

# Feature Selection for Improvement the Performance of an Electric Arc Furnace

Amado Sánchez Sánchez<sup>1</sup>, José Crispín Hernández Hernández<sup>1</sup>, Haydee Patricia Martínez Hernández<sup>1</sup>, David Ibarra Guzmán<sup>2</sup>, Arturo Contreras Juárez<sup>2</sup>, Arturo Aguila Flores<sup>2</sup>, and Perfecto Malaqías Quintero Flores<sup>1</sup>

<sup>1</sup>Instituto Tecnológico de Apizaco,  
Laboratorio de Investigación en Tecnologías Inteligentes, Apizaco, Tlaxcala,  
Mexico

<sup>2</sup>Universidad Politécnica de Tlaxcala Región Poniente,  
Hueyotlipan, Tlaxcala,  
Mexico

amado.sanchez1982@gmail.com, josechh@yahoo.com

**Abstract.** Feature selection has as principal goal to find a representative space of minimal size from original set of larger size. Several research works have been developed on this problem. This paper presents Support Vector Machine-Recursive Feature Elimination (SVM-RFE), Genetic Algorithms (GA), and Differential Evolution (DE) algorithms for feature selection from a database of an Electric Arc Furnace (EAF) for locating variables related to energy consumption. The proposal suggests merging the coefficients generated by LDA and SVM, employing them in RFE to obtain the ranks for each discriminant variables. The measure of accuracy and error rate for each algorithm is presented like a decisive score for choosing the subset obtained by the algorithm with the best performance. The variables selected were adjusted for the EAF control system achieving the reduction of the energy consumption to 3.5% in a steel castings and 1 minute reduction of the connected EAF in a steel casting.

**Keywords:** Feature selection, electric arc furnace, DEFS, GA, SVM-RFE, LDA, energy consumption

## 1 Introduction

The impact of feature selection methods applied to data analysis process has been exposed on several research such as medical diagnosis, image recognition, credit to a bank, text classification, where the results allow given a dataset described by  $N$  features to find the minimum number  $n$  of relevant features for describing the data as well as the original set of features [5, 3].

In this paper, the application of feature selection algorithms on data of an Electric Arc Furnace (EAF) is presented in order to identify the variables that

discriminate the merger process and adjust them to the melting process for improvement of the performance in steel casting.

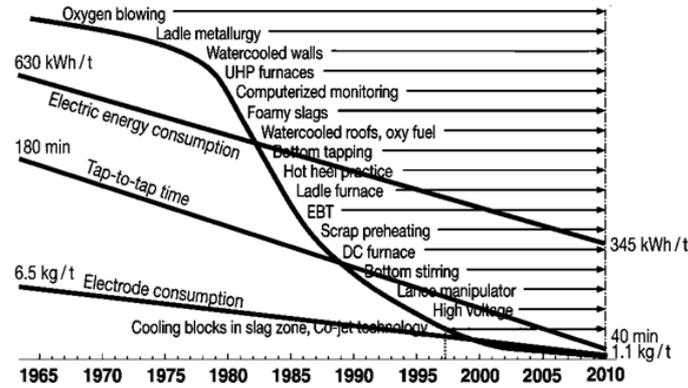
The steel melting process in electric arc furnace is too complex and its nature consists of  $N$  number of variables that require a precise control to achieve high performance. Although electrical energy is the principal resource in EAF, additional chemical energy is supplied; this is added by injecting oxygen, natural gas and graphite in the EAF through multiple injection units adapted to the wall to achieve thermal homogeneity within the furnace. The release of heat generated by the chemical energy ( $CO + CO_2 + \Delta E$ ) is exploited to get to the parts where the arc is less (cold spots).

The electricity needed to generate the arc represents a high cost for the steel-making process, that is, the 80% of the steel mill consumption total. Therefore decreasing by at least 1% of the electricity consumption, the involved cost would be positively impacted.

