Electrofused Magnesium Oxide Classification Using Digital Image Processing and Machine Learning Techniques

A B M Shawkat Ali
School of Computing Sciences
CQ University Australia
Queensland 4702 Australia
Email: s.ali@cqu.edu.au

W K Daniel Pun
School of Computing Sciences
CQ University Australia
Queensland 4702 Australia
Email: d.pun@cqu.edu.au

Abstract—This research is focused on using digital image processing and machine learning techniques to classify Electrofused Magnesia for industry automation. We generate the data from different images by using a modern digital image process. This research proposes a new method to construct the digital image database. The propose new method is based on simple histogram mode and intensity deviation. A group of six popular machine learning algorithms has been tested to build up an automatic system for industry. We have concluded that the best suited algorithm for magnesia industry automation from this group is the PART algorithm.

I. INTRODUCTION

Magnesia Industry uses high quality magnesite to produce Electrofused Magnesia (EM). Electrofused Magnesium Oxide (EMO) is produced by transporting the calcined magnesia to the electrofusion plant where it is melted at 3000°C in electric arc furnaces. The material adopts the structure of periclase - a white relatively non-reactive solid which presents exceptional dimensional stability and strength at high temperatures. Ingots formed in this process are cooled, stripped and broken up with a mobile rockbreaker. The product then passes through a series of crushing, sorting, and screening processes before being bagged and containerised. It is then transported through Rockhampton for shipment around the world. During the sorting and screening processes, EMO will then be manually divided by operators into three categories as follows: Electorfused Higgens 1 (EFH1), shown as Fig. 1, and Electorfused Higgens 2 (EFH2), shown as Fig. 2, and High Lime Core (HLC), shown as Fig. 3. The name Higgens comes from the type of furnace it is made in. After some months of investigation and discussion with plant's engineers and other personnel, we have found that practical research can be applied to the crushed electrofused material sorting area, that is, value adding though intelligent automation into electrofused material sorting area. This research study therefore is focused on the EM products classification and identification.

The layout of this paper is that we start to focus on EMO and its categories. The problem description is reviewed in the next section. Section 3 describes how digital image data can be generated. Section 4 identifies the algorithms used for the

Fig. 1. Magnesia: Electorfused Higgens 1 (EFH1)

Fig. 2. Magnesia: Electorfused Higgens 2 (EFH2)

experiment. Six algorithms have been briefly explained. The experimental outcome is shown in section 5. In the last section, we summarise the methodology, its contribution and conclude with recommendations for future research directions.

II. PROBLEM DESCRIPTION

Since the commencement of production the manual sorting process has been continuous - manual exercise relying on human skill. Our research has identified that there are various

Fig. 3.   Magnesia: High Lime Core (HLC)

human and environmental factors associated with continuous operations, such as:

- concentration spans of humans;
- visual skill of the individual;
- noise levels and
- speed of the conveyor belt.

These factors are recognised and compensated for by various methods - they still remain even at a reduced level. All of above factors have the potential to cause:

- loose concentration;
- variance in quality throughout the day, especially before and after lunch breaks;
- people take sick leave and an inexperienced person may fill the role;
- reduced staff due to absenteeism and resignation; and
- new staff training.

Any of the above points will affect the standard of quality sorting. We believe that automation could lead to process benefits. Figure 4 visualises how the problem of sorting is being affected by human factors, the operating environment and the needs for quality control.
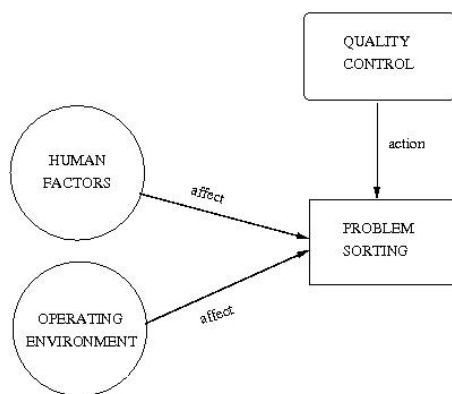

Fig. 4.   The problem of sorting

Currently EMO are still being sorted by operators who are human in the factory (see Fig. 5). From the discussion

with these operators, they determine what is EFH1 and what is HLC, and the remainder would be classified as EFH2. Operators are human beings who do not operate perfectly every single time they sort and screen. They are fallible and prone to mistakes. The objective of this study is to improve this situation in the industry plant.

