

Naive Bayes Multi-Label Classification Approach for High-Voltage Condition Monitoring

Imene Mitiche*, Alan Nesbitt*, Philip Boreham[‡],
Brian G. Stewart[†] and Gordon Morison*

* School of Engineering and Built Environment

Glasgow Caledonian University,
Glasgow, United Kingdom

[†] Institute of Energy and Environment

University of Strathclyde,
Glasgow, United Kingdom

[‡] Innovation Centre for Online Systems

Bere Regis,
United Kingdom

Abstract—This paper addresses for the first time the multi-label classification of High-Voltage (HV) discharges captured using the Electromagnetic Interference (EMI) method for HV machines. The approach involves feature extraction from EMI time signals, emitted during the discharge events, by means of 1D-Local Binary Pattern (LBP) and 1D-Histogram of Oriented Gradients (HOG) techniques. Their combination provides a feature vector that is implemented in a naive Bayes classifier designed to identify the labels of two or more discharge sources contained within a single signal. The performance of this novel approach is measured using various metrics including average precision, accuracy, specificity, hamming loss etc. Results demonstrate a successful performance that is in line with similar application to other fields such as biology and image processing. This first attempt of multi-label classification of EMI discharge sources opens a new research topic in HV condition monitoring.

I. INTRODUCTION

Electromagnetic Interference (EMI) is a measurement technique exploited mainly for Partial Discharge (PD) and other insulation fault detection in High-Voltage (HV) systems [1]. PD is an electrical discharge which appears in insulation degradation, and leads to asset breakdown if it is consistent over time [2]. This technique measures other discharge sources such as arcing and exciter pulses. EMI diagnosis is performed by experts who analyse the measured signals to detect faults and assess the condition of an HV site [3].

Utility companies are suffering from a lack in expert knowledge and staff for condition monitoring, due to retirement of an ageing workforce, as well as sickness and periods of annual leave. This can have an impact on tasks that depend on expert knowledge, specifically the analysis of data measured on site in order to identify potential faults in a particular asset. Companies can be left with no other option but to hire or train more staff, however this requires a considerable amount of investment.

Intelligent systems are computer based algorithms that are generally trained to perform certain human duties quicker. This reduces the dependence on staff while possibly helping experts and non-experts in gaining more confidence and

information on data interpretations. Furthermore, this brings the benefit of minimising the consequences associated with asset failure, such as maintenance costs, due to poor condition monitoring. Such intelligent systems are based on machine learning which involves training and classifying two or more classes where each class is associated with one or multiple labels.

Single label classification is popular in many problems such as insulation and mechanical fault detection. For instance, in [4] authors proposed single label classification for EMI discharge sources. Similar work was performed in [5]–[7] where signals containing multi-labels were treated as a single class. The inconvenience with this approach is that the classifier outputs one label per input instance only. In real world applications this is not always the case. For example, there could be multiple faults captured in a single signal such as PD and arcing or exciter pulses.

The aim in this work is to develop a machine learning algorithm with multi-label classification ability for condition monitoring of HV plants. The work exploits the naive Bayes approach which is based on [8]. This approach was successful in multi-label classification of natural scene images and yeast genes. The classification task is complex on raw data, therefore it is desired to employ feature extraction techniques to retrieve important information and unique fingerprints. In this work, a combination of two feature extraction techniques, called 1D-Histogram of Oriented Gradients (HOG) [9] and 1D-Local Binary Pattern (LBP) [10], are employed on EMI time signals. These two methods each provide a histogram of features.

This paper is structured as follows. Section 2 provides detailed mathematical description of the employed evaluation metrics, feature extraction and classification algorithms. Section 3 introduces the experimental set-up including EMI signal measurement followed by classification results. Section 4 concludes the main findings and provides future recommendations.

II. METHODOLOGY

A. Proposed algorithm's outline

In this work we aim to develop a multi-label classification framework for EMI discharge sources classification, inspired by the naive Bayes classifier proposed in [8]. Multi-label classification can be defined as the task of potentially detecting the two or more labels within a single signal. The data analysed was collected from real world operating power plant using EMI measurement methods in the form of time signals. Prior to classification, 1D-HOG and 1D-LBP are employed to retrieve temporal and magnitude information. Before describing the employed feature extraction and classification algorithms, definitions on multi-label learning and methods for evaluation are provided.