## **2 Related Works for Improvement of the Performance of EAF**

Some soft computing methods, focused on industrial and biological processes have been applied in EAF. Bernardo et al [2], propose the application of multi-objective algorithm for controlling temperature in a steelmaking process. Erik et al [13], propose using multivariate prediction models (Principal Component Analysis, Multiple Linear Regression, Partial Least Squares), to estimate the properties of scrap and achieving adequate chemical conditions of steel. Miroslaw et al [10], propose a combination of a regression tree with a neural network to optimize the input data, and the use of evolutionary algorithms to find optimal weights to be used in the neural network; both were to predict the temperature of an EAF. Khan et al [8], they offer excellent proposal for the analysis of information from a EAF using data mining in order to find rules that could be used for the development of automated intelligent systems, at the same time they introduced the concept of chemical energy and perform an analysis on the behavior of carbon, gas and oxygen through the development of an energy balance model, achieving at various stages of the merger process, determine the average gas required into the EAF.

Each work has contributed through the passage of time to improve the performance of the EAF's as shown in figure 1.



**Fig. 1.** Several innovations have been introduced on EAF to improve its performance. (Extracted from [8])

The figure 1 depicts the performance of EAFs. During previous years, different innovations were introduced to EAFs. These technological developments caused significant improvements in different key performance indicators.

### 3 Methods for Feature Selection Used

There are two main approaches to data reduction: feature transformation and features selection. The feature selection algorithms are organized into three main categories: filter methods, envolventes methods, and embedded methods [9].

The first method, select the features without running any learning algorithm, they are evaluated through the intrinsic properties of the data. In envolventes methods, learning algorithms are used to evaluate the selected subset, based on learning classifier for each of the candidates of the subset. Embedded methods incorporate learning algorithms and search features in an optimization problem. Feature selection known as the selection of variables, is the problem of selecting a subset of the original features, in contrast to methods based on the transformation wich, allowing the modification of the input characteristics to a new feature space; in features selection, the original representation variables do not change [9, 6, 11].

According with the literature, the most popular methods for feature selection are: SVM-RFE, GA, DE and LDA for reduction data.

#### 3.1 SVM-RFE Algorithm

The *SVM – RFE* algorithm is a technique that combines Support Vector Machine and Recursive Feature Elimination, methods developed by V. Vapnik et al, and I. Guyon et al, respectively. These predictive models have the capacity

to generate knowledge for the interpretation of the obtained solution [16, 7]. The general esqueme of the algorithm to SVM-RFE is illustrated in figure 2.

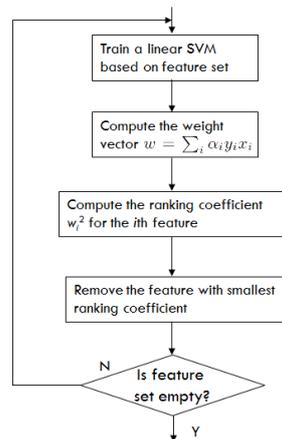


Fig. 2. SVM-RFE general scheme.

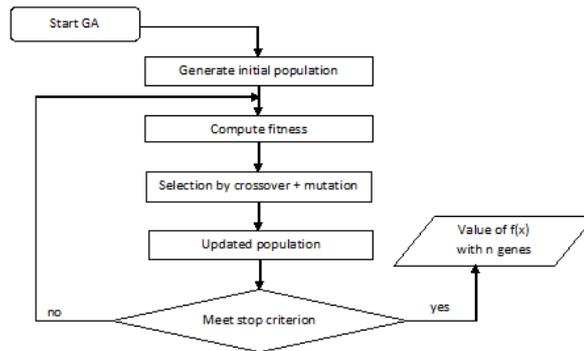
Where  $x$  is the feature expression vector of a sample,  $x_i$  is that of sample  $i$  in the training set ( $i = 1, 2, \dots, n$ ),  $y_i \in \{ +1, -1 \}$  is its corresponding class label,  $w = \sum_{i=1}^n \alpha_i y_i x_i$ , is the vector of weights of the features, and  $b$  is a scalar offset. The  $\alpha_i$ 's and  $b$  are estimated from the training set. Only those samples closest to the separating boundary (called support vectors). The criterion ( $w_i$ ), estimates the effect of eliminating a feature in an objective function, but it becomes a suboptimal procedure when it decides to eliminate many features. Performing these reductions are necessities in some cases for obtain a small subset of features. For more details see [7].