Our research has shown that the feature of magnesia recognition should be integrated into a Smart Recognition System that is being developed for this project.

Fig. 5.   Manually classification

The suggested method for magnesia object detection is black and white color segmentation. In the process, objects which are not similar with magnesia are located as well. These objects are filtered later during the classification stage.

Each object goes through image processing, and is represented in the feature space. It is then labeled into its category, according to the system mode. Labeling is done either by the user while building the database, or by the system call during classification. Figure 6 shows the concept of a machine learning algorithm training scheme.

Initially, the chosen classifying algorithm is Support Vector Machine (SVM). The algorithm was shaped to the optimal classification model, based on the database. New objects were classified to different magnesia by the optimal model.

In this project, the classifier was composed of a three phase task. The first phase distinguished magnesia from other objects. It was followed by assigning the magnesia to groups by shape and colour in the second phase. The last phase completed the fine classification between magnesias in their group. Figure 7 demonstrates the concept of the classification scheme.

III. Digital Image Data Generation

The following process demonstrates how the digital image data was generated and processed. In step one, a 6-megapixel **Kodak**^{TM} EasyShare Z650 zoom digital camera was used to take photographs of EMO. Before taking raw data from all pictures taken, we removed some noise background from the pictures. In step two, after evaluating similar backgrounds for all the images over the experiment, we changed the image background to black. This process removed all the noise
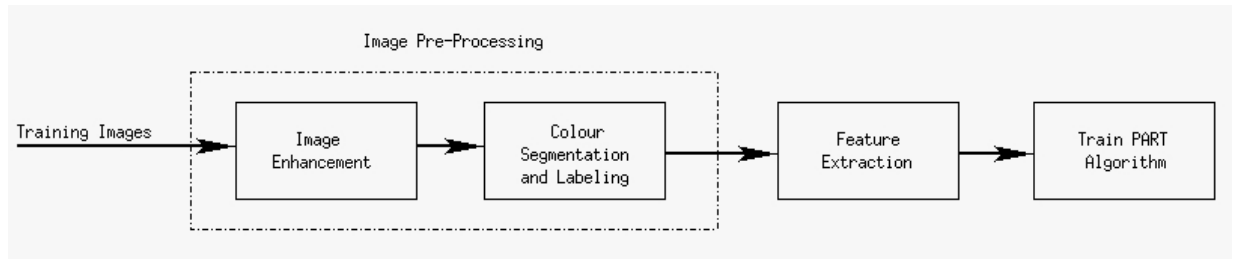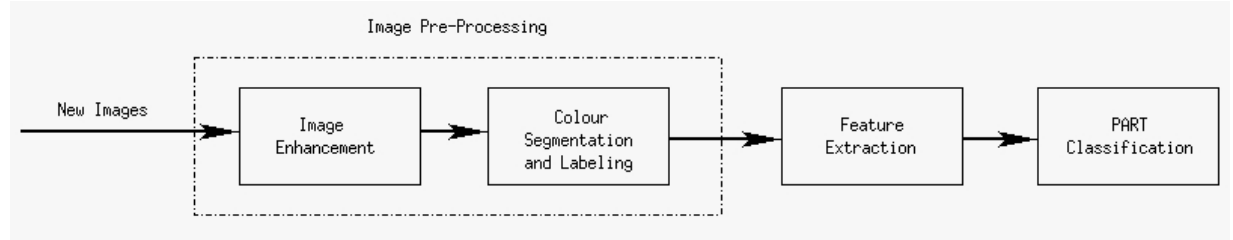
Fig. 6. Training scheme

Fig. 7. Classification scheme

information from background. Figure 8 and Figure 9 show the difference between the two backgrounds.

Fig. 8. EFH2 with original background

Fig. 9. EFH2 with background removal

The next step was to map the images as Red-Green-Blue (RGB) and gray scale representations. After that we extracted only gray scale information from the combined figure. Figure

10 and Figure 11 display the data extraction process. Finally we constructed the histogram of each image. From the histogram we collected mode values and intensity deviation for each image. We tried to construct a system faster than before. Our aim is to make a simple database for the automation task.