B. Multi-label learning

The mathematical notation in this paper denotes scalars by lower case, vectors by bold lower case and matrices by bold upper case. For input data $\mathbf{x} = \mathbb{R}^d$, there is a limited set of labels $\mathbf{y} = \{1, 2, \dots, Q\}$. The objective in multi-label learning is to produce a function $f : \mathbf{x} \rightarrow 2^{\mathcal{Y}}$ in the form of $\mathbf{x} \times \mathbf{y}$ from a training data set $\mathbf{D} = (x_i, \mathbf{y}_i); 1 \leq i \leq m$, where x_i is an input instance and \mathbf{y}_i is its relative set of labels. A successful learning might provide more labels that are in \mathbf{y}_i than those that are not. In order to solve this issue, a ranking function is introduced to the learning in order to map the learning outputs and transforms $f(x_i, y_1) > f(x_i, y_2); y_1 \in \mathbf{y}_i, y_2 \notin \mathbf{y}_i$ to $rank_f(x_i, y_1) < rank_f(x_i, y_2)$. The performance evaluation of a multi-label classifier is complex and different from a single-label classifier. In this work, five evaluation metrics, proposed in [8], known as hamming loss, one-error, coverage, ranking loss, average precision, are employed. These metrics are specific to multi-label classification and are calculated for a test set $\mathbf{S} = (x_i, \mathbf{y}_i); 1 \leq i \leq p$ as follows:

$$hloss(h) = \frac{1}{p} \sum_{i=1}^p \frac{1}{Q} |h(x_i) \Delta \mathbf{y}_i| \quad (1)$$

$$one - error(f) = \frac{1}{p} \sum_{i=1}^p \left[\left[\underset{y \in \mathbf{y}_i}{argmax} f(x_i, y) \right] \notin \mathbf{y}_i \right] \quad (2)$$

$$coverage(f) = \frac{1}{p} \sum_{i=1}^p \max_{y \in \mathbf{y}_i} rank_f(x_i, y) - 1 \quad (3)$$

$$rloss(f) = \frac{1}{p} \sum_{i=1}^p \frac{1}{|\mathbf{y}_i| |\bar{\mathbf{y}}_i|} \left| \{(y_1, y_2) | f(x_i, y_1) \leq f(x_i, y_2), (y_1, y_2) \in \mathbf{y}_i \times (\bar{\mathbf{y}}_i)\} \right| \quad (4)$$

$$avgprec(f) = \frac{1}{p} \sum_{i=1}^p \frac{1}{|\mathbf{y}_i|} \sum_{y \in \mathbf{y}_i} \frac{\left| \{y' | rank_f(x_i, y') \leq rank_f(x_i, y), y' \in \mathbf{y}_i\} \right|}{rank_f(x_i, y)} \quad (5)$$

where Δ is the symmetric difference between two sets, $\llbracket \cdot \rrbracket = 1$ if the predicate holds otherwise it is zero, and $\bar{\mathbf{y}}_i$ is the complementary of \mathbf{y}_i . The hamming loss is evaluated on the classifier and its value defines the number of times an instance is misclassified. The remaining metrics are evaluated over the learning function to determine the ranking degree of excellence of the multiple labels for each instance. One-error determines the number of times the top ranked label is incorrect. Coverage defines the required number of steps to jump to the next label within the label list in order to identify all labels for each instance. Ranking loss computes the average fraction of label pairs that are arranged incorrectly for the instance. Average precision computes the average fraction of labels that are ranked above the actual top ranked label. The smaller the values for these metrics the better, except for average precision which should be high.

