Fisher's Linear Discriminant Analysis Based Prediction using Transient Features of Seismic Events in Coal Mines

Başak Esin Köktürk Güzel, Bilge Karaçalı
Department of Electrical and Electronics Engineering
İzmir Institute of Technology
İzmir, Turkey
Email:{basakkokturk,bilge}@iyte.edu.tr

Abstract—**Identification of seismic activity levels in coal mines is important to avoid accidents such as rockburst. Creating an early warning system that can save lives requires an automated way of predicting. This study proposes a prediction algorithm for the AAIA′16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines that is based on transient activity features along with average indicators evaluated by a Fisher's linear discriminant analysis. Performance evaluation experiments on the training datasets revealed an accuracy level of around 0.9438 while the performance on the test dataset was at a level of 0.9297. These results suggest that the proposed approach achieves high accuracy in predicting danger seismic events while maintaining low complexity.**

I. Introduction

ONE OF the most important subjects in coal mining is to detect specific gas emissions and seismic activity rates. The miners can suddenly find themselves in dangerous situations due to methane explosions or rockburst [1]. The most common accident cause in coal mines are cave-ins (roof, rock and coal) which account for 70.6 of all injuries [2]. Safety of miners' lives substantially depends on an early warning mechanism that can potentially be constructed using specific alert measurements for seismic activities as well as physical conditions of the mine. In order to create such an early warning mechanism, warning signals can be triggered if the measured values or energy levels, taken measurement from reference points of the mine, exceed a preset hazard threshold. However, seismic activity datasets that are observed from coal mines are very high dimensional and hard to process due to the fact that they are measured from a wide range of points and for a long duration. Since the dataset is very complex and high dimensional, expert knowledge-based systems can fail for foresight of the dangerous activities. The automation of early warning systems in coal mines has vital importance to prevent interpretation differences between mining experts and make analysis more rapid.

In the literature, there are several automated methods that have been proposed to recognize hazardous seismic activity patterns. Neural networks are the most popular method for prediction of seismic events in coal mines [3]. Identification of neural network parameters and layer numbers, however, is complicated, and entails substantial cost because of its "black-box" structure [4]. As a simpler and practical alternative, we propose a hazardous seismic ebent activity prediction method based on Fisher′s Linear Discriminant Analysis [5] that operates on an encoding of transient seismic activity are on 24 hour period along with average seismic activity parameters and conventional risk assessment methods. Performance evaluation of the method on the AAIA′16 Data Mining Challenge Dataset suggest that the approach offers accurate prediction of hazardous seismic activity around %92.97 levels.

In the next section, we provide a detailed description of the dataset and explain the proposed approach. In the third section, we present performance evaluation results of our method on both initial training dataset as well as the additional training datasets. At the conclusion section, we summarize our algorithm and discuss the results.

II. Material and Methods

The dataset used in this study was provided by Research and Development Centre EMAG for AAIA′16 Data Mining Challenge: Predicting Dangerous Seismic Events in Coal Mines. The dataset consists of total energy measurements for 24 hour period from different sensors and the counts for seismic bumps perceived at longwalls. In addition, the dataset contains hourly readings for 24 consecutive hours that are related with the most recent assessments of the conditions determined by mining experts. In the dataset each sample has one ID of the main working site where the measurements were taken and 540 features which contains 12 average risk parameters and 528 risk assessments measures. Finally, the respective labels ("normal" or "warning") are provided for each individual sample to assist on the training.