### 3.2 Genetic Algorithms

Genetic algorithms (GA) are adaptive search techniques, based on the analogy with biology, in which a set of possible solutions evolves via natural selection. Genetic Algorithms were introduced by John Holland in 1975, based on natural selection proposed by Charles Darwin; ie, they mimic the mechanisms of gene duplication and natural selection. In nature individuals compete among themselves to achieve survival, genes of the fittest individuals are propagated to subsequent generations, including on occasion the children adapt better parents. So species evolve generation after generation. The set of all genes are encoded in a string of values, called chromosome. In the early works of John Hollan coding is performed with a string of zeros and ones. And in fact currently representations

are coded with an integer, real or float value which allows the development of genetic operators can be made more specific [3, 12, 18].

The general scheme of an simple genetic algorithm is depict in the figure 3:



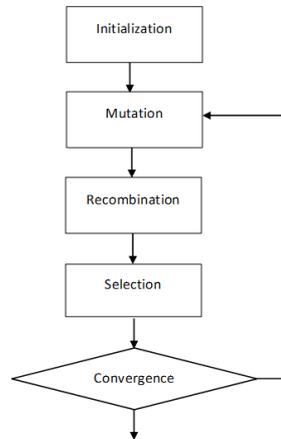
**Fig. 3.** General scheme of Genetic Algorithm

In carrying out the genetic algorithm for selecting features, it must; firstly, ensure that the strings are randomly generated to build the initial population. Each string represents a subset of features and values in every position in the string are coded as either the presence or absence of a particular feature. Then fitness must be calculated, which is a measure indicating how well a set of characteristics survives evaluation criteria specified for each subset. Best feature subsets are more likely to be selected to form a new subset through a crossing or mutation. The mutation changes some values (by adding or removing features) in a random subset. Crosses made to the different characteristics of a pair of subassemblies are combined into a new subset. This feature selection based on genetic algorithm is an iterative process in which; each successive generation is generated by applying genetic operators for members of the current generation. Thus, good subsets are evolved over time until stopping criteria are met.

### 3.3 Differential Evolution

Differential evolution (DE) is a simple optimization method that has parallel, direct search, easy to use, good convergence, and fast implementation properties. The first step in the DE optimization method is to generate a population of NP members each of D-dimensional real-valued parameters, where NP is the population size, and D represents the number of parameters to be optimized [15].

The general squeme is depict in the figure 4.



**Fig. 4.** General esqueme of Differential Evolution

Feature selection with differential evolution (DEFS) is rooted in the general DE algorithm . The main goal of DEFS is to select a subset of features with a low rate of misclassification or high precision in the classification [17]. Therefore the error rate is used as a fitness function; ie, a minimization function:

$$Fit = errorRate = \frac{FP + FN}{TP + TN + FP + FN} \quad (1)$$

where FP, FN, TP and TN are the measurements for false positives, false negative, true positives and true negatives, respectively.

Feature selection plays a central role in the data analysis process since irrelevant features often degrade the performance of algorithms devoted to data characterization, rule extraction and construction of predictive models, both in speed and in predictive accuracy.

## 4 Methodology Proposal

The propose is a method to reduce the initial dimension of EAF database to select a subset of discriminant variables related with the energy consumption. The model involves the SVM-RFE that guides the feature elimination process. The proposal suggested to merge the coefficients generated by LDA and SVM and employ them in RFE to obtain the ranks for each discriminant variables in the database EAF.

### 4.1 Linear Discriminant Analysis

LDA is one of the most commonly used technique for data classification and dimension reduction [1].