C. 1D-Histogram of Oriented Gradients (HOG) descriptor

The original HOG was first proposed to retrieve information from 2D data images for object detection [11]. However, since in this work the analysed data of interest is time signals a 1D version of HOG is implemented to capture the temporal information from the time series data. The implementation steps are as follows. Divide a time series $\mathbf{x}(t); t = 1, 2, \dots, N$ with length N , into n constant length overlapping intervals. In each interval, with temporal range $[t_1 \ t_2]$, the magnitude of gradient at time $t(t_1 \leq t \leq t_2)$ is obtained as:

$$|g(t)| = \left| \frac{\sigma}{2} (\mathbf{x}(t+1) - \mathbf{x}(t-1)) \right| \quad (6)$$

where σ is a scaling factor to ensure that the angular orientation of the gradient ($\tan^{-1}(g(t))$) lies approximately evenly within $[-90^\circ \ 90^\circ]$. The next step involves the use of a kernel smoothed voting on gradient within all orientation bins, where the voting magnitude for each orientation bin b_i is defined as:

$$I = |g(t)| e^{-\frac{1}{2\hat{\sigma}^2} (\tan^{-1}(g(t)) - \angle(b_i))^2} \quad (7)$$

where $\angle(b_i)$ is the orientation of the bin and $\hat{\sigma}$ represents the decay rate of the Gaussian smoothing kernel. The HOG is finally obtained across the bins. This method is robust against noise, however it extracts only temporal information. Therefore, 1D-LBP is employed along the 1D-HOG to extract information on the amplitude variations of the analysed EMI data, as the more information is presented to the classifier, the better the performance will be.

D. 1D-Local Binary Pattern (LBP)

This technique is derived from the original 2D-LBP for image feature extraction [12]. The aim in 1D-LBP is to process a single vector of data samples, and map each data sample to an LBP code through comparison against neighbouring samples. The mathematical concept of 1D-LBP is as follows. For a time series $\mathbf{x}(t); t = 1, 2, \dots, N$ with P neighbour samples, the 1D-LBP is obtained by:

$$LBP = \sum_{r=0}^{\frac{P}{2}-1} s[\mathbf{x}(t+r-\frac{P}{2}) - \mathbf{x}(t)]2^r + s[\mathbf{x}(t+r+1) - \mathbf{x}(t)]2^{r+\frac{P}{2}} \quad (8)$$

given the sign function

$$s[x] = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

where r is the distance between the centre sample and the neighbouring one. Here $P = 8$ and $r = 1$ are chosen which provides 2^8 LBP codes. The amplitude of the centre sample C is compared against 8 neighbours, 4 occurring before C and 4 after. This provides a binary code which is then multiplied by the binomial weights of the respective samples. The sum of the resulting values provides the LBP code. The variations between the centre sample and its neighbours are quantified by the LBP value, in that the value tends towards zero for signals with slow amplitude variations and it is large for signals with multiple narrow pulses.

E. Naive Bayes classification

Given an input $x \in \mathbb{R}^d$ with its relative label(s) $\mathbf{y} \subseteq \mathbb{Y}$, a class vector $\vec{\mathbf{y}}_x$ is provided for the input instance, in that the l^{th} component $\vec{\mathbf{y}}_x(l) = 1$ if $l \in \mathbf{y}$, else $\vec{\mathbf{y}}_x(l) = 0$. Given a test input $t = 1, 2, \dots, d$, assume that H_1^l is the case in which t is associated with label l and H_0^l is the case in which t is not associated with label l . The class vector is then obtained using the following Maximum A Posteriori (MAP) concept:

$$\vec{\mathbf{y}}_t(l) = \arg \max_{b \in \{0,1\}} P(H_b^l|t), \quad l \in \mathbb{Y} \quad (9)$$

Based on the Bayesian rule and assuming class conditional independence within features, Equation 9 can be reformulated to:

$$\begin{aligned} \vec{\mathbf{y}}_t(l) &= \arg \min_{b \in \{0,1\}} \frac{P(H_b^l)P(t|H_b^l)}{P(t)} \\ &= \arg \min_{b \in \{0,1\}} P(H_b^l) \prod_{k=1}^d P(t_k|H_b^l) \end{aligned} \quad (10)$$

where $P(t_k|H_b^l)$ is computed as:

$$P(t_k|H_b^l) = g(t_k, \mu_k^{lb}, \sigma_k^{lb}), \quad 1 \leq k \leq d \quad (11)$$

