To improve the performance of classification, specificity, and sensitivity
4	Support Vector Machine and Classification regression tree	SVM-CART	2013	Lufimpu-Luviya Yannick et al.	MAT LAB	The Regression gives the extreme confusion	To identify the age band of 2D image face
5	Decision Tree and Naive Bayes	DT-NB	2014	Dewan Md. Fardi et al.	WEKA3	Provides less compact solution	To improve the classification accuracy of DT-NB classifier for the classification of multi class problems.

IV. COMPARATIVE RESULTS

In this portion, testing datasets and experimental results for all the proposed hybrid approach are to be discussed.

A. Datasets

The performance of all the hybrid approaches is tested on different image data sets. The table II describes the datasets are used in the different hybrid algorithm.

TABLE II. SELECTED DATASETS

Table 2	Dataset Description	
S. No	Hybrid algorithm	Testing Dataset
1	GA-SVM	Cancer Image Dataset
2	EKM – EELM	Face Image Dataset
3	Naive Bayes – SVM	Retinal Image Dataset
4	Decision Tree – Naive Bayes	Iris plants, Breast Cancer Image, Tic-Tac-Toe
5	SVM – CART	Human Face Images

B. Genetic Algorithm and Support Vector Machine

The primary focus of this algorithm is to optimize the classification process precision, specificity analysis and cross validation on real benchmark. The assigned SVM classifier in feature vector will generate three types of SVM classifier. This classifier is trained during the learning stage. The main influence of this paper implementation of Genetic Algorithm to select the discriminating feature. This algorithm reduces the size of feature vector by 25% and also the recognition rate has increased.

C. Extreme K-Means Algorithm and Effective Extreme Learning Machine

The objective of this method is aimed to improve the classification accuracy in human face images. The EKM method is used to cluster dimensionality in each database. This method has an image decomposition with reduced dimensionality, curvelet transforms with locality alignment. It achieves better accuracy in image classification.

D. Naive Bayes and Support Vector Machine

This technique is mainly targeted to improve the performance of classification, sensitivity, and specificity in retinal images. NB and SVM are combined using weighted probabilistic ensemble. This method gives some important information from the retinal image and presented some automated system for detection of drusen in color images.

E. Support Vector Machine and Classification and regression Tree

This method is aimed to develop identify the age band of the person from 2D face image. The main idea of this approach first calculates the age, then predict the threshold values used

to get the age band of 2D face images. In this method both classification and regression approaches are to be tested. For this experiment only male face images only taken. This proposed method (SVM-CART) produce 84% accuracy in image classification.

F. Decision Tree and Naive Bayes

The purpose of this hybrid method is to improve the classification accuracy in multi class problems. This method is used to eliminate the noisy instances from training data sets. The enactment of this method is tested against traditional NB and DT classifiers and using classification accuracy. This method (DT-NB) shows 90% accuracy in image classification.

V. COMPARISON OF ACCURACY

In this section, performance of proposed hybrid approaches is taken for the analysis. The table III shows the overall performance of the hybrid method.

TABLE III. PERFORMANCE OF HYBRID APPROACH

Table 3	Hybrid Methods	
S. No	Algorithms	Accuracy
1	GA-SVM	89%
2	EKM-EELM	90%
3	NB-SVM	98%
4	SVM-CART	84%
5	DT-NB	90%

In figure 2 the classification accuracy result is analyzed based on the performance of hybrid techniques that combines GA-SVM, EKM-EELM, NB-SVM, SVM-CART and DT-NB.