LDA considers maximizing the following objective:

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \quad (2)$$

where  $S_B$  is the between classes scatter matrix and  $S_W$  is the within classes scatter matrix. Due to the fact that scatter matrices are proportional to the covariance matrices,  $J$  is defined using covariance matrices. The definitions of the scatter matrices are:

$$S_B = \sum_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T \quad (3)$$

$$S_W = \sum_c \sum_{i \in c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (4)$$

where  $\bar{x}$  is the overall mean of the data cases,  $c$  represent the classes,  $S_B$  is the scatter of class 1 with respect the scatter of class 2 and hence corresponds to computing the scatter relative to a different vector. An important property to notice about the objective  $J$  is that is is invariant w.r.t. rescalings of the vectors  $w \rightarrow \alpha w$ . Hence, is possible choose  $w$  such that the denominator is simply  $w^T S_W w = 1$ , since it is a scalar itself [4, 14].

The coefficients of the eigen vector calculated by LDA are used to evaluate the relevancy of each feature for class discrimination. LDA obtains the projection vector  $w$ , wich is used in RFE algorithm.

#### 4.2 SVM Weight Vector for Classification

Support vector machines are an example of a linear two-class classifier. The data for a two-class learning problem consist of objects labeled with one of two labels corresponding to the two classes; for convenience we assume the labels are +1 (positive examples) or -1 (negative examples).

A concept required for defining a linear classifier is the dot product between two vectors, defined as  $w^T = \sum_i w_i \cdot x_i$ . A linear classifier is based on a linear discriminant function of the form:

$$f(x) = w^T x + b \quad (5)$$

The vector  $w$  is known as the weight vector, and  $b$  is called the bias. Consider the case  $b = 0$  first. The set of points  $x$  such that  $w^T x = 0$ , are all points that are perpendicular to  $w$  and go through the origin to a line in two dimensions, a plane in three dimensions, and more generally, a hyperplane. The bias  $b$  translates the hyperplane away from the origin. The hyperplane divides the space into two according to the sign of the discriminant function  $f(x)$  defined in Equation (5). This weighth vector  $w$ , guide the RFE method to remove the most discriminant features [16, 7].

### 4.3 LDA\_SVM-RFE

LDA and SVM are proposed to train and to obtain the vector weights and eigen value of each feature by removing one by one with the smallest weight.

In each iteration is used a validation method to train SVM and LDA classifiers to calculate the eigen value vector definite as  $w_1$  and weights vector definite as  $w_2$  of each feature of the data, both are averaged to obtain a single vector  $w$ . Thus, RFE find a final subset that contains the most discriminant variables related with the energy consumption. The method is described in the figure 5:

---

**Pseudocode 1: LDA\_SVM-RFE**

---

Given the original set with respective labels, set  $X=\{x_1y_1, x_2y_2, \dots, x_ny_n\}$

Train LDA and SVM classifiers

Obtain discriminant coefficients of each feature from  
LDA and SVM classifier

average the vector  $w_1$  from LDA an vector  $w_2$  from SVM  
to obtain a single vector  $w$

validate the new  $w$  vector en RFF method

Compute the ranking criteria

Find the feature with smallest ranking criterion

Update feature ranked list

Eliminate the feature with smallest ranking criterion

Feature ranked list  $r$

---

**Fig. 5.** Pseudocode for feature selection with LDA\_SVM-RFE.

This proposal is based on the SVM-RFE strategy proposed by Guyon et al. [7], where each feature is evaluated for its corresponding coefficient.

## 5 Experimental Results

This model was evaluated with EAF data, previously these were treated and normalized using the minimum and maximum expression values of each feature given as allows  $y_i = x_i - \min(x_i) / (\max(x_i) - \min(x_i))$ .

The performance of the algorithms is evaluated by 10-FOLD Cross Validation. The algorithms were coded on Matlab using a laptop intel® core™ i5-2540M CPU @ 2.60GHz y 4.00 GB en RAM.

The total data is 1254 examples with 75 variables. In the table 1 is shown the accuracy and error rate for each algorithm used.

**Table 1.** Comparison of the algorithms.

Algorithm	Accuracy (%)	Error Rate (%)	Std. Des.
DEFS	67.69	32.31	0.0729
GA	65.27	43.73	0.0380
SVM-RFE	84.31	15.69	0.0019
LDA_SVM-RFE	84.37	15.63	0.0043

We observe that we model obtain te best accuracy respect DEFS, GA and SVM-RFE. GA has the worse accuracy, barely 65.69% following by DEFS with 67.69%. SVM-RFE is very close with 84.31% and LDA\_SVM-RFE offer an accuracy rate of 84.37%.

