Communications on Applied Nonlinear Analysis

ISSN: 1074-133X

Vol 33 No. 1s (2026)

Multiclass Classification Methods: A Review

Toshiba Mansoori, Prof.Anil Kumar Gupta

Computer Science and Applications Department

Barkatullah University, Bhopal

Computer Science and Applications Department

Barkatullah University, Bhopal

Article History:

Received: *29-01-2026*

Revised: *15-03-2026*

Accepted: *22-03-2026*

Abstract:

Artificial intelligence is playing vital role in various domains presently. Multiclass classification is an important task of AI. The concept of multiclass classification has been adopted in various domains such as medical science, banking sector, corporate, cyber security etc. Still researchers are applying such concept in different fields and trying to enhance the accuracy and efficiency of the models. The efficiency and accuracy of models are depend on various factors of such as dataset, selection of features, selection of algorithms, selection of hyper-parameters. The main objective of this paper is to present the various machine learning algorithms that are used for multiclass classification besides their merits and demerits. The present study also states the various research gaps in the existing multiclass classification models that are need to be resolved. This study will be useful to the researchers who are applying the multiclass classification in a specific domain to classify the data with great accuracy and efficiently. [The efficacy of models used for classification, recognition, diagnosis, or clustering of data of different domains is determined by their performance in real-world applications. Evaluating such models require detail understanding of their underlying equations and features to discern whether they perform "well" or "poorly." Such methods have become increasingly important to researchers in recent years. While a wide range of statistical methods has been applied, there remains a gap in guiding researchers toward the most appropriate method for their specific applications. This paper collects and analyzes the most significant classification methods used in research. It provides a comparative analysis of these methods, focusing on their mathematical foundations, purposes, and limitations. The study highlights that feature selection plays a significant role in classification performance, and the results provide guidance on selecting features for classification, recognition, diagnosis, or clustering tasks.]

Keyword*:* Data collection**,** Feature extraction; Classification, Classification, Recognition, Diagnosis, Clustering.

Introduction

Classification is a supervised machine learning technique in which an unseen data sample is assigned to a appropriate class [1]. The concept of multiclass classification is presented in the figure 1. Initially dataset is loaded into the memory. There are two methods of collecting data [2]: Quantitative data collection which is based on taking tools to collect structured data appropriate for the expertise of the categories of predefined responses and random sampling, which results in results that are easy to abstract, compare and published [2]. and Qualitative data collection which is a process plays a paramount role in effect assessment by providing useful data to understand methods based on observable outcomes and change assessment [3].

The researcher must register data systematically to ensure it is useful and accurate through observations, field notes, photography, and other appropriate means. Methods of collecting data must take into account research ethics. Several methods for collecting information for research projects include [2, 3]: Questionnaires, Portfolios, Direct observations, Inter views, Focus group interviews, Case-studies, Critical incidents, Diaries, Document and other materials.

Loaded data is required to preprocess first. Preprocessing is an essential task in data analysis. Various tasks such as find and replace missing values, normalize the data and identify the outlier are performed at this point.

Features selection is a way of opting the attributes that are useful in predictive analysis and leave the irrelevant features. It enhance the predictive accuracy. The goal of this study is to compare the performance of classification methods regarding feature selection, particularly when the size of training patterns is limited, and the number of features is significant. We consider various algorithms to identify common features that optimize classification performance.

Dataset is partitioned into two disjoint sets namely training dataset and test datasets. Since multiclass classification is a supervise machine learning techniques where each samples has class label. Multiclass classification based models are trained with the help of training dataset. Once the model is trained, it is evaluated with the help of test dataset. Model’s performance is evaluated on various predictive performance metrics.

