ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

Data Classification Particle Swarm Optimization and Gravitational Search Algorithm

Shivani Shrivastri[1], Rahul Deshmukh[2]

P.G. Student, Department of Computer Science and Engineering, MIST, Bhopal, Madhya Pradesh, India[1]

Associate Professor, Department of Computer Science and Engineering, MIST, Bhopal, Madhya Pradesh, India[2]

Abstract: Classification is an important problem in data mining. Under the guise of supervised learning, classification has been studied extensively by the AI community as a possible solution to the "knowledge acquisition" or "knowledge extraction" problem. Briefly, the input to a classifier is a training set of records, each of which is a tuple of attribute values tagged with a class label. A set of attribute values defines each record. Many different techniques have been proposed for classification, including Bayesian classification, neural networks, genetic algorithms and tree-structured classifiers. They have been successfully applied to wide range of application areas, such as medical diagnosis, weather prediction, credit approval, customer segmentation, and fraud detection and many more. PSOGSA based data classification can also be apply, might yield more efficient and promising results, work which possesses classification of standard data using gravitational search algorithm with optimize manner. So classification of data done by the famous widely used method Feed-forward neural network with gravitational search algorithm. Particle swarm optimization is a popular heuristic algorithm that had been applied on many optimization problems over the years including data classification problem. The modified PSO is combined with gravitational search algorithm to solve its slow Execution time in the last iterations, making the hybrid PSOGSA algorithm.

Keywords: Particle swarm optimization, Gravitational search, Classification, Bayesian classifier.

I. INTRODUCTION

Now a day, machine learning is one of the most innovative concepts in present research scenario. Therefore exploring machine learning along with data mining and its learning algorithms has lots of scope to work. In machine learning and data mining, classification is best for producing accurate, rapid and straight forward results and hence among several techniques of machine learning, classification has been selected. Machine learning denotes changes in the system that is adaptive in the sense that they enable the system to do the same task more effectively the next time. In recent year many successful machine learning applications have been developed, ranging from data mining program that learn user reading preferences to autonomous vehicles. Machine learning is also used in various fields of real life application like, statistics, artificial intelligence, philosophy, information theory, biology, cognitive science, computational complexity and control theory, medical, finance, engineering, aeronautics and A comparative study of well-known classification methods are presented. It is concluded that there is no single best pruning method [6]. Even though the divide-and-conquer algorithm is quick, efficiency can become important in tasks with hundreds of thousands of instances. The most time-consuming aspect is sorting the instances on a numeric feature to find the best threshold t. This can be expedited if possible thresholds for a numeric feature are determined just once, effectively converting the feature to discrete intervals, or if the threshold is determined from a subset of the instances and many more. The goal of this work is to build a classification algorithm for classifying stock market data and reach a certain decision point that whether an analyst should sell, purchase or hold shares of a particular company. Elomaa and Rousu stated that the use of binary discretization with C4.5 needs about the half training time of using C4.5 multi splitting. In C4.5 multi-splitting of numerical features does not carry any advantage in prediction accuracy over binary splitting. Decision trees are usually univariate since they use splits based on a single feature at each internal node [6].

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

Data classification is a branch of Artificial Intelligence and has proved itself very useful in constructing a content-based image retrieval system. Data Classification plays an important role in the field of data analysis. Data classification is basically an attempt of labelling an image the appropriate identifiers. For determining these identifiers we specify the area of interest. This specification can be generalized, like considering arbitrarily taken pictures (say, the ones from the internet) for the classification, or may limit to some specific domain, say, medical images or geographical images (remotely sensed images).Image classification is a kind of image annotation, in which involves tagging or indexing the image in linguistic manner. However, image annotation uses far larger vocabulary (number of classes) as compared to image classification. Image classification can be defined as a process of assigning all the pixels contained by a digital image to a particular class from the set of classes on the basis of their characteristics. Actually need to distinguish between two types of classes: information classes and spectral classes.

i) Information classes: It is the area of interest or the content that the analyst is particularly trying to identify in the given imagery, for example, different kinds of crops, different types of forests or different species of trees, different rock types or geological areas, etc.

ii) Spectral classes: It contains the groups of pixels that are uniform (or almost-similar) with respect to their reflectance and brightness values corresponding to the different spectral channels of the data.

