

# Electric Field Intensity for Nonlinear Classifier: A Novel Approach

Thiago R.F. Mendonça, Milena F. Pinto and André L.M. Marcato

Federal University of Juiz de Fora – UFJF  
Juiz de Fora, MG – Brazil  
thiago.mendonca@engenharia.ufjf.br

**Abstract**— The classification through specific characteristics is of great relevance for development of machine intelligence systems in order to improve decision making capability. Spite of the existence of several classifier algorithms, they all have its drawbacks and might not be suitable for a certain application. Within the supervised learning classifiers, the definition of some parameters is a common issue, such as the learning step, number of hidden neurons in a multilayer perceptron, the margin length for the support vector machines, the overfitting in nonlinear classifiers, among other situations that in some way depend on a subjective decision. In order to overcome these issues, it is proposed in the present work a novel approach for a supervised classifier, which is capable of working with linear and nonlinear situations by calculating a hyperplane based on electric field intensity. To validate the method's performance, a well-known classification problem found in pattern recognition literature is tested and compared with another classical method. The obtained results have proven its good performance.

**Keywords**— *Classification algorithms, Electrical Fields, Nonlinear classification, Pattern recognition.*

## I. INTRODUCTION

Pattern recognition algorithms are considered an important part in most machine intelligence systems developed for decision making based on features [1]. The feature extractions may be performed via several forms, such as image processing techniques, signal waveforms acquired from different sensors, among others.

According to [2], from discovering regularities in data by using computer algorithms and providing actions, it is possible to classify a determined data into different categories through pattern recognition process. The main interest in this area revolves in the emerging applications, i.e. personal identification based on physical characteristics, efficiently searching documents, databases, machine vision systems for automating process and it could also be applied to aid in fast diagnoses, classifying based on features extracted from blood samples or other medical exams. For instance in [3], it is discussed an optimal classifier for Electroencephalogram (EEG) waveforms based on patterns. It is noteworthy that the image and signal waveform analysis and its interpretation represent an important foothold in various scientific areas, with applications in materials science, robotics, biophysics and in biomedicine [4].

With the aim to determine classes by its set of properties, it is of great importance to choose the most suitable classifier algorithm for a certain application. Usually, the selection is based on the percentage of misclassifications, computational effort and time required for convergence of the method. It is also possible to optimally combine a number of algorithms to design novel approaches in order to work better for certain requirements.

It is worth to notice that many supervised methods for classification of patterns have some constraints. In the multilayer perceptron for instance, it is not easy to define the learning step and number of hidden neurons. In the support vector machines algorithm could be cited the margin length. Additionally, in general, there is the overtraining consideration. These definitions are in some way based on the user experience and therefore subjective, resulting in different hyperplanes for each user.

Furthermore, owing the considerable applications of classification algorithms and as an attempt to overcome the above constraints, it is proposed in this work a novel approach for a supervised classifier that is capable of working with linear and nonlinear situations, calculating a hyperplane based on electric field intensity. Therefore, it avoids the subjectivity in parameter definition as well as least computational effort after training stage. The proposed method is also proved to be more constant, independent of initial conditions.

This paper is organized as follows. Section II presents an overview concerning some well-known methods, describing its limitation and attractiveness. Section III discusses the proposed method's concept and how it overcomes the mentioned limitations. In Sections IV are presented the results and comparison with a feedforward backpropagation algorithm as well as other validation tests. Finally, it is presented in Section V the conclusion and suggestion for future works.

## II. RELATED WORKS AND BRIEF OVERVIEW

Due to the vast applicability of classification algorithms in the most diversified scientific areas, several methods and modifications were published with the aim to overcome some constraints and to provide better results in specific situations.

The Perceptron algorithm is one of the first classification methods presented in literature and was firstly proposed in [5].

Since this algorithm is unable to deal with non-linearly separable problems, the Multilayer Perceptron (MLP) arises as an improvement, using an artificial neural network based on the supervised learning approach. However, the major problem of these methods consists in choosing the optimal values for the learning and momentum rate, the number of layers and neurons used in the neural network.