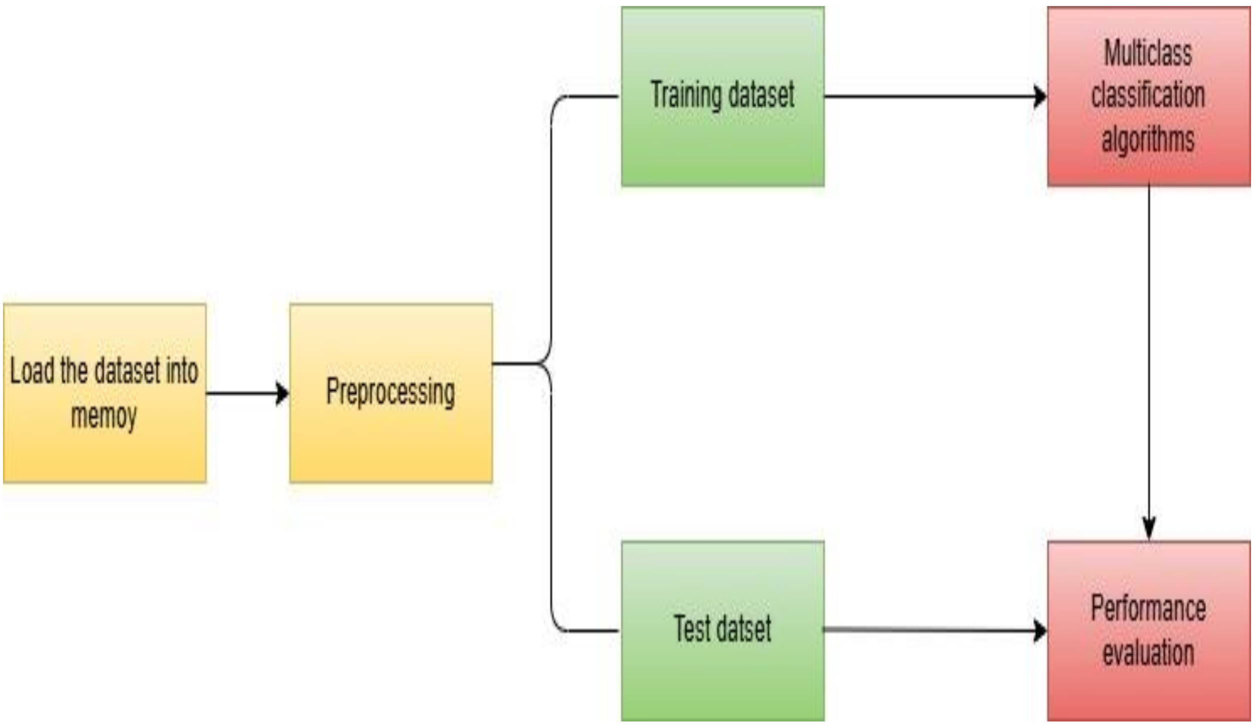

Figure1: Block diagram of Multiclass classification task

1. Multiclass classification methods

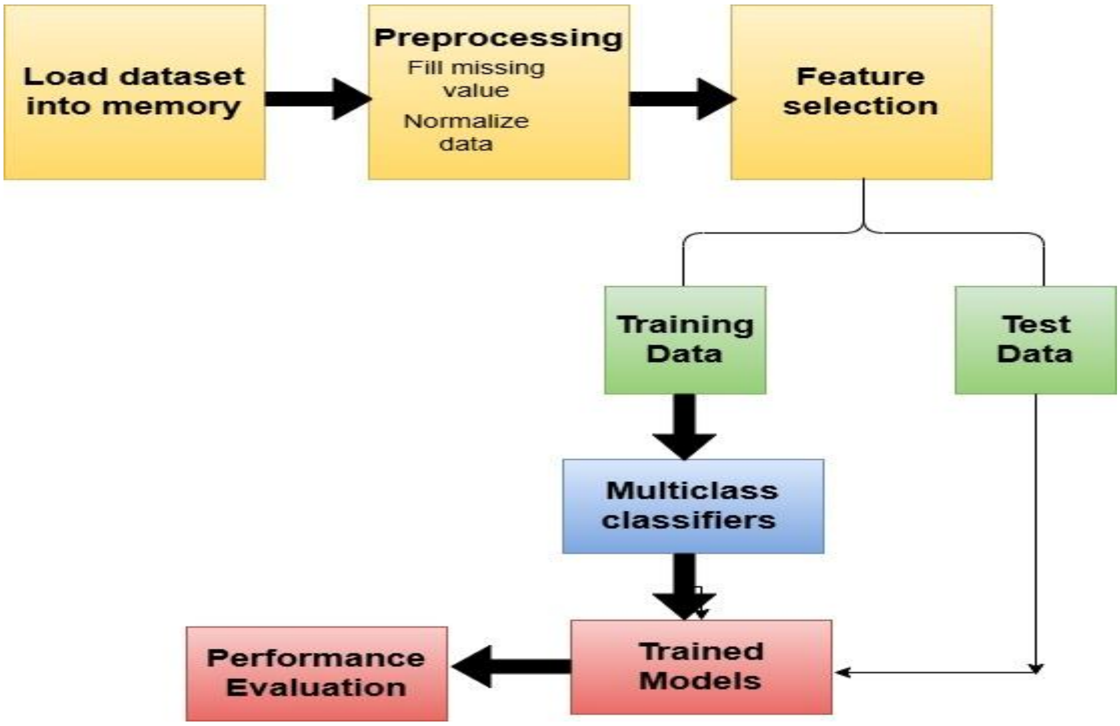

Figure1: Block diagram of Multiclass classification task

This section briefly describes the various multiclass classification techniques that have been applied in various domain to classify the unknown sample into the correct class.

A K-Nearest Neighbors (KNN) Classifier: This classifier has been shown to be efficient in action recognition studies. It selects the class relative to the instance under question using the Euclidean distance metric in the multidimensional feature space. Many studies utilize the KNN classifier because of its implementation simplicity, flexibility, and the fact that it allows for analysis of the classification resolution [4, 5]. KNN is one of the oldest non-parametric classification techniques. To classify an unknown example, the distance is measured (using measures such as Euclidean distance) to the training examples. The shortest distances of k are specified, and the most frequent category among these neighbors is assigned as the output class. The k value is typically determined using a validation set [6].

Support vector machines method: SVMs are among the most robust and effective classification techniques [7]. They are established based on the concept of maximizing the margin, i.e., maximizing the minimum distance from the separating hyperplane to the nearest example. The basic SVM supports only binary classification, but extensions [8] have been suggested to deal with multiclass classification states. In these extensions, further parameters and constraints are added to the optimization problem to handle the segmentation of various categories [7].

Decision Tree Method: Decision trees are a powerful classification method. They construct a binary classification tree where every node corresponds to a binary predicate on one feature; one branch corresponds to the positive cases of the predicate and another to the negative cases. Every node corresponds to a sequence of predicates and their values appearing on the descending path from the root. Every leaf is categorized by a class. To predict the class label of an input, a path to a leaf is traced based on the value of the predicate at every node visited. The predicates are selected by calculating the information gain of every feature, which is the expected decrease in entropy created by splitting the samples based on the feature [9].