For a successful classification the prime requirements are the suitable classification system with the sufficient number of training samples. Three major problems that were identified by Foody [14], when the data used was of a medium spatial resolution, for the purpose of vegetation classifications defining proper hierarchical levels for the purpose of mapping. Defining a variety of land-cover units that are discernible by the selected data.

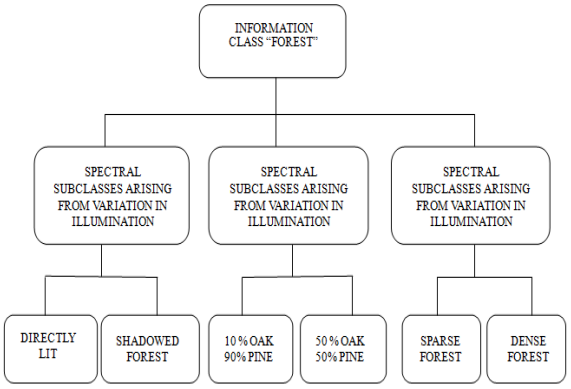

Figure 1.1: Example of Classification

Selecting the representative training sites. The presence of sufficient number of the training samples with their representativeness is critical for the classification of an image.

The training samples are mostly collected from the fieldwork or can be obtained from the fine spatial resolution of the aerial and satellite photographs. In case, landscape of the study area is somewhat complex and heterogeneous, then it becomes difficult to select sufficient training samples. This problem gets complicated when medium or coarse kind if spatial resolution data are applied for the classification, as they lead to the occurrence of a large amount of the mixed pixels.

Requires "training pixels", the pixels where there is knowledge of both the spectral principles and the class that is, the Land cover classes are defined in advance. There are sufficient reference data made available which are used as training samples. Thus, the signatures generated from training samples are then used to train the classifier to classify the spectral data into a thematic map as shown in the figure 1.2. Thus here we define information categories and then examine their spectral severability. Supervised classification requires prior information before testing process and it must collected by analyst. In this analyst identifies representative training sites for each informational class and also here algorithm generates decision boundaries.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

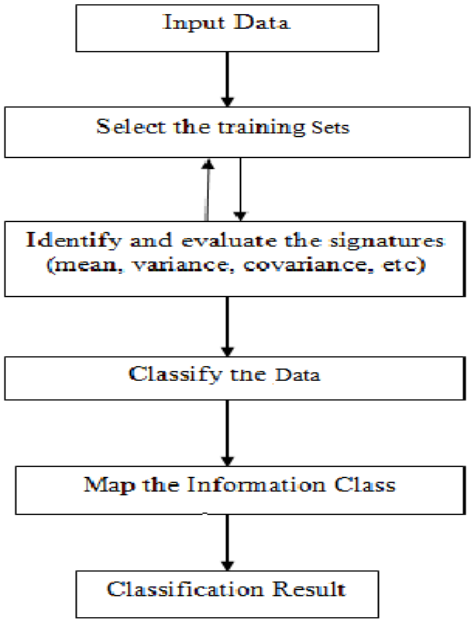

Figure 1.2: General Steps of Data Classification

II. RELATED WORK

Decision trees classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. The feature that best divides the training data would be the root node of the tree. There are numerous methods for finding the feature that best divides the training data such as information gain and gini index. While myopic measures estimate each attribute independently. However, a majority of studies have concluded that there is no single best method. Comparison of individual methods may still be important when deciding which metric should be used in a particular dataset. The same procedure is then repeated on each partition of the divided data, creating sub-trees until the training data is divided into subsets of the same class There are two common approaches that decision tree algorithms can use to avoid over fitting training data one is stop the training algorithm before it reaches a point at which it perfectly fits the training data and other is prune the induced decision tree.
Multi-layer neural network consists of large number of units (neurons) joined together in a pattern of connections. Units in a net are usually segregated into three classes: input units, which receive information to be processed, output units, where the results of the processing are found; and units in between known as hidden units during classification the signal at the input units propagates all the way through the net to determine the activation values at all the output units. Each input unit has an activation value that represents some feature external to the net. Then, every input unit sends its activation value to each of the hidden units to which it is connected. Each of these hidden units calculates its own activation value and this signal are then passed on to output units.