in that $g(\cdot, \mu_k^{lb}, \sigma_k^{lb})$ is the Gaussian probability density function, with mean and standard deviation of μ_k^{lb} and σ_k^{lb} respectively, for the k^{th} feature with H_b^l condition. Using Equation 11 in 10, The MAP can be calculated as follows:

$$\begin{aligned} \vec{\mathbf{y}}_t(l) &= \arg \max_{b \in \{0,1\}} P(H_b^l|t) \exp(\phi_b^l), \\ \phi_b^l &= - \sum_{k=1}^d \frac{(t_k - \mu_k^{lb})^2}{2\sigma_k^{lb2}} - \sum_{k=1}^d \ln(\sigma_k^{lb}) \end{aligned} \quad (12)$$

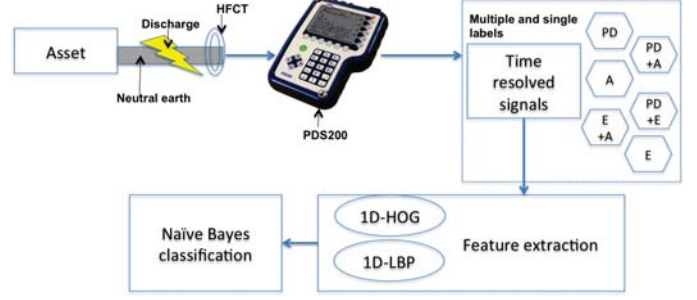


Fig. 1. Diagram representing procedure for EMI signal measurement and identification of the label(s) (Partial Discharge (PD), Arcing (A), Exciter (E)) contained within the signal.

In the case of high input dimension d in practice, ϕ_b^l may take a large negative value making exponentiation calculation challenging for computers. Thus, the probability $P(H_1^l|t)$ is first calculated as:

$$\begin{aligned} P(H_1^l|t) &= \frac{P(H_1^l)P(t|H_1^l)}{P(H_0^l)P(t|H_0^l)} = \frac{P(H_1^l)}{P(H_1^l) + P(H_0^l) \frac{P(t|H_0^l)}{P(t|H_1^l)}} \\ &= \frac{P(H_1^l)}{P(H_1^l) + P(H_0^l) \exp(\phi_0^l - \phi_1^l)} \end{aligned} \quad (13)$$

This provides a computable $\exp(\phi_0^l - \phi_1^l)$ as the difference would be reasonable. Next, $P(H_0^l|t)$ can be calculated as $1 - P(H_1^l|t)$ and finally find $\vec{\mathbf{y}}_t(l)$ using Equation 9.

III. EXPERIMENT

A. EMI data set

TABLE I
DATA SET INFORMATION.

Event(s)	Site	No. of Files	Total no. of segments
Arcing (A)	4	12	546
PD	1 to 11	113	4847
Exciter (E)	2, 3, 5, 7, 9 to 11	20	728
A+PD	1, 5, 7	10	590
A+E	7	1	59
PD+E	1 to 4, 8, 12	20	1122

The signals analysed in this work were captured using the EMI technique which follows the CISPR16 standard [13]. The measurement and recording set up was performed in real world operating HV power plants. Figure 1 summarises this approach. First, a high frequency current transformer sensor was connected around the neutral earth connection of the asset. The measured sensed signals were then recorded in a device called the Partial Discharge Surveyor (PDS)200 at a sampling rate of 24kHz. With the help of EMI expert analysis, which is usually performed through hearing and visualisation of the time resolved signals or EMI frequency spectrum, the signals that contain faults or an event are labelled and selected for classification. This includes single-label and multi-label signals which contain PD, arcing, exciter, PD and arcing, PD and exciter, exciter and arcing events. Figure 2 shows examples of time resolved signals for each type and Table