We propose a method that evaluates the average risk parameters along with the risk assessment measures separately from the hourly measurements of the provided parameters over a 24-hour period for the prediction task at hand. To this end, we extracted the hourly measurements of the 22 different parameters provided in the training dataset
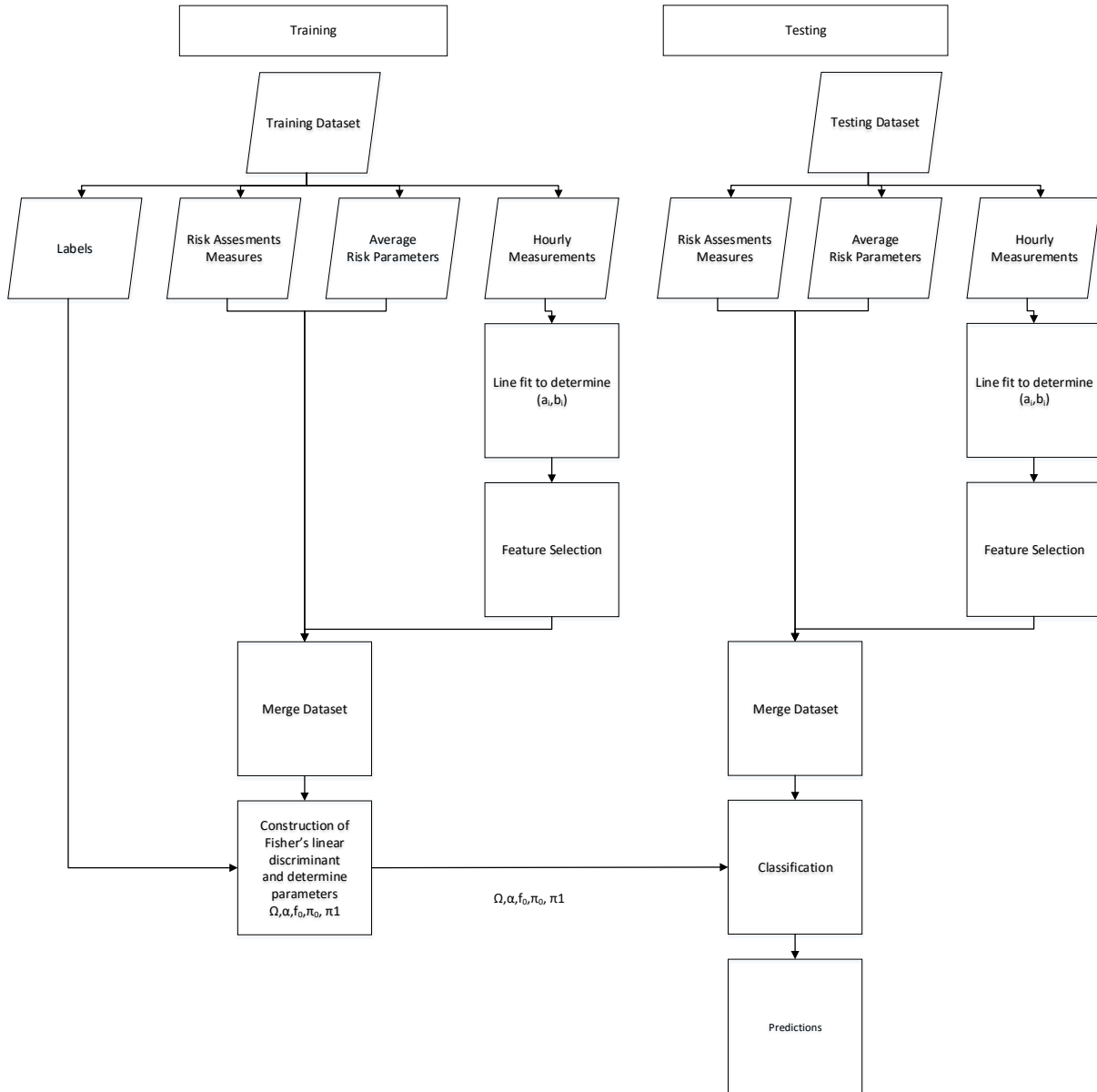
Fig. 1. Block Diagram of the Proposed Method

in indices from 14 through 541 and performed a line fit to determine the parameters (a_i, b_i) such that the fit error