An RBF network is a three-layer feedback network, in which each hidden unit implements a radial activation function and each output unit implements a weighted sum of hidden units outputs. Its training procedure is usually divided into two stages. First, the centres and widths of the hidden layer are determined by clustering algorithms and second is the weights connecting the hidden layer with the output layer are determined by singular value decomposition (SVD) or least mean squared (LMS) algorithms. The problem of selecting the appropriate number of basis functions remains a critical issue for RBF networks.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

II.I Literature survey

Statistical approaches are characterized by having an explicit underlying probability model, which provides a probability that an instance belongs in each class, rather than simply a classification. Linear discriminant analysis (LDA) and the related Fisher's linear discriminant are simple methods used in statistics and machine learning to find the linear combination of features which best separate two or more classes of object. LDA works when the measurements made on each observation are continuous quantities. When dealing with categorical variables, the equivalent technique is discriminant correspondence analysis proposed by Mika in 1999. Maximum entropy is another general technique for estimating probability distributions from data. The overriding principle in maximum entropy is that when nothing is known, the distribution should be as uniform as possible, that is, have maximal entropy. Bayesian networks are the most well known representative of statistical learning algorithms.

Naive Bayes classifiers

Naive bayesian networks (NB) are very simple bayesian networks which are composed of directed acyclic graphs with only one parent and several children with a strong assumption of independence among child nodes in the context of their parent. The basic independent, the major advantage of the naive Bayes classifier is its short computational time for training [6]

Bayesian network (BN) is a graphical model for probability relationships among a set of variables (features) The Bayesian network structure S is a directed acyclic graph (DAG) and the nodes in S are in one-to-one correspondence with the features X. The arcs represent casual influences among the features while the lack of possible arcs in S encodes conditional independencies .Typically, the task of learning a Bayesian network can be divided into two subtasks: initially, the learning of the DAG structure of the network, and then the determination of its parameters.

Support vector machines (SVMs) are the newest supervised machine learning technique. SVMs revolve around the notion of a margin" either side of a hyper plane that separates two data classes. Maximizing the margin and thereby creating the largest possible distance between the separating hyper plane and the instances on either side of it has been proven to reduce an upper bound on the expected generalisation error the model complexity of an SVM is unaffected by the number of features encountered in the training data.

CART is introduced by Leo Breiman, Jerome Friedman, Richard Olsen and Charles Stone in 1984. It is a data mining decision tree classification algorithm. In CART algorithm following concept is simply used for making a decision tree. It is a classification method it uses historical data to construct decision trees. Decision tree is then used to classify new data. No of classes should know priori to perform classification [2]. A set of historical data with pre assigned classes for all observations. For example, learning sample for credit scoring system would be fu
ndamental information about previous variables matched with actual payoff results as classes [2].

Extraction of the land-cover information from the remotely sensed image is one of the most widely used applications concerning the field of remote sensing. Despite of great potential of the remote sensing method as a source of obtaining the land-cover information, various problems are encountered, like lower classification accuracy for users [1], [2], etc. Such problems have been attached with the research issues on the topics like the classification methods, the design of the sensor, the class definition, etc. Here, we have focused on the various supervised classification methods. Many parametric classifiers that are based on the statistical theory have also been successfully applied for the classification of remotely sensed images, like maximum likelihood, the minimum distance, the Mahalanobis distance, etc.

Artificial neural network (ANN) is also referred as simply "Neural Network" (NN) that is a process model supported as biological neural networks. They consist of an interconnected collection of artificial neurons. An artificial neural network is an adjective system that changes its structure supported information that flows through the artificial network during a learning section. The neural network classifier, which avoids several problems such as statistics algorithms and adopts a nonparametric approach, has been used in the field of remote sensing [14]-[17]. In NN it is tried to find the parameters which minimizes the mean square error prediction error with respect to a set of training examples.