Another well-known method is the clustering algorithm called K-means [6], which is responsible to cluster a set of instances based on their similarities, generating k-parameters. Thus, one problem concerning this method is the optimal selection of k, aiming for less effort.

The Genetic programming method, which is described in [7], is applied in several situations rather than the classification problem, but it can be optimized to generate a set of classification rules to a given dataset, using evolutionary algorithms to mix and mutate the existing candidates into a new rules generation. The main drawback of this method is the fitness function, which must be adapted to a determined problem to achieve optimal performance.

The Support Vector Machine algorithm works by calculating a higher dimension hyperplane to separate two classes [8], presenting a great generalization. However, requires a good kernel function to generate new dimensions based on the current dataset.

Although these mentioned methods are well established, their inherent limitations still motivates researches toward better algorithms. Nowadays, there are technologies and methods that may be combined and added in existing methods in order to enhance performance. For instance, in [9] is proposed the utilization of multi-objective genetic algorithm (MOGA) in machine learning to improve generalization, learning and optimization abilities. Also focusing on genetic algorithms, in [10] is introduced an approach for classifier and feature selection in a multi-classifier system. In [11], a novel classification algorithm is proposed for a specific application of classifying thermostable protein by using Hurst exponent and SVM. Aiming for improvement of voice based identification techniques, is discussed in [12] a fusion of classifier algorithms. Finally, in order to obtain a faster classification of high amount of spatial data, [13] presents a novel parallelized remote sensing classifier modification of SVM.

As it can be seen, this consolidates a trend in this research topic, where the search for an optimal classifier, which requires low computational effort and low convergence time, is pursued.

### III. CONCEPT OF THE PROPOSED APPROACH

The proposed method is a supervised algorithm, which requires a training stage in order to define a separation hyperplane that is used to classify future unknown samples. This hyperplane is based on the null equipotential value of an electric field generated from the training set. After this training stage, it is analyzed in which side of the equipotential lies the new sample.

The electric field intensity is calculated according to (1), where K is constant, Q is the charge and d is distance. By

considering each sample from one class to have a predetermined value of charge and the other class to have the opposite, then the combination of their electric field intensity disturbs the field in such a way that the equipotential represents a good separator.

$$E = \frac{K \cdot |Q|}{d^2} \quad (1)$$

Consider as an illustrative example, two classes with a Gaussian distribution in a two dimensional feature space. The equipotential lines of the electric field intensity of the overall system calculated by (1) are depicted in Fig. 1. The matrix containing the magnitude values of this field is used with the validation samples, comparing if it is in the positive or negative side.

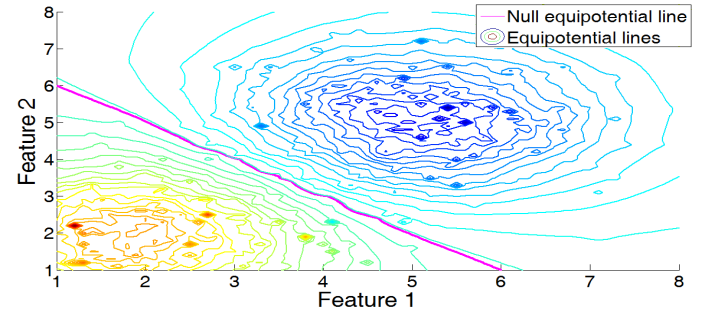


Fig. 1. Equipotential lines of the overall system containing two classes in a two dimensional feature space, highlighting the null equipotential line.

It is noteworthy that this method has a high computational effort in the training process, requiring the computation of the electric potential field distribution for each feature coordinate of the training set and adding them together, which results in the final matrix. The total electric field intensity distribution is used as a “lookup table” during the validation process. In other words, after the training phase, when the matrix is calculated, the new samples are classified just by referring to the lookup table, demanding, in this way, low computational burden.