$$\frac{1}{2} \sum_{h=1}^{24} (p_{i,h} - (a_i h + b_i))^2 \qquad (1)$$

is minimal for each parameter p_i with hourly measurements $p_{i,h}$, for $i = 1, 2, \ldots, 22$. This resulted in a time evolution dataset with 44 features. To further refine this dataset, we have calculated the Kolmogorov-Smirnov statistic [6] between the empirical cumulative probability distributions of the two groups over each of the 44 time evolution parameters (a_i, b_i) for $i = 1, 2, \ldots, 22$, and ranked the pa-

rameters in the order of decreasing statistic value, with the understanding that the larger values of the statistic indicate more pronounced separation between the groups. Next, we have carried out Fisher's linear discriminant analysis [7] on the top ranked $1, 2, \ldots, 44$ time evolution parameters, and calculated the area under the resulting receiver operating characteristics curve obtained on the original training dataset and the associated labels. This analysis identified 39 time evolution parameters with the greatest Kolmogorov-Smirnov statistic providing the highest area under the curve on the original training dataset, that were then collected to form the

calculated time evolution dataset containing the transient features.

The final prediction was obtained by merging the average risk parameters and the risk assessment measures provided in indices 2 through 13 in the training dataset with the calculated time evolution dataset, and constructing a Fisher's linear discriminant over the merged dataset. This entailed calculating the average vectors and covariance matrices

$$\mu_0 = \frac{1}{\ell_0} \sum_{j \in J_0} x_j \tag{2}$$

$$\mu_1 = \frac{1}{\ell_1} \sum_{j \in J_1} x_j \tag{3}$$

$$\Sigma_0 = \frac{1}{\ell_0 - 1} \sum_{j \in J_0} (x_j - \mu_0)(x_j - \mu_0)^T \tag{4}$$

$$\Sigma_1 = \frac{1}{\ell_1 - 1} \sum_{j \in J_1} (x_j - \mu_1)(x_j - \mu_1)^T \tag{5}$$

over the parameter vectors of the merged dataset $\{x_j\}$ with respect to the index sets J_0 and J_1 defined by

$$J_0 = \{j | y_j = 0\} \tag{6}$$

and

$$J_1 = \{j | y_j = 1\} \tag{7}$$

in terms of the training labels $\{y_j\}$ with $\ell_0 = |J_0|$ and $\ell_1 = |J_1|$. This allowed expressing the discriminant function $f(x)$ for a new parameter vector x through

$$f(x) = w^T x \tag{8}$$

with

$$w = (\Sigma_0 + \Sigma_1)^{-1}(\mu_1 - \mu_0). \tag{9}$$

As the final step of the analysis, we have identified the parameters α and f_0 to convert the values of the discriminant function into an empirical log-likelihood ratio for the two groups via the expression

$$L(x) = \alpha(f(x) - f_0) \tag{10}$$

so that the collection of values $\{L(x_j)\}$ over the merged training dataset $\{x_j, y_j\}$ achieved the smallest average training errors on the two groups with respect to a threshold of 0, or the average of the Type I and Type II training error rates, and the corresponding empirical posterior probabilities given by

$$P_1(x_j) = \frac{\pi_1}{\pi_0 e^{-L(x_j)} + \pi_1} \tag{11}$$

satisfied

$$\frac{1}{\ell_0 + \ell_1} \sum_{j=1}^{\ell_0 + \ell_1} P_1(x_j) = \pi_1 \tag{12}$$

with π_0 and π_1 denoting the prior probabilities of the respective groups in the training dataset. This was carried out by finding the f_0 value that achieved the equality above for a specific value of α for $\alpha = 2^{-5}, 2^{-4.5}, 2^{-4}, \ldots, 2^5$. Calculating the errors over the resulting (α, f_0) pairs identified the best

values for α and f_0 for the smallest training error while maintaining the required prior probabilities.

In the next section, we present the results that we have obtained on different training and test dataset combinations.