The ANN is based on principle of learn by example. There are, however the 2 classical types of the neural networks, perceptron and also the multilayer perceptron, where we target to execute the perceptron algorithm. The basic concept of perceptron algorithm is to determine a linear function for the feature vector $f(x) = wT\ x + b$, resulting $f(x) > 0$ for the vectors of one category, and $f(x) < 0$ for the vectors of another. The vector $w = (w1, w2 \ldots wm)$ is the vector of the

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

coefficients (weights) for the function, and the supposed bias is b. If we denote the categories as +1 and -1, then we can state that it is required to find a decision function d(x) = sign (wT x + b). The algorithm of perceptron learning is completed iteratively. This starts with at randomly chosen parameters (w0, b0) for the decision and iteratively, updates them. In the nth iteration of the rule the training sample (x, c) is so chosen that the decision function is not able classify the sample properly (i.e. sign (wn x + bn)! = c).

PSO is also a population-based stochastic optimization technique and is well adapted to the optimization of nonlinear functions in multidimensional space. It modelsthe social behaviour of bird flocking or fish schooling. PSO has received significant interest from researchers studying in different research areas and has been successfully applied to several real-world problems [14].

The classical example of a swarm is bees' swarming around their hive but it can be extended to other systems with a similar architecture. Some approaches have been proposed to model the specific intelligent behaviours of honeybee swarms and they have been applied for solving combinatorial type problems [15–17]. Tereshko considered a bee colony as a dynamical system gathering information from an environment and adjusting its behaviour in accordance to it. Tereshko and Loengarov established a robot idea on foraging behaviour of bees. Usually, all these robots are physically and functionally identical, so that any robot can replace any other robot. The swarm possesses a significant tolerance; the failure of a single agent does not stop performance of the whole system. Like insects, the robots individually have limited capabilities and limited knowledge of the environment. On the other hand, the swarm develops collective intelligence. The experiments showed that insect like robots are successful in real robotic tasks.

III. PROPOSED SOLUTION

In the proposed algorithm, agents are considered as objects and their performance is measured by their masses. All these objects attract each other by the gravity force, and this force causes a global movement of all objects towards the objects with heavier masses. Hence, masses cooperate using a direct form of communication, through gravitational force. The heavy masses – which correspond to good solutions – move more slowly than lighter ones, this guarantees the exploitation step of the algorithm.

In GSA, each mass (agent) has four specifications: position, inertial mass, active gravitational mass, and passive gravitational mass. The position of the mass corresponds to a solution of the problem, and its gravitational and inertial masses are determined using a fitness function.

In other words, each mass presents a solution, and the algorithm is navigated by properly adjusting the gravitational and inertia masses. By lapse of time, we expect that masses be attracted by the heaviest mass. This mass will present an optimum solution in the search space.The GSA could be considered as an isolated system of masses. It is like a small artificial world of masses obeying the Newtonianlaws of gravitation and motion. More precisely, masses obey the following laws:

Law of gravity: each particle attracts every other particle and the gravitational force between two particles is directly proportional to the product of their masses and inversely proportional to the distance between them, R. We use here R instead of R2, because according to our experiment results, R provides better results than R2 in all experimental cases.

Law of motion: the current velocity of any mass is equal to the sum of the fraction of its previous velocity and the variation in the velocity. Variation in the velocity or acceleration of any mass is equal to the force acted on the system divided by mass of inertia.

In both GSA and PSO the optimization is obtained by agents movement in the search space, however the movement strategy is different. Some important differences are as follows:

– In PSO the direction of an agent is calculated using only two best positions, pbesti and gbest. But in GSA, the agent direction is calculated based on the overall force obtained by all other agents.

– In PSO, updating is performed without considering the quality of the solutions, and the fitness values are not important in the updating procedure while in GSA the force is proportional to the fitness value and so the agents see the search space around themselves in the influence of force.