The algorithm flowchart for the proposed method is presented in Fig. 2. It is noticed that in the training phase the class is known and therefore, it is determined a positive and negative charge for samples from class 1 and 2, respectively. The electric field intensity in the feature space is calculated for each sample of the training set by applying (1) and adding it to the total electrical field intensity of the system.

The dimension of the computed matrix is related to the number of features used to classify. For instance, in the case presented in Fig. 1, the classification is based on two features. Therefore, the lookup table, generated in the training phase will be two dimensional. If three features were used, then the matrix would be three dimensional and so on. The examples presented in this paper are limited to a maximum of three parameters situations just as a form to visually represent the samples, not being a constraint to the proposed method. It is noteworthy that as the number of features increases, the computational effort, required in the training phase, also

increases. However, the classification stage is not affected as much.

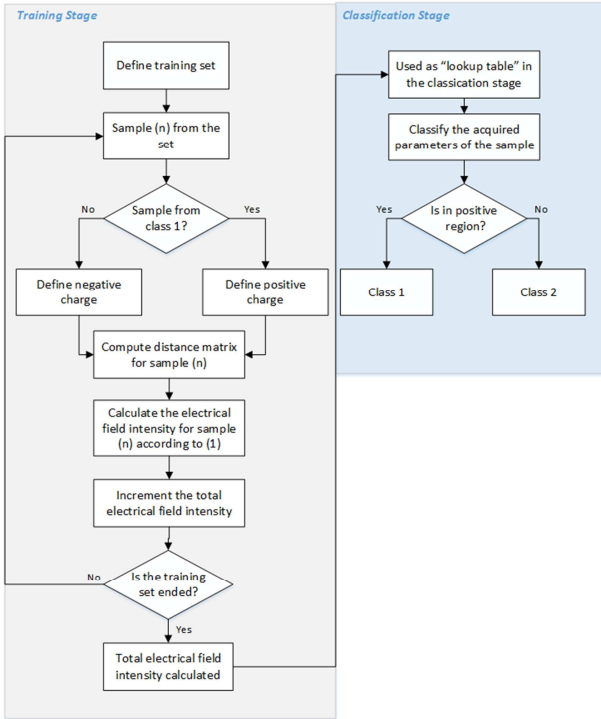


Fig. 2. Flowchart of the proposed method algorithm.

Another interesting fact about this method is the possibility of different decaying curves utilization. Instead of the inverse of distance, it could be used the Gaussian function, exponential, among others, resulting in hyperplanes that are optimal for each application. This is better discussed in Section IV, which shows results for two decaying curves.

The proposed method is more appropriate for classifying two classes, since it is trained based on positive or negative charges. However, its application can be expanded for multiclass problems by generating one hyperplane considering one class confronting the others. This is done for each class in order to obtain a hyperplane that separates one class from the remainder. Through a comparison process, it is checked in which hyperplane the sample is in the positive region. Fig. 3 depicts a classification process based on multiple separation curves. Note that there are regions outside of any positive side, where more misclassifications occur.

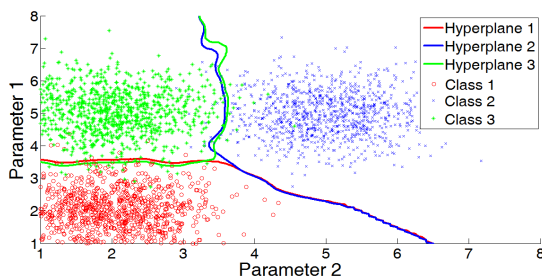


Fig. 3. Classification of multiclass problem based on multiple separation curves.

## IV. RESULTS

In order to validate the proposed method, a well-known dataset containing information for classifying types of plants based on some physical features is used. This dataset is found in [14] and includes three classes of plants, Setosa, Versicolor and Virginica. They are classified based on three to four parameters, Sepal length; sepal width; petal length and petal width. It is worth to notice that to visualize in a feature space, only three attributes were chosen. Fig. 4 depicts the distribution of the classes in a three dimensional feature space.