III. RESULTS

We have been tested our proposed method using the five different dataset and respective warning level labels that were provided by AAIA′16 Challenge committee. These datasets were named as: `training dataset`, `additional training dataset 1`,`additional training dataset 2`, `additional training dataset 3` and `additional training dataset 4`. Different combinations of these datasets were used to train the algorithm and the others were used for testing its performance.

Firstly, we used the original training dataset and we estimated the posterior probabilities for both additional training datasets as well as the original training dataset. The receiver operating characteristic (ROC) curves for this trial is shown in Figure 2.

The greatest AUC was obtained on the additional train dataset 1 at 0.9619 followed by at 0.9422 and at 0.9345 and at 0.9088 and at 8943. Next, we merged each additional
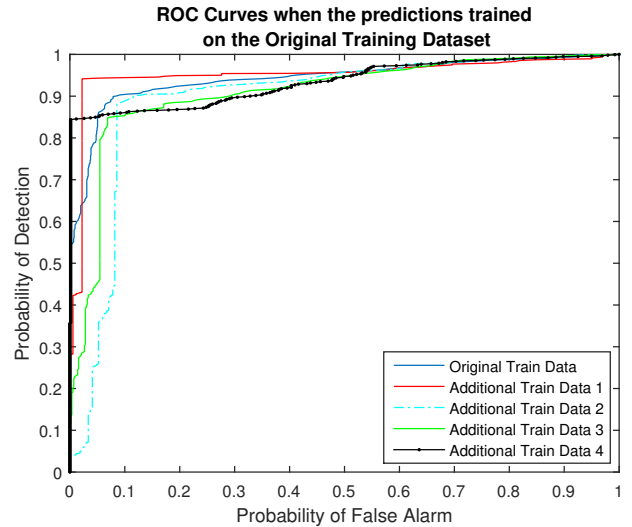

Fig. 2. The ROC curves when the algorithm trained by original training dataset. Best prediction is performed on additional training dataset 1

training dataset with the original training dataset and used the combined data to train the algorithm, and tested the resulting prediction on all datasets. The average area under curve (AUC) values are shown in Table I.

The area under the receiver operating characteristics curve obtained on the test dataset was 0.9297 as reported by the evaluation committee of the AAIA'16 Challenge when the training data was combination of the original training data and additional training data 1. The ROC curves are shown in Figure 3 for each dataset.

TABLE I
AVERAGE AREA UNDER CURVE VALUES FOR DIFFERENT TRAINING SET
COMBINATIONS

| Train Data	Average AUC Value
Original Train Data	0.9264±0.0240
Original Train Data , Additional Train Data 1	0.9311±0.0292
Original Train Data , Additional Train Data 2	0.9280±0.0253
Original Train Data , Additional Train Data 3	0.9298±0.0240
Original Train Data , Additional Train Data 4	0.9290±0.0224
Original Train Data , Additional Train Data 1, Additional Train Data 2	0.9320±0.0283
Original Train Data , Additional Train Data 1, Additional Train Data 3	0.9334±0.0261
Original Train Data , Additional Train Data 1, Additional Train Data 4	0.9337±0.0275
Original Train Data , Additional Train Data 2, Additional Train Data 3	0.9336±0.0250
Original Train Data , Additional Train Data 2, Additional Train Data 4	0.9303±0.0228
Original Train Data , Additional Train Data 3, Additional Train Data 4	0.9327±0.0220
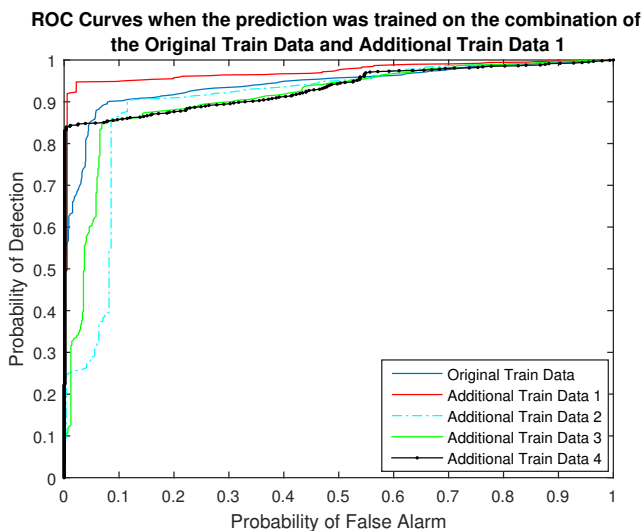
Fig. 3. The ROC curves when the algorithm was trained on the merge of the original training dataset with the additional training dataset 1. Best prediction performance is achieved on the additional training dataset 1