In this practice, all variables of the EAF were used to confirm the performance, given the dimension of the data. In the table 2 the top 5 ranking variables generated for each algorithm are listed.

**Table 2.** Top 5 ranking variables determined by the runs algorithms

Algorithm	Top 5 ranking variables
DEFS	(1) Kilowatts hour per charge metal tons. (2) Amount of $O_2$ injected in Melting 1. (3) Time On of EAF during steel casting. (4) °C/minute in heating step. (5) Mega Watts in stage flat bath.
GA	(1) Time On of EAF during steel casting. (2) Amount of graphite ( $kg$ ) injected in Melting 2. (3) Amount of oxygen ( $m^3$ ) injected in Melting 2. (4) Tons of scrap charged to the EAF, to the Melting 2. (5) °C/minute in heating step.
SVM-RFE	(1) Ratio $C_{Tot}$ , $O_{2Tot}$ , $CH_{4Tot}$ throughout the steel casting. (2) Amount of graphite ( $kg$ ) injected in Melting 2. (3) Amount of oxygen ( $m^3$ ) injected in Melting 2. (4) Tons of scrap charged to the EAF, to the Melting 2. (5) °C/minute in heating step.
LDA_SVM-RFE	(1) Ratio $C_{Tot}$ , $O_{2Tot}$ , $CH_{4Tot}$ throughout the steel casting. (2) Ratio $O_2$ , $CH_4$ in Melting 1. (3) Amount of $CH_4$ injected in Melting 1. (4) Amount of $O_2$ injected in Melting 1. (5) $O_2$ required to burn the $CH_4$ effectively.

Although the performance of each algorithm was different, they share some common variables. With this information, the flow injection of the chemical energy in the EAF were adjusted so as to create conditions for effective combustion reactions, which allows the system to obtain thermal energy instead of losing by excess or lack of the elements reagents involved in the formation of chemical energy and chiefly in the formation of foamy slag.

### 5.1 Test Results

From previous information was determined the adjuste for the elements on the EAF system control. The adjustment was executed supported by the process melting team. This new flow allow chemical balance in the combustion reaction. That chemical energy helps to the electric energy supply by the transformer through the electrodes. The figure 6, depict the adjuste for the flows at the chemical elements.

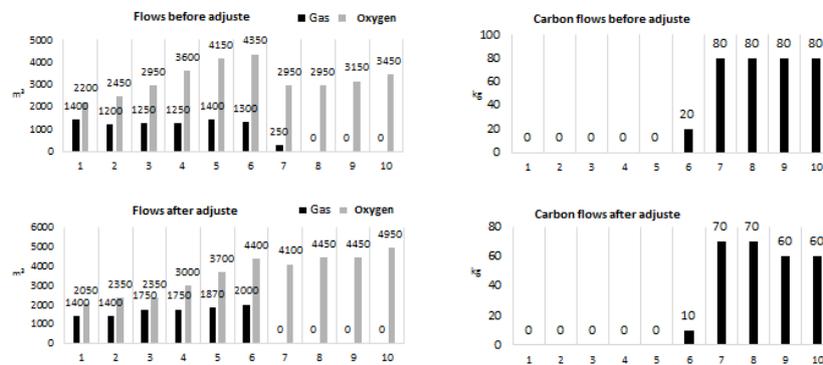
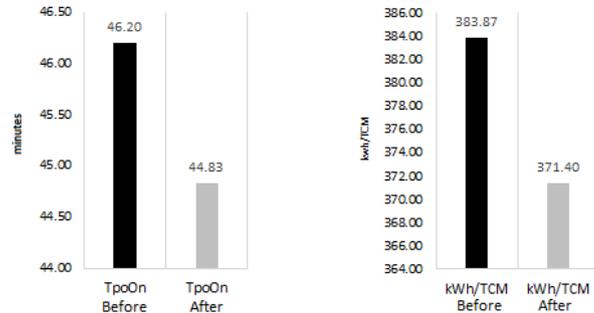


Fig. 6. Representative graphics for the adjusted on EAF system control

According with the graphics, we can see that the flows before the adjuste have irregular behavior since the phase 6, where the chemical energy is not constant. The flows adjusted claim supply the oxygen, natural gas and carbon in progressive form for generate the fusion after the heating of the scrap, this condition allow that the chemical energy stay agressive on the last steps. After the adjuste on the EAF control system were accused steel casting tests. The behavior for the kWh and time on for the EAF are presented in the figure7.