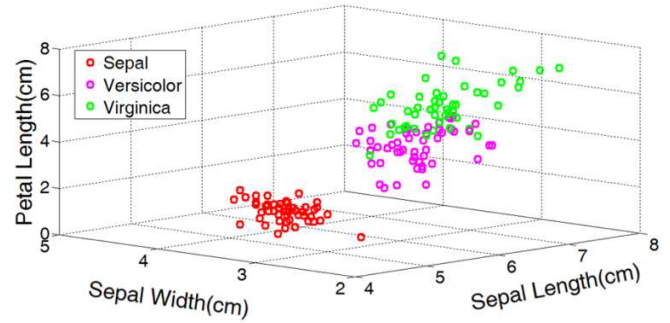


Fig. 4. Classification of three types of plants in the feature space.

Note that one class is linearly separable from the others. However, two classes are not linearly separable from each other. The dataset contains 150 samples, being 50 samples from each class. The procedure for training the classifier consists of selecting 40 random samples for each class in order to compute the hyperplanes. The 10 samples that remained are used to validate the training stage. After performing several iterations of training and classifications, with different sets of samples, the average error using the proposed method was zero, i.e. no misclassification occurred.

To compare with another method for nonlinear classification, it was performed the same number of iterations of the same procedure using the feed-forward backpropagation network method. It is defined a neural network with three inputs, ten hidden layers (chosen empirically based on the least error in classification) and one output ranging from 1 to 3. The training function used is the Levenberg-Marquardt backpropagation, which updates weights and bias values according to Levenberg-Marquardt optimization. Although it requires more memory than other methods, it is one of the fastest backpropagation algorithms and it is capable to solve nonlinear classification. With several tests using this method, usually 0 to 3 misclassifications occurred in the validation set. The time spent in the training phase was around 1.4 to 1.7 seconds with 13 to 19 iterations of backpropagation until convergence to optimal network. The misclassifications always occurred in the nonlinearly separable classes, i.e. Versicolor and Virginica types.

Instead of referring to a real problem for classification, it is considered fictional cases of separating two classes with AND, XOR and Gaussian distributions. The separation curves generated considering the proposed method for each distribution are presented in Fig. 5 though Fig. 8.

The case shown in Fig 5 represents a common case of two classes linearly separable. Class 1, represented in red, is generated with a Gaussian distribution with average at points [2, 2] and class 2, represented in blue, is generated also with a Gaussian distribution, but with average at [5, 5]. Both classes contain the same standard deviation equals to one. It is noteworthy that the separation curve is near to the optimal solution given by Bayes decision theory, however, without the need to know a priori the distribution function as required by the optimal solution.

All the classes in red outside the green shaded area are considered a misclassification of class 1, and all classes in blue inside the green shaded area are considered a misclassification of class 2. The total errors used to measure performance are the sum of misclassifications of both classes.

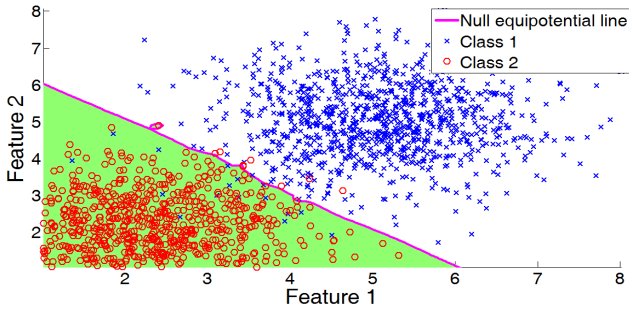


Fig 5. Fictional case considering two classes with Gaussian distribution in the feature space.

In the AND case presented in Fig. 6, the averages of class 1 and 2 are situated in such a way that resembles the logical “and” output. Note that the separation curve obtained by the calculation of a null equipotential line as proposed in this work, separates in two regions, achieving errors in the same proportion for both classes.