IV. CONCLUSION

In this paper, we have proposed a prediction algorithm for dangerous seismic events in coal mines using a combination of existing risk assessment parameters, average seismic energy measurements as well as hourly seismic activity measurements and Fisher's linear discriminant analysis. The method fits a line to capture the seismic activity information provided by hourly measurements and uses the two line-fit parameters as features for the ensuing prediction subjected to feature selection using Kolmogorov-Smirnov statistics and area under the curve measures on the training data. In order to produce the final predictions, we have applied a mathematical conversion on the outputs of the discriminant function to produce empirical posterior probabilities that ranged between 0 and 1, indicating the likelihood of a future seismic event. At an additional level of complexity, we have also evaluated the performance of the predictions subject to different training datasets, as training datasets themselves vary in the level at which they represent the actual prediction problem. In the performance comparison tests over the training data, we observed varying accuracy levels for the different training datasets used, and submitted the best performing configuration to the AAIA'16 challenge, that achieved an area under the curve level of 0.9297 on the test data that was withheld from the challenge participants. The strengths of our proposed method lie first in the manner with which the hourly seismic activity measurements are evaluated and merged with the average measurements as well as existing risk assessment parameters. In addition, the simplicity of the Fisher's linear discriminant function offers a greater potential for generalizability of the demonstrated high performance to other seismic activity prediction cases as it minimizes the risk for overtraining. Finally, the conversion of the prediction results into posterior probabilities allows processing the results in conjunction with other probabilistic insights that one may have on the prediction problem at hand such as site-specific conditions and associated risks no reflected in the measurements. This also reflect the weakness of the method proposed here as it does no take into account any site-specific information, though this can be remedied in future applications.

REFERENCES

[1] A. Janusz, M. Sikora, Ł. Wróbel, S. Stawicki, M. Grzegorowski, P. Wojtas, and D. Ślęzak, "Mining data from coal mines: Ijcrs'15 data challenge," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing.* Springer, 2015, pp. 429–438.
[2] M. Sari, H. S. B. Duzgun, C. Karpuz, and A. S. Selcuk, "Accident analysis of two turkish underground coal mines," *Safety Science*, vol. 42, no. 8, pp. 675–690, 2004. doi: http://dx.doi.org/10.1016/j.ssci.2003.11.002
[3] J. Van Zyl and C. W. Omlin, "Prediction of seismic events in mines using neural networks," in *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*, vol. 2. IEEE, 2001. doi: http://dx.doi.org/10.1109/IJCNN.2001.939568 pp. 1410–1414.
[4] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of clinical epidemiology*, vol. 49, no. 11, pp. 1225–1231, 1996. doi: http://dx.doi.org/10.1016/S0895-4356(96)00002-9
[5] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936. doi: http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x
[6] B. Bagwell, "A journey through flow cytometric immunofluorescence analyses—finding accurate and robust algorithms that estimate positive fraction distributions," *Clinical Immunology Newsletter*, vol. 16, no. 3, pp. 33–37, 1996. doi: http://dx.doi.org/10.1016/S0197-1859(00)80002-3
[7] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification.* John Wiley & Sons, 2012.