**Fig. 7.** Results obtained for the electric energy consumption (kWh/TCM) and the time connecting of the EAF(TpoOn)

The figure 7 depicts the results back to the settings made in the control system of EAF, where the predicted was decreased by at least 1% of electricity consumption and reduce time connected the EAF at least 1 minute. Achieving a savings of 3.5% in consumption and one minutes less on average than TpoOn.

## 6 Conclusions

In this paper, some algorithms for feature selection and the performance of each were presented. Similarly, a new it uses the eigen values generated by LDA and the SVM weight vector to guide the recursive feature elimination to obtain a subset containing the most discriminantes variables related to the energy consumption on an EAF. The LDA\_SVM-RFE performance was better than the performance of the algorithms used for comparison. The results of test steel castings were completely satisfactory, achieving a 3.5% reduction in electricity consumption and 1 minute less for the EAF connected time in a steel casting.

## References

1. Balakrishnama, S., Ganapathiraju, J.: Linear discriminant analysis for signal processing problems (2002)
2. Bernardo, P.R.d.C., Leonardo, C.d.A.: New trends of soft computing methods for industrial and biological processes (2010)
3. Chtioui, Y., Bertrand, D., Barba, D.: Feature selection method using genetic algorithm for the classification of small and high dimension data. In: First International Symposium on Information and Communications Technologies. pp. 1–4 (2004)
4. Chu, D., Zhang, X.: Sparse uncorrelated linear discriminant analysis (2013)
5. Claeskens, G., Croux, C., Kerckhoven, J.: An information criterion for variable selection in support vector machines pp. 541–558 (2008)
6. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157–1182 (2003)

7. Guyon, I., Weston, J., Barnhill, S.: Gene selection for cancer classification using support vector machines. *Machine Learning* 46, 389–422 (2002)
8. Khan Muhammad Badruddin, K., Yagi, I., Terano, T.: Applying data mining techniques to assess steel plant operation conditions. In: *Data Mining Found. and Intell. Paradigms*. pp. 343–361 (2011)
9. Kohavi, R., H. John, G.: Wrappers for feature subset selection. *Artificial Intelligence* 97, 273–324 (May 1996)
10. Kordos, M., Blachnik, M., Wiczorek, T.: Evolutionary optimization of regression model ensembles in steel-making process (2010)
11. Masaeli, M., Fung, G., G. Dy, J.: From transformation-based dimensionality reduction to feature selection (2010)
12. Mitchell, M.: *An Introduction to Genetic Algorithms*. Massachusetts Institute of Technology, London, England, 1st edn. (1996)
13. Sandberg, E.: *Energy and Scrap Optimisation of Electric Arc Furnaces by Statistical Analysis of Process Data*. Ph.D. thesis (2005)
14. Singh, S., Silakari, S.: Generalized discriminant analysis algorithm for feature reduction in cyber attack detection system pp. 173–180 (2014)
15. Storn, R., Price, K.: Differential evolution a simple and efficient heuristic for global optimization over continuous spaces. *Global Optimization* 11(4), 341–359 (1997)
16. Vapnik, V.N.: *The nature of statistical learning theory* (1995)
17. Xue, B., Fu, W., Zhang, M.: Differential evolution (de) for multi-objective feature selection in classification. In: *GECCO 14: Conference on Genetic and Evolutionary Computation*. pp. 83–84. ACM Press (2014)
18. Younes, C., Dominique, B., Dominique, B.: Feature selection by a genetic algorithm application to seed discrimination by artificial vision (1998)