– PSO uses a kind of memory for updating the velocity (due to pbesti and gbest). However, GSA is memory-less and only the current position of the agents plays a role in the updating procedure.

– In PSO, updating is performed without considering the distance between solutions while in GSA the force is reversely proportional to the distance between solutions.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

– Finally, note that the search ideas of these algorithms are different. PSO simulates the social behaviour of birds and GSA inspires by a physical phenomena.

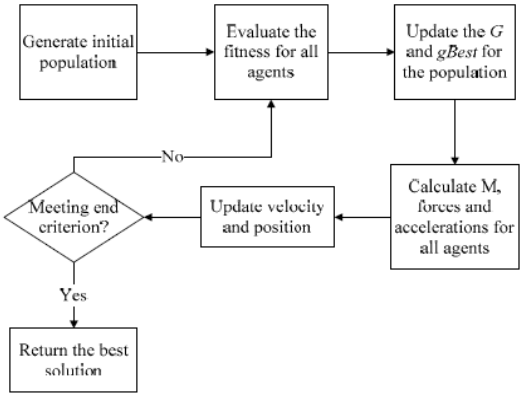

Figure1.3: Pictorial representation of evaluating steps

IV. EXPERIMENTAL ANALYSIS

In the research work, the performance is evaluated by data classification Accuracy measured. This accuracy assessment is done using the Iris Plants Database datasets in experiment 1 and randomly chosen user arbitrary chosen image dataset in Experiment 2 . The two experiments are performed for testing the performance of the proposed method. On the basis of the results the comparisons are made among the FNN-GSA ,FNN-PSO and FNNPSOGSA it is applied to standard benchmark functions . Iris Plants Database has been used for the result analysis which is updated on Sept 21 by C.Blake - Added discrepancy information. The database has been created by R.A. Fisher and donated by Michael Marshall in July, 1988 .This is perhaps the best known database to be found in the pattern recognition and classification. Fisher's paper is a classic in the field and is referenced frequently now days.

The data set contains 3 classes of 50 instances each, here each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other. Predicted attribute: class of iris plant,- This is an exceedingly simple domain. "Iris-setosa" where the error is in the fourth feature. The 38th sample: 4.9,3.6,1.4,0.1,"Iris-setosa" where the errors are in the second and third features. Number of Instances: 150 (50 in each of three classes) . Number of Attributes: 4 numeric, predictive attributes and the class .

The user and the producer accuracy are calculated for a particular class by dividing the number of classified data with the total number of data in that class. The results of the two experiments conducted on the two different datasets are analysed.

Table1: Result Comparison

Data Classification Methods	IRIS Dataset classification Efficiency (in %)	Glass Data Set Classification Efficiency (in %)	Random Data set Classification Efficiency (in %)
FNN-PSO	95.33	94.47	66.67
FNN-GSA	96	95.20	69.00
FNNPSO-GSA (Proposed method)	**98.66**	**97.02**	**71.00**

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

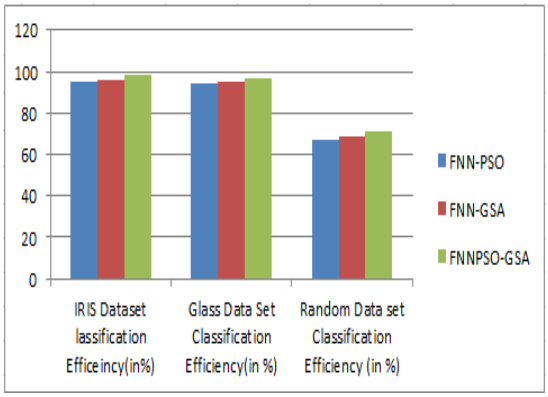

Figure4: Comparison of the accuracies for the proposed and the existing work.

V. CONCLUSION

As the results showed in chapter 5 the local search operator prop used as a kind of mutation did not improve very much the performance of the PSO and PSO-GSA algorithms. Although, we have so far not tested other local algorithms such as Levenberg-Marquardt and Back propagation, so we cannot state that the use of local operators in particle swarm optimizers is always of few use. Also, here applied some of the algorithms tested in this work to a more complete neural network optimization methodology based on the simultaneous adjustment of weigh and architectures of multi-layer perceptrons (MLP).