Naive Bayes Method: Naive Bayes is an effective classifier depends on the standard of Maximum A Posteriori (MAP). Specified a trouble with K categories {C1, . . . , CK} with so- called previous probabilities P(C1), . . . , P(CK), we can specify the category label c to an obscure example with characters x = (x1, . . . , xN) like c = argmaxcP(C = ckx1, . . . , xN), that is select the category with the maximum a rear probability specific the observed information. This aposterior eventuality can be subedit, by Bayes algorithm, as follows: P(C = c\\x1, . . . , xN) = P (C=c)P (x1,...,xN\\C=c) /P (x1,...,xN) . Where the denominator is the same for whole categories, it can be discarded from the comparison. We will count the supposed category conditional probabilities of the characters specified the obtainable categories. This can be completely complicated taking into computation the dependencies among characters. The naive bayes method is to presume category conditional independence i.e. x1, . . . , xN are independent given the category. This facilitates the numerator to be P(C = c)P(x1\\C = c). . . P(xN \\C = c), and then selecting the category c that maximizes this value total the categories c = 1, . . . , K. Clearly this method is surely extensible to the state of having more than two categories, and was obvious to implement quite despite of the underlying simplifying suggestion of conditional independence [10].

C4.5 Classifier: This algorithm is an improvement of the ID3 algorithm developed by Ross Quinlan. C4.5 addresses both categorical and continuous attributes in decision tree construction. To handle continuous features, C4.5 splits attribute values into two sections based on a specified threshold. It also deals with missing feature values. C4.5 uses “gain ratio” rather than just information gain) as a measure to determine the best features to create a decision tree [11].