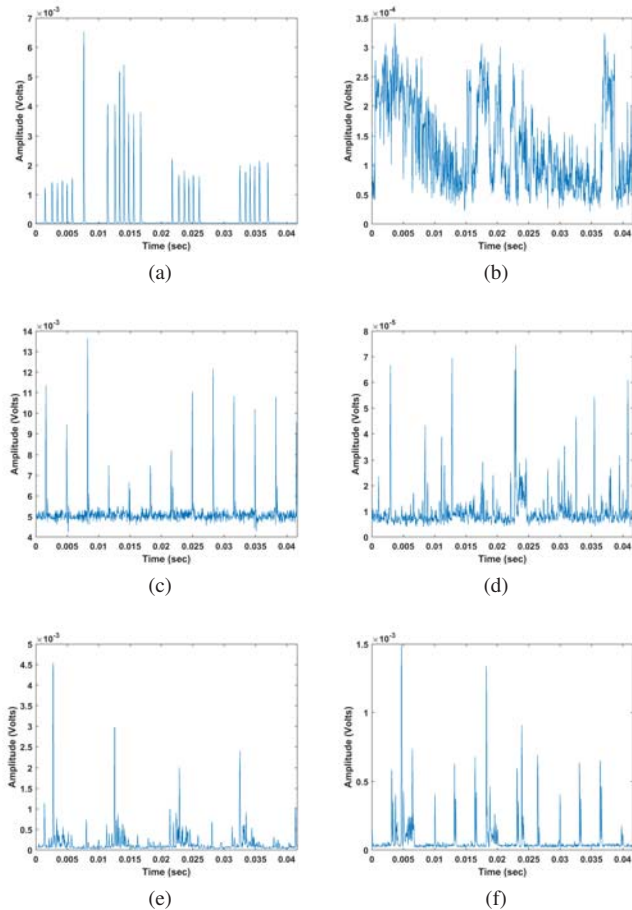


Fig. 2. EMI time resolved signal examples of (a) Partial Discharge (PD) (b) arcing (c) exciter (PN) (d) PD and arcing(RN) (e) PD and exciter (f) exciter and arcing (E).

I provides a more details on the data set. Prior to feature extraction, the signals were segmented with a 4000 samples window. Two time series feature extraction techniques known as 1D-HOG and 1D-LBP are employed independently and jointly. The feature vector obtained for each instance is implemented in the naive Bayes classifier. Here, the dataset is randomly split into 70% for training the classifier and the remaining 30% are used for testing. This was performed over ten iterations and the average performance is presented in the next subsection.

B. Results

Table II shows the classification results over the dataset in terms of the five metrics: average precision, coverage, hamming loss, one-error and ranking loss. It is observed that 1D-HOG has a slightly better performance than 1D-LBP, however the combination of both provides improved results, where the losses, coverage and error are low and average precision is higher. This results are as desired and are in line with previously published results on multi-label classification using naive Bayes method on yeast data and natural scene image classification. On the other hand, the obtained coverage and average precision are higher than the yeast and natural

TABLE II
MULTI-LABEL CLASSIFICATION RESULTS AND COMPARISON TO NATURAL SCENE IMAGE DATA AND YEAST DATA CLASSIFICATION USING MLNB [8].

	avg. prec.	coverage	hloss	one-error	rloss
1D-LBP	0.75	1.06	0.40	0.41	0.38
1D-HOG	0.77	1.04	0.36	0.34	0.35
1D-HOG +1D-LBP	0.78	1.02	0.35	0.33	0.34
Natural scene data	0.76	1.09	0.19	0.36	0.20
Yeast data	0.75	6.45	0.20	0.23	0.17

scene image results [8]. Overall, this outcome reveals that the proposed method is successful in terms of all evaluation metrics. This classification could support expert analysis in decision making. However, by training the proposed model with more data instances, it will gain more confidence and therefore it will be able to perform an auto-classification without the requirement of expert decision.

IV. CONCLUSION

To the authors knowledge, this is the first study to investigate multi-label classification in HV fault diagnosis using the naive Bayes approach. Feature extraction techniques were implemented to retrieve important and non-redundant information and to enhance the classification performance. Here, two time series based methods known as 1D-HOG and 1D-LBP were employed. Experimental results, on a dataset of three fault related signals and the combinations of thereof, demonstrated a successful classification on all evaluation metrics. The work carried out in this