The work is effective and efficient; however, in the future work more exploration of different types of the classification methods that can be applied with the fuzzy topological space will be done. The different methods for the determination of the threshold value will be carried out. As future works, plan to improve the early stopping approach, still based on the GL5 stop criteria, to permit a more exploration of the search space for the PSO optimizers without loosing control of the generalization capability of the trained models with support vector machine.

REFERENCES

[1] A.E. Eiben and C.A. Schippers, "On evolutionary exploration and exploitation," Fundamenta Informaticate, vol. 35, no. 1-4, pp. 35-50, 1998.
[2] DH. Wolpert and WG. Macread, "No free lunch theorems for optimization," IEEE Transactions on Evolutionary Computation, vol. 1, no. 1, pp. 67-82, 1997.
[3] X. Lai and M. Zhang, "An efficient ensemble of GA and PSO for real function optimization," in 2nd IEEE International Conference on Computer Science and Information Technology, 2009, pp. 651-655.
[4] A. A. A. Esmin, G. Lambert-Torres, and G. B. Alvarenga, "Hybrid Evolutionary Algorithm Based on PSO and GA mutation," in proceeding of the Sixth International Conference on Hybrid Intelligent Systems (HIS 06), 2007, p. 57.
[5] L. Li, B. Xue, B. Niu, L. Tan, and J. Wang, "A Novel PSO-DE-Based Hybrid Algorithm for Global Optimization," in Lecture Notes in Computer Science.: Springer Berlin / Heidelberg, 2008, pp. 785-793.
[6] Elomma Lean Yu , and King Rousu , "Evolving Least Squares Support Vector Machines for Stock Market Trend Mining", IEEE transactions on evolutionary computation, vol 13 , IEEE, 2009,pp: 87-102
[7] N. Holden and AA Freitas, "A Hybrid PSO/ACO Algorithm for Discoverin Classification Rules in Data Mining," Journal of Artificial Evolution and Applications (JAEA), 2008.
[8] E. Rashedi, S. Nezamabadi, and S. Saryazdi, "GSA: A Gravitational Search Algorithm," Information Sciences, vol. 179, no. 13, pp. 2232- 2248, 2009.
[9] J. Kennedy and RC. Eberhart, "Particle swarm optimization," in Proceedings of IEEE international conference on neural networks, vol. 4, 1995, pp. 1942–1948.
[10] Y. Shi and R.C. Eberhart, "A modified Particle SwarmOptimiser," in IEEE International Conference on Evolutionary Computation, Anchorage, Alaska, 1998.
[11] Isaac Newton, In experimental philosophy particular propositions are inferred from the phenomena and afterwards rendered general by induction, 3rd ed.: Andrew Motte's English translation published, 1729, vol. 2,1984.
[12] E. G Talbi, "A Taxonomy of Hybrid Metaheuristic," Journal of Heuristics, vol. 8, no. 5, pp. 541-546, 2002.
[13] X. Yao, Y. Liu, and G. Lin, "Evolutionary programming made faster," IEEE Transactions on Evolutionary Computation, vol. 3, pp. 82-102, 1999.
[14] G.M. Foody, "Status of land cover classification accuracy assessment," Remote Sens. Environ., vol.80, no. 1, pp. 185-201, April 2002.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

[15]Maria C. Alonso, Jose A. Malpica, Alex Aggire "Consequences of the Hughes phenomenon on some classification techniques," ASPRS annual conference, may 1-5, 2011.
[16]Cilhar, J. Xiao, Q.Chen, J.Beaubien, J.Fung, K.and Latifovic, "Classification by progressive generalization: a new automated methodology for remote sensing multispectral data," International Journal of Remote Sensing, 19, pp. 2685-2704, 1998.

[17]R. S. De Fries and Jonathan Cheung-Wai Chan, "Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from Satellite Data," Remote Sens. Environ., vol. 74, pp. 503-515, Apr.2000.