The intensity deviation is a similar measure of standard deviation of a discrete random variable is the *root-mean-square* (RMS) deviation of its values from the mean. If the random variable X takes on N values $x_1, \cdots, x_N$ (which are *real numbers*) with equal probability, then its intenisty deviation I can be calculated as follows:

1) Find the mean, $\bar{x}$, of the values.
2) For each value x_i calculate its deviation $(x_i - \bar{x})$
3) Calculate the squares of these deviations.
4) Find the mean of the squared deviations. This quantity is the *variance I^2*.
5) Take the square root of the variance.

This calculation is described by the following formula:

$$I = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{1}$$

Some of EFH1 images were converted as histrogram mode and intensity deviation around the mean were also summarised in Table I. We tried to feed the raw image data with the system. But the computational cost was too high and more over classification performance was very poor. The numbers 212.65 and 64.98 shown in Table I are the average values of Histogram Mode and Intensity Deviation.

IV. COMMON USEFUL ALGORITHMS

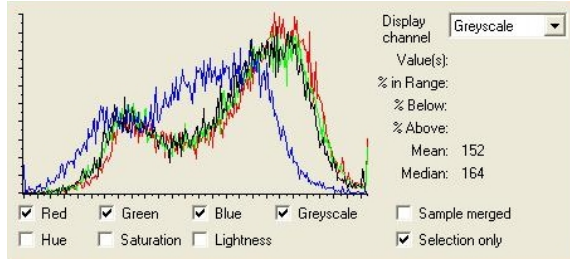In this experiment six different machine learning algorithms namely Naive Bayes (NB), Support Vector Machine (SVM),
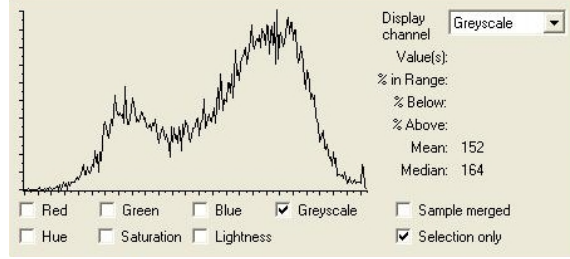
Fig. 10.   Histogram of RGB channels

Fig. 11.   Histogram of gray scale image

TABLE I
HISTOGRAM MODE AND INTENSITY DEVIATION AROUND THE MEAN

Histogram Mode	Intensity Deviation
232	52.3
218	62.8
234	54.4
192	73.2
168	74.8
232	81.8
226	62.1
231	54.4
211	54.5
230	44.1
205	59.3
221	59.9
186	63.4
204	63.2
205	72.6
217	73.4
220	82.0
194	65.0
197	74.5
230	71.9
212.65	**64.98**

IBK, ABM1, Decision Tree C4.5, and PART were used to classify the EMO to do the automation of the sorting process.

A. Naive Bayes (NB)

The Naive Bayes (NB) algorithm [7] based on the Bayesian theorem is suitable for high dimensional data classification. Given a set of variables, $X = x_1, x_2, x_3, \cdots, x_n$, our goal was to construct the posterior probability for the event y_j among a set of possible targets $Y = y_1, y_2, y_3, \cdots, y_n$. Following the traditional method, X is the independent variables and Y is the set of categorical levels i.e., dependent variable. Bayes' rule can be formulated as follows:

$$p(Y_j | [x_1, x_2, x_3, \cdots, x_n] \cdot [x_1, x_2, x_3, \cdots, x_n]') \infty$$

$$p([x_1, x_2, x_3, \cdots, x_n] \cdot [x_1, x_2, x_3, \cdots, x_n]' | Y_j) p(Y_j) \quad (2)$$

where $p(Y_j | [x_1, x_2, x_3, \cdots, x_n] \cdot [x_1, x_2, x_3, \cdots, x_n]')$ represents the posterior probability of class membership.

Naive Bayes algorithm considers the conditional probabilities of the independent variables to be statistically independent. We can now decompose the likelihood to a product of terms:

$$p(X | Y_j) \infty \prod_{k=1}^{mm} p(x_k | Y_j) \quad (3)$$

Finally, with the help of Bayes' rule, we classified a new example X with a class level C_j that earns the highest posterior probability.