Genetic algorithm: Genetic algorithms aim to obtain the best appropriate parameters utilizing the technique of genetic evolution and survival of the fittest in natural selection. They permit the removal of false provisions in algorithms and improve the accuracy of text classification. This is a global algorithm for potential optimization, mimicking biological evolution. It is widely utilized for its simplicity and power. Currently, many researchers utilize this method to improve text classification [12].

Neural Network: The ANN, which is commonly called the neural network (NN), is a mathematical pattern or arithmetic pattern depends on the biological neural networks, in another word, a simulation of the biological nervous system. It comprise of a coherent set of artificial neurons processing data utilizing a contact method to calculate. In generality status, ANN is an adaptive method that converts its structure depends on the outside or inside data that streams over the network through the learning stage [13].

Fuzzy Logic: Fuzzy logic is a decision-making process used for expert systems and operational rules. Classical Boolean logic varies from Fuzzy logic in that Fuzzy logic allows partial membership in a group. Classical Boolean logic has only two values: a member either belongs or does not belong (1 or 0). Fuzzy logic allows partial membership, which may be any value between zero and one, representing a degree of membership [14].

K-Means Classifier: K-Means is a clustering algorithm (unsupervised learning technique) that automatically creates groups. Items possessing similar characteristics are placed in the same group. The algorithm is called k-means because it creates k distinct clusters. The average of the values in a certain group acts as the center (centroid) of that group [15].

Bayesian Classifier: Bayesian classifiers are statistical classifiers. They can predict the probability of class membership, such as the probability that a certain instance belongs to a particular category. The classification is based on Bayes' theorem. Bayesian classifiers have shown high accuracy and speed when applied to large databases [16].

ID3 Decision Tree Classifier: The ID3 algorithm uses one of the attributes to be the root of the tree and to create a sub node for each possible value of the attributes. If the node is not a sheet, the selection process is repeated and a sub-contract is created through the training data set, otherwise the process of splitting this section of the tree will end. When the decision tree is finished, a new test state is predicted by comparing the values of the new state attributes with the tree nodes starting from the main root down to the prediction value. The ID3 algorithm uses the Entropy code and wins the information gain to calculate and determine the most appropriate attribute of the division [17] : $Entropy(s) \; \square \; \square \; f_{\square} \; \log_2 \; f_{\square} \; \square \; f_{\square} \; \log_2 \; f_{\square}$

Entropy is calculated for each value of the values of the studied attribute:

Where:

S: Value of studied character

f +: Percentage of positive cases at the value of the studied character (number of positive cases / total number of cases)

f −: Number of negative cases at the value of the studied character (number of negative cases / total number of cases)

The general information gain (for the whole dataset) is calculated:

Info(main) = - f(+) / (f + m) * log2 (f(+)/ (f + m)) - f(-)/ (f+ m) *.Log2 (f(-)/(f+ m)

Information gain is calculated for each value of the studied attribute values:

Info(Attrib) = f(s) / (total)* Entropy(Attrib)

Where:

f: positive ratio ratio f

m: negative character ratio M S: Value of studied character is calculated Gain (Attrib) = Info (main) - info (Attrib)

, thus selecting the highest profit.

Linear Discriminant Analysis: LDA is a well-known method used to determine sample groups in a specific set of data. It attempts to separate groups (data classes) with a linear function so that the classes are as far apart as possible, while maintaining the distance between individual data samples within a single category as small as possible. The method assumes that the data in each class is normally distributed, but it is still applied successfully in many pattern recognition issues [18].

Quadratic Discriminant Analysis: As the name suggests, QDA is closely related to LDA. However, QDA does not assume that the covariance of each class is identical. Instead of a linear function separating the groups, QDA uses a quadratic function. It can be considered a generalization of LDA. Due to the more complex boundaries, it requires more computational time [19].

PCL Classifier**:** PCL is based on the idea of emerging patterns. It requires a feature selection process before the pattern is established. Selected features are discretized. Emerging patterns are then derived from the discretized training data. An emerging pattern is a set of conditions involving various features where most patterns of one class satisfy the conditions, but patterns of other classes do not. Thus, an emerging pattern acts as a multi-feature discriminator. PCL uses feature sets and does not assume feature independence; it can provide more than just a prediction, offering interpretable basics [20].