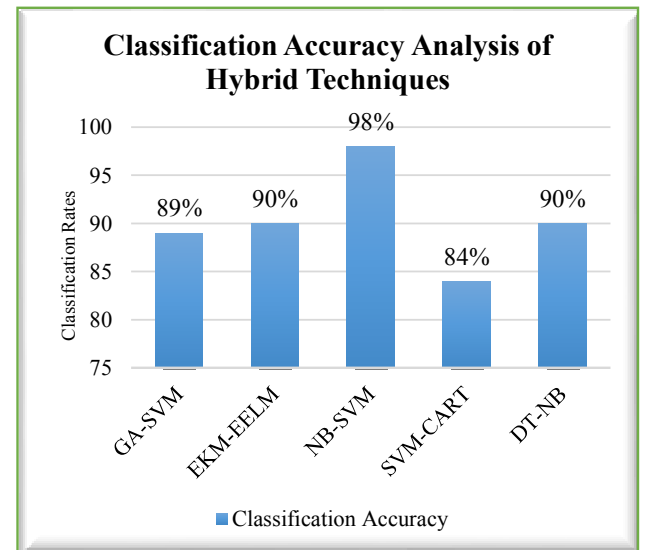


Fig. 2. Classification Accuracy of Hybrid Methods

In health applications, the data mining algorithms are used to maximize the accuracy of medical images for accessing disease of the patients [24]. The accuracy of classification from this study was found to be 89 % (GA-SVM), 90 % (EKM-EELM), 98 % (NB-SVM), 84 % (SVM-CART) and 90 % (DT-NB) respectively. From this study, we have analyzed some hybrid data mining techniques based on image classification. From those hybrid methods naive bayes and support vector machine combined method provides better accuracy in image classification when compared with other hybrid data mining algorithms. The hybrid intelligent system provides better performs when compare with enhanced data mining algorithm, but it's difficult to hybrid the methods.

VI. CONCLUSION AND FUTURE WORK

Image Classification is the most major part of image mining and digital image processing. In this study, we have deliberated evolution of classification method based on different hybrid approach. The performance of hybrid methods was analyzed based on the classification accuracy, advantages and characteristics of the reference data points. From this study the hybrid method NB-SVM shows better accuracy in image classification when compared to other hybrid approaches. In future, we aims to combine atleast two data mining algorithms. By applying the proposed hybrid algorithm, it is intended to find better image classification accuracy and furthermore reduce the computational time complexity than other hybrid methods.



Mr. P. Thamilselvan, received the M. Phil. Degree in Computer Science from Bharathidasan University, Tiruchirappalli, Tamilnadu, India in 2013. He also received M.C.A. Degree in Computer Applications from Anna University, Chennai, Tamilnadu, India in 2012. He is currently pursuing the Ph.D. in Computer Science from Bharathidasan University, Tiruchirappalli Taminadu, India. His area of research interests includes Artificial Neural Networks and Image Mining. E-mail: thamilselvan1987@gmail.com



Dr. J. G. R. Sathiaselvan is the Head of Computer Science Department at Bishop Heber College, Tiruchirappalli. He has 25 years of teaching experience. He has presented more than 20 research papers in International conference publication which are published in IEEE, ACM, Springer, and reputed journals. Dr. Sathiaselvan has authored a book entitled as, "Programming In C#, .Net", which was published in PHI, New Delhi, in 2009. His research areas include Web Services security, Data mining, image processing and big data analytics. E-mail: jgrsathiaselvan@gmail.com

REFERENCES

- [1] Osmar R. Zaiane, Antonie Maria-Luiza, Alexandru Coma. "Mammography classification by association rule based classifier," International conference on multimedia data mining ACM SIGKDD, pp 62–69, 2002.
- [2] Xia Xuanyang, Gong Yuchang, Wan Shouhong, Li Xi. "Computer aided detection of SARS based on radiographs data mining," The 27th IEEE Conference on engineering in medicine and biology, Shanghai, China, pp 7459–7462, 2005.
- [3] Shuyan Wang, Mingquan Zhou, Guohua Geng. "Application of fuzzy cluster analysis for medical image data mining," International IEEE conference on Mechatronics and automation, Niagara Falls, Canada, pp 36–41, 2005.
- [4] R. Jensen, Qiang Shen. "Semantics preserving dimensionality reduction: rough and fuzzy-rough based approaches," The IEEE transactions on knowledge and data engineering, pp 1457–1471, 2004.
- [5] I. Christyanni, E. Dermatas, G. Kokkinakis. "Fast detection of masses in computer aided mammography," International IEEE Conference on signal processing, pp 54–64, 2000.
- [6] Nabiha Azizi, N. Zemmam, M. Sellami, N. Farah. "A new hybrid method combining genetic algorithm and support vector machine classifier: Application to CAD system for mammogram images," 2014 IEEE International Conference on Multimedia computing and systems, pp 415–420, 2014.
- [7] Feilong Cao, Bo Liu, Dong Sun Park. "Image classification based on effective extreme learning machine," Elsevier, Neurocomputing 102, pp 90–97, 2013.
- [8] Ghazanfar Raza, M. Rafique, A. Tariq, M. U. Akram. "Hybrid classifier based drusen detection in colored fundus images" 2013 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT), ISBN 978-1-4799-2303-8, pp 8–13, 2013.
- [9] Dewan Md. Farid, Li Zhang, Chowdhury Mofizur Rahman, M. A. Hossain, Rebecca Strachan. "Hybrid decision tree and Naive bayes classifiers for multi-class classification tasks," Elsevier, Expert systems with applications, pp 1937–1946, 2014.
- [10] Lufimpu-Luviya Yannick, P. Sebastien, M. Djamel, F. Farnard. "Combining Classification and Regression approaches for age band estimation from human faces," IEEE 8th International Symposium on Image and Signal Processing and Analysis, pp 136–141, 2013.
- [11] Usman M. Akram, Shoab A. Khan. "Automated detection of dark and bright lesions in retinal images for early detection of diabetic retinopathy," Springer, Journal of Medical Systems (JOMS), vol. 36, no. 5, pp 3151–3162, 2012.
- [12] Anam Tariq, M. Usman Akram, Arslan Shaukat, Shoab A. Khan. "Automated Detection and Grading of Diabetic Maculopathy in Digital Retinal Images," Springer, Journal of Digital Imaging, vol. 26, no. 4, pp. 803–812, 2013.
- [13] M. U. Akram, A. Tariq, M. A. Anjum, M. Y. Javed "Automated Detection of Exudates in Colored Retinal Images for Diagnosis of Diabetic Retinopathy," OSA Journal of Applied Optics, volume. 51 no. 20, pp 4858–486, 2012.
- [14] M. Usman Akram, Shehzad Khalid, Shoab A. Khan. "Identification and Classification of Microaneurysms for Early Detection of Diabetic Retinopathy," Elsevier, Pattern Recognition, vol. 46, no. 1, pp 107–116, 2013.
- [15] Dewan Md. Farid, Li Zhang, Alamgir Hossain, Chowdhury Mofizur Rahman, Rebecca Strachan, Graham Sexton, Keshav Dahal. "An adaptive ensemble classifier for mining concept drifting data streams," Elsevier, Expert Systems with Applications, 40, pp 5895–5906, 2013.