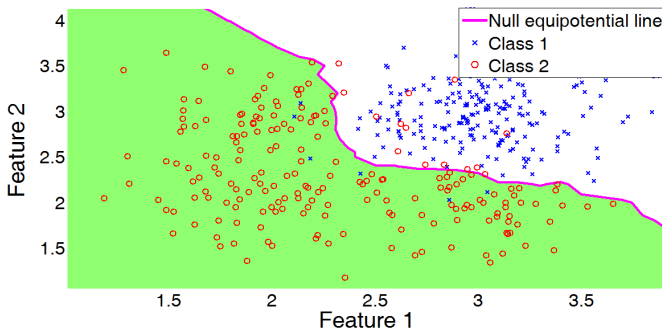


Fig 6. Fictional case considering two classes resembling a logical “and” output.

The same happens for the XOR case, presented at Fig 7, in which the localization of classes in the feature space is similar to the logical “exclusive or” output. This is a well-known and the simplest discussed example of a nonlinear separation. The separation curve calculated from the proposed method classifies based on quadrants, resulting in the same amount of misclassifications for all classes.

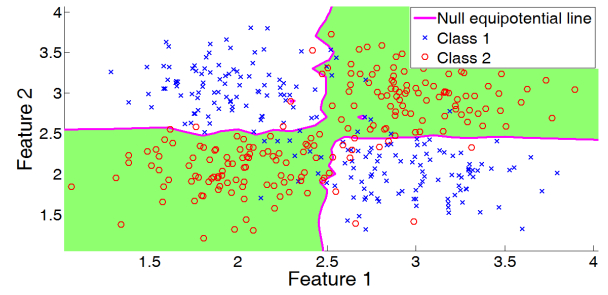


Fig 7. Fictional case considering two classes resembling the “exclusive or” output.

Another well discussed nonlinear classification problem is depicted in Fig. 8, where two classes with same average, but with different standard deviation, overlaps with one another. However, the separation curve obtained seems satisfactory.

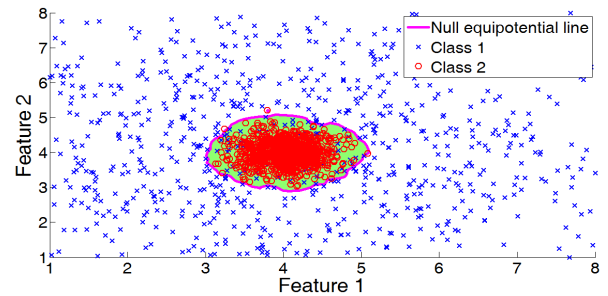


Fig 8. Fictional case of two classes with gaussian distribution overlapping with the same average but different standard deviation matrix.

The error rates of the proposed method for each case compared with the same feed-forward backpropagation network are presented in Table 1 through Table 4. For cases of Fig. 5, Fig. 6 and Fig. 7, the training set was generated with 1000 samples for each class considering a multivariate normal distribution with specific mean values for each case and the symmetric covariance matrix defined as  $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ . For the case presented in Fig. 8, the mean values were overlapped at [4, 4] and for each class, it was defined the symmetric covariance matrix as  $\begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \end{bmatrix}$  and  $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$  for classes 1 and 2, respectively.

From these comparisons, it is possible to conclude that the average error for the proposed method is slightly lower. However, the training time is always close to 1 second, independent of the separation case. The required time for the proposed method is proportional to the number of samples and the complexity of the decaying curve, since the calculation is performed for each training sample. This explains the reason the training time with the inverse of distance is lower than with Gaussian decaying curve.

TABLE 1. RESULTS FOR CLASSIFICATION OF CASE PRESENTED IN FIG. 5.

Case: Linearly separable Gaussian distribution		
Classifier	Error [%]	Training time [s]
Multilayer Perceptron	1.7375	1.9009
Proposed Method	Gaussian Curve	1.1325
	Inverse of distance	0.7117

TABLE 2. RESULTS FOR CLASSIFICATION OF CASE PRESENTED IN FIG. 6.

Case AND			
Classifier		Error [%]	Training time [s]
Multilayer Perceptron		6.6500	1.9314
Proposed Method	Gaussian Curve	6.5750	1.0718
	Inverse of distance	6.5250	0.7130

TABLE 3. RESULTS FOR CLASSIFICATION OF CASE PRESENTED IN FIG. 7.