Comparative analysis of multiclass classification

In this section the Explained multiclass classification techniques have been compared and shows their suitability in classification task.

Table (1): Comparison of multiclass classification methods

Comparative analysis of multiclass classification Algorithms			
NO	Classification Algorithms	Positive aspect	Limitations
1	KNN	It is ease of understanding Training is so quick. 3. strong for noisy training data.	Slow Process Classification time is lengthy tricky to obtain the optimal value

2	SVM	1. most active ways in classification, especially common in text classification 2. contrast with ANN, it takes the basic characters of the data. 3. Gives great efficiency in classification	More difficult to classify 2. Tricky to explain for resolve, parameter sample. 3. There is a need for a number of key parameters to get the best rating result
3	Decision tree	1. The decision tree has a stellar velocity of learning and velocity of classification. 2. Backing transparence of knowing /classification. Backing multi- classification.	1. A small number of difference in the data can imply so various looking trees. 2. Development of a decision tree may do affect very for unnecessary attributes. ke predicaments of XOR, parity or multiplexer.
4	**Naive Bayes**	1. To improve the classification execution by eliminating the unrelated options. performance is good 3. The calculation time is short	1. it needs a great number of registers in order to get good results 2. Requires adjustment of its threshold values
5	**C4.5**	1. appropriat for real -world problems. address lost values 3. Divide the statement high accurately	More easy rules. 2. Requires high training samples 3.Unsatisfactory in practice
6	**Genetic algorithm**	1. it can get the best solutions in a quite short time 2. The random mutation It fairly guaranteed extent that we view a great range of solutions 3. Its coding is simple compared with other algorithms that perform the very process	1. The difficulty of reaching a good way of guidance reflects in fact the work of the algorithm 2. It may not find the optimal resolution for the issue specified in all status. 3. Difficult to choose parameters like population sizeو a number of generations, etc.

7	**ANN**	1. In complex the range, it supplies good result best for continued domain 3. The testing operation is speedy	Slow in the training process
8	**Fuzzy logic**	1. The notions of mathematics used in fuzzy reasoning are easy. Fuzzy logic notions are simple to understand. 2. Fuzzy logic is elastic for any system and is easy to gather more functions without starting from the start and easy management 3. Fuzzy logic afford inaccurate data. 4. Fuzzy logic may be merged with standard control ways. 5. Fuzzy logic is depended on natural language.	Fuzzy logic is a and an appropriate method to set the input range to the output range, but it may not be comfortable in all conditions. If there is a simpler solution, ambiguous logic may not be advisable
9	**K-mean**	1. Computational complexity is low Simple and easy	1. Depends on a large number of parameters 2. This way does not guarantee the best solution
		implementation 3.Dealing with a total of data on a large scale	3.Failure with sets nonlinear data
		1. Bayesian classifiers are efficient as decision trees and neural network classifiers. This Very high accuracy	1. Calculations and probabilities using a complex Bayes are often characterized by accuracy and caution must be taken to calculate them correctly The calculate can be NP-hard

10	**Bayesian**	shows a very high speed 4. The independence of the saved layer appears 5. Simplifies the arithmetic process	3. The quality of the results depends on the quality of the beliefs or the previous model 4. All sections should be calculated for the possibility of one branch calculate
11	**ID3**	1. It produces s a high precision algorithm of the C4.5 algorithm. 2. It gives omission rate decreased and false alarm rate , Increases detection rate and reduces space consumption 3. ID3 algorithm use nominal attributes for classification with no missing values	Search time is long 2. It takes a large amount of memory, which is bigger than the C4.5algorithm to execute a large program
12	**LDA**	It work for searches for all vectors in the underlying space and find the best discriminate among classes	omplex matrix
13	**QDA**	Close to LDA but no here is no assumption that the covariance of each of the classes is identical, more accurate.	mplex matrix
14	**PCL**	Used convert data from higher- dimensional into a lower-dimensional space	

Discussion

[**describe table here**]

Conclusion

Extensive review of literature revealed that multiclass classification can be useful tool in the present scenario for classification of various data belongs to different domain. Various machine learning algorithm have been applied along with hybrid models. Such models shows outstanding performance at the time of classification. However there are still various domains where the concept of multiclass classification may be applied to infer useful information that can be used for the betterment of society, organizations and environment.