B. Support Vector Machine (SVM)

The SVM algorithm, invented by Vladimir Vapnik and his group in nineteen-nineties, is one of the most successful learning and classifying supervised algorithms. It has been widely used and popular in businesses, engineering, medical, and science communities. The main function of SVM is to construct the optimal hyperplane (OH) in the training space using the proper estimation [9] [4] of a weight vector ω and the scalar bias factor b. All of the training patterns are said to be linearly separable if there exists ω and b such that the following inequalities satisfied:

$$(\omega \cdot \mathbf{X}_i) + b \geq 1; \quad \text{if } y_i = 1 \quad (4)$$

$$(\omega \cdot \mathbf{X}_i) + b \leq -1; \quad \text{if } y_i = -1 \quad (5)$$

where $\mathbf{x}_i = x_i, x_2, x_3, \cdots, x_n$ are vectors representing the data points, and y_i are the class attributes.

The prediction function of SVM is expressed [9] as follows:

$$\hat{f}(\mathbf{X}) = sign(\omega^o \cdot \Phi(\mathbf{X}) + b)$$

$$= sign\left(\sum_{i=1}^{l} \alpha_i^o y_i \Phi(\mathbf{X_i}) \cdot \Phi(\mathbf{X}_j) + b^o\right) \quad (6)$$

where α is the optimisation parameter, and the product of $\Phi(\mathbf{X_i}) \cdot \Phi(\mathbf{X}_j)$ is a scalar quantity. In machine learning literature this product is called *kernel function*. In our experiments we used **rbf** (radial basis function) kernel function with SVM.

C. IBK

Instance Based Learning IBK is a very commonly used classification method with the exception that it is possible to define the number of nearest neighbours that is considered in the K-nearest neighbour component of the algorithm. It works on the principle that first plot each training instance and then measure the distance of each test instance to the training instances. The class of the training instance with the least distance between it and the test instance is the class that

we assigned to the test instance. Basically k was chosen to be an odd number, and we took the smallest average distance of the k instances [1].

D. ABM1

Boosting is popular machine learning technique, which can improve the performance of any learning algorithm. **Adaboost.M1** (ABM1) is a upgrade version of boosting algorithm proposed by Freund and Schapire in 1996 using classification trees as single classifiers [6] [2]. We summarise the algorithm as follows:

Input: - sequence of m examples $\langle(x_1, y_1), \cdots, (x_m, y_m)\rangle$ with labels $y_i \in Y = \{1, \cdots, k\}$

weak learning algorithm **WeakLearn**

integer T specifying number of iterations

Basically boosting algorithm has access to another unspecified learning algorithm and called as WeakLearn.

Initialise $D_1(i) = 1/m$ for all i.

Do for $t = 1, 2, \cdots, T$

1) Call **WeakLearn**, providing it with the distribution D_t.
2) Get back a hypothesis $h_t : X \rightarrow Y$.
3) Calculate the error of h_t :

$$\epsilon_t = \sum_{i:h_t(x_i) \neq y_i} D_t(i). \qquad (7)$$

If $\epsilon_t > 1/2$, then set $T = t - 1$ and abort loop.

4) Set $\beta_t = \epsilon_t/(1 - \epsilon_t)$.
5) Update distribution D_t :

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} \beta_t & \text{if } h_t(x_i) = y_i \\ 1 & \textbf{otherwise} \end{cases} \qquad (8)$$

where Z_t is a normalisation constant (chosen so that D_{t+1} will be a distribution).

6) **Output** the final hypothesis:

$$h_{fin}(x) = \arg \max_{y \in Y} \sum_{t:h_t(x)=y} \log \frac{1}{\beta_t} \qquad (9)$$

E. Decision Tree C4.5 algorithm

Decision Tree is a member of the supervised learning algorithms, which was first used by Quilan [8] for data classification. A decision tree is a simple rule based learning algorithm which is easy to understand.