Case XOR			
Classifier		Error [%]	Training time [s]
Multilayer Perceptron		10.1750	1.9219
Proposed Method	Gaussian Curve	10.1000	1.0353
	Inverse of distance	11.0750	0.7645

TABLE 4. RESULTS FOR CLASSIFICATION OF CASE PRESENTED IN FIG. 8.

Case: Non linearly separable Gaussian distribution			
Classifier		Error [%]	Training time [s]
Multilayer Perceptron		5.7125	2.8037
Proposed Method	Gaussian Curve	4.8500	1.0096
	Inverse of distance	5.2125	0.7084

## V. CONCLUSIONS

In the present work is discussed the importance of a reliable and fast classifier for applications in several research fields. A brief overview of some commonly used classifier methods is outlined. It is highlighted their advantages and drawbacks as well as how other works have proposed new approaches to overcome nonlinear classifier's limitations and difficulties imposed in the training stage. Following this research trend, it is proposed in this work a novel approach for a supervised classifier, which is capable of working with linear and nonlinear situations by calculating a hyperplane based on electric field intensity. Therefore, it only requires the definition of charges magnitudes for each class, removing the user's responsibility in defining training constraints. The results have shown good performance for both training time and misclassification percentage.

For future works, it is intended to study the performance for other decaying curves and evaluate for each scenario if it will have good performance. It is also intended to embed this algorithm in hardware to validate in a more practical application the proposed classifier.

## ACKNOWLEDGMENT

The authors would like to thank the Federal University of Juiz de Fora, CAPES, FAPEMIG, CNPq, INERGE and the Brazilian Government for all financial, technical and support.

## REFERENCES

- [1] Theodoridis, S. and Koutroumbas, K. "Pattern Recognition," Second Edition. Elsevier, 2003.
- [2] Bishop, C.M. "Pattern Recognition and Machine Learning", Springer, 2006.

- [3] Lotte, F., Congedo, M., L'ecuyer, A. and Lamarche, F. "A review of classification algorithms for EEG-based brain-computer interfaces," Journal of Neural Engineering, Institute of Physics: Hybrid Open Access, 2007.
- [4] Anil K. J., Robert P.W. D. and Jianchang M. "Statistical Pattern Recognition: A Review," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, no. 1, 2000.
- [5] Rosenblatt, F. "The perceptron, a perceiving and recognizing automaton" Project Para. Cornell Aeronautical Laboratory, 1957.
- [6] Macqueen, J. et al. "Some methods for classification and analysis of multivariate observations". In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1967. p. 281-297.
- [7] Koza, J. "Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems". Stanford University, Department of Computer Science, 1990.
- [8] Boser, B., Guyon, I. and Vapnik, V. "A training algorithm for optimal margin classifiers". In: Proceedings of the fifth annual workshop on Computational learning theory. ACM, 1992. p. 144-152.
- [9] Litaio, Z., Tiejun, W., Xi, J. and Jin, J. "The Machine Learning Classifier based on Multi-Objective Genetic Algorithm" 7th International Conference on Computing and Convergence Technology (ICCT), 2012.
- [10] Nguyen, T. et al. "A novel genetic algorithm approach for simultaneous feature and classifier selection in multi classifier system", IEEE Congress on Evolutionary Computation (CEC), 2014.
- [11] Liu, H. et al. "A Novel Classification Algorithm of Thermostable Proteins by Using Hurst Exponent and SVM Classifier", Fourth International Conference on Natural Computation, ICNC 2008.
- [12] Deriche, M. A. and Naseem, I. A. "A novel speaker identification algorithm using classifiers fusion", 10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA), 2010.
- [13] Pan, X. and Zhang, S. "A Novel Parallelized Remote Sensing Image SVMClassifier Algorithm", 5th International Congress on Image and Signal Processing (CISP), 2012.
- [14] Bache, K. and Lichman, M. "UCI Machine Learning Repository", Irvine, CA: University of California, School of Information and Computer Science.