References

1. J.Jelina, R.Sasikala, '' **Multi Class Classification Methods on Data Analysis Using Data Mining Techniques''**, IOSR Journal of Engineering (IOSRJEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719, PP 01-04,2018.
2. Male, T, ''**Collecting Quantitative Data**'', In I. Palaiologou, D. Needham and T. Male (Eds). Doing Research in Education: Theory and Practice. London: SAGE, 192-208.],2016.
3. Zubin Austin ,Jane Sutton, '' **Qualitative Research: Data Collection, Analysis, and Management** '', Can J Hosp Pharm , 68(3),pp:226–231,2015.
4. Preece, SJ, Goulermas, JY, Kenney, LPJ and Howard, D,'' **A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data''**, University of Salford,2009.
5. L. Bao and S. S. Intille, **"Activity recognition from user-annotated acceleration data",**
6. Pervasive Computing, Lecture Notes in Computer Science, vol. 3001, pp. 1-17, 2004.
7. U. Maurer, A. Rowe, A. Smailagic and D. Siewiorek, **"Location and activity recognition**
8. **using eWatch: A wearable sensor platform," , Lecture Notes in Computer Science, vol. 3864, pp. 86-102, 2006.**
9. Mohamed Aly,**'' Survey on Multiclass Classification Methods''**, November 2005.
10. Yoonkyung Lee, Yi Lin, and Grace Wahba.'' **Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data''.** Journal of the American Statistical Association, 99(465):67–81, March 2004.
11. Tao Li, Chengliang Zhang and Mitsunori Ogihara,' **''A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression''**, Advance Access publication April 15, 2004.
12. Irina Rish. , **''An empirical study of the naive bayes classifier''**, In IJCAI Workshop on Empirical Methods in Artificial Intelligence, 2001.
13. Ian H. Witten, Eibe Frank, & Mark A. Hall, "**Data Mining Practical Machine Learning Tools and Techniques**", Third Edition, Morgan Kaufmann, 2011.
14. Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee, Khairullah khan, **''A Review of Machine Learning Algorithms for Text- Documents Classification**'', JOURNAL OF ADVANCES IN INFORMATION TECHNOLOGY, VOL. 1, NO. 1, pp:4-20,FEBRUARY 2010.
15. S.Neelamegam, Dr.E.Ramaraj, **'' Classification algorithm in**

Data mining: An Overview'', International Journal of P2P Network Trends and Technology (IJPTT) , Volume 3, Issue 5 September to October 2013.

16. J.F. Baldwin, "**Fuzzy logic and fuzzy reasoning**" ,in Fuzzy Reasoning and Its Applications, E.H. Mamdani and B.R. Gaines (eds.), London: Academic Press, 1981.
17. Ayon Dey, " **Machine Learning Algorithms: A Review**", (IJCSIT) International Journal of Computer Science and InformationTechnologies, Vol. 7 (3) , pp: 1174-1179,2016.
18. Pang-Ning Tan, Michael Steinbach, & Vipin Kumar, "**Introduction to Data Mining**", Pearson Addison-Wesley, March 2006.
19. Yingying Wang , Yibin Li , Yong Song , Xuewen Rong and Shuaishuai Zhang **,** "**Improvement of ID3 Algorithm Based on Simplified Information Entropy and Coordination Degree**", Algorithms, 10, 124;doi:10.3390/a10040124 ,2017.
20. P DohnᏗlek, P Gajdoš, T Peterek and V SnᏗšel, "**An Overview of Classification Techniques for Human Activity Recognition**",VIBROENGINEERING PROCEDIA, VOLUME 2, ISSN 2345-0533 ,NOVEMBER 2013.
21. MQ Zhang , " **Identification of protein coding regions in the human genome by quadratic discriminant analysis**", Proceedings of the National Academy of Sciences ,94(2) ,PP:565–8**,**1997**.**
22. Huiqing Liu, Jinyan Li, Limsoon Wong,**'' A Comparative Study on Feature Selection and Classification Methods Using Gene Expression Profiles and Proteomic Patterns''**, Laboratories for Information Technology, 21 Heng Mui Keng Terr, 119613 Singapore, Genome Informatics 13: 51-60 ,2002