Given the similar data matrix $[x_1, x_2, x_3, \cdots, x_n]$ and a probability distribution $P = (p_1, p_2, p_3, \cdots, p_n)$ the **Entropy** P is expressed as follows:

$$I(P) = -(p_1 \log(p_1) + p_2 \log(p_2) + p_3 log(p_3) + \cdots$$
$$+ p_n \log(p_n) + \cdots \cdots) \quad (10)$$

If a set T of records is partitioned into disjoint exhaustive targets $y_1, y_2, y_3, \cdots, y_k$ based on categorical attribute, then the information needed to identify the target of an element of T is $Info(T) = I(P)$, where P is the probability distribution of the target i.e., partition $(y_1, y_2, y_3, \cdots, y_k)$ and P can be computed using the following equation:

$$P = (|y_1|/|T|, |y_2|/|T|, |y_3|/|T|, \cdots, |y_n|/|T|) \qquad (11)$$

Now, we can define the quantity **Gain** (X, T) for a variable as:

$$Gain(X, T) = Info(T) - Info(X, T) \qquad (12)$$

Finally, the measure **Gain** to rank variables was used to build decision trees depending on where each node was located and the variable with greatest **Gain** among the variables was not yet considered in the path from the root.

F. PART

PART is a more recent scheme for producing sets of rules called "decision lists", which are ordered sets or rules. PART is developed from the C4.5 and RIPPER algorithms and is a partial decision tree algorithm. However, unlike C4.5 and RIPPER, PART does not have to perform global optimisation in order to generate rules. This algorithm works by forming pruned partial decision trees (built using C4.5's heuristics), and immediately converting them into a corresponding rule [3] [5].

V. Comparison of the Algorithms Performance

10-fold cross validation technique was used for evaluation in the experiment and we obtained good results. Table II shows the six algorithms performance for EMO classification through digital image processing.

TABLE II
RESULTS FROM DIFFERENT ALGORITHMS

Dataset	(1) NB	(2) SVM	(3) IBK	(4) ABM1	(5) C4.5	(6) PART
Mgdata	77.62	73.71	83.76	78.48	80.69	84.31

Also, a summary of the 10-fold cross validation results is shown in Table II. The reported values are indicating the percentage of accuracy for the classification task.

We have found that the PART algorithm is the number one choice for our task. The second choice is the IBK algorithm. The SVM performed worst in the experiment. However, we did not try the different types of kernels in the SVM implementation. We chose polynomial kernel with polynomial degree 2. Since the data has two attributes, therefore the system is extremely faster than before.

VI. Conclusion

In this research project, an EMO classification system is being implemented. The classification process contains three phases: digital imaging, data extraction by using statistical methods, and classification by using traditional learning algorithms. The system performance for each classification algorithm was evaluated by the 10-fold cross validation method. Finally, we found that the PART algorithm is the best choice for the magnesia classification task. The SVM performance could be improved by different types of kernels and it's parameter tuning. The future aim is to extend this project by considering more images with suitable data extraction techniques. The optimal model can be found to have low error rates with better equipment in the near future research.

REFERENCES

[1] D. Ada and D. Kibler, "Instance-based learning algorithms", *Machine Learning*, vol. 6, pp. 33-66, 1991.

[2] E. Alfaro, M. Gamez, and N. Garcia, "Multiclass corporate failure prediction by Adaboost.M1", *International Advances in Economic Research*, Springer, vol. 13, no. 3, pp. 301-312, August, 2007.

[3] S. Ali and K. A. Smith, "On learning algorithm selection for classification", *Journal on Applied Soft Computing*, ELSEVIER, vol. 1, pp. 119-138, 2006.

[4] B. E. Boser, I. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers", *Proceedings of the Fifth Annual Workshop of Computational Learning Theory*, ACM Press: Pittsburgh, pp. 144-152, 1992.

[5] S. J. Cunningham and G. Homes, "Developing innovative applications in agriculture using data mining", *Tech Report*, Dept. of Computer Science, University of Waikato, New Zealand.

[6] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm", In *Proceedings of the Thirteenth International Conference on Machine Learning*, Morgan Kaufmann, pp. 148-156, 1996.

[7] S. Mehran, "Learning limited dependence Bayesian classifiers", In *the Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, AAI Press, pp. 335-338, 1996.

[8] R. Quilan, *C4.5: Programs for Machine Learning*, CA: Morgan Kaufman Publishers, 1993.

[9] V. Vapnik, *Statistical Learning Theory*, USA: John Wiley and Sons, 1998.