- [16] Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Hsiao. "Data mining techniques and applications A decade review from 2000-2011," Elsevier, *Expert Systems with Applications*, 39, pp 11303–11311, 2012.
- [17] E. W. T. Ngai, Li Xiu, D. C. K. Chau. "Application of data mining techniques in customer relationship management: A literature review and classification review article," Elsevier, *Expert Systems with Applications*, 36 pp 2592–2602, 2009.
- [18] Alireza Taravat, F. Del Frate, C. Cornaro, S. Vergari. "Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky, ground-based images," *IEEE Geoscience and remote sensing letters* vol. 12, no. 3 pp 666–670, 2014.
- [19] A. Kannan, V. Mohan, N. Anbazhagan "An Implementation of Hybrid Algorithm for Diagnosing MRI Images Using Image Mining Concepts," Springer-Verlag Berlin Heidelberg, *ICIEIS*, Part II, CCIS252, pp. 140–150, 2011.
- [20] Bor-Chen Kuo, Hsin Hua Ho, Cheng Hsuan Li, Chih Cheng Hung, Jin-ShiuhTaur. "A Kernel-Based Feature Selection Method for Support Vector Machine With Radial Basis Function Kernel for Hyperspectral Image Classification," *IEEE Journal for Applied Earth Observations And Remote Sensing*, Vol. 7, No. 1, pp 317-326, 2014.
- [21] M.R. Homaeinezhad, S. A. Atyabi, E. Tavakkoli, H. N. Toosi, A. Ghaffari, R. Ebrahimpour. "ECG arrhythmia recognition via a neuro-SVM–KNN hybrid classifier with virtual QRS image-based geometrical features," Elsevier, *Expert Systems with Applications* 39, pp 2047-2058, 2012.
- [22] Bichen Zheng, Sang Won Yoon, Sarah S. Lam. "Breast cancer diagnosis based on feature extraction using a hybrid of K-means Algorithms and support vector machine algorithm," Elsevier, *Expert Systems with Applications*, pp 1-7, 2013.
- [23] Mohammad Reza Zare, A. Mueen, M. Awedh, W. Chaw Seng "Automatic classification of medical X-ray images: hybrid generative-discriminative approach," *IET Image Process.*, 2013, Vol. 7, Issue. 5, pp 523-532, 2013.
- [24] Dr. M. Durairaj, P. Thamilselvan. "Applications of artificial neural network for IVF data analysis and prediction," *Journal of Engineering, Computers, and Applied Sciences (JEC and AS)*, Volume 2, No.9, ISSN No. 2319-5606, pp 11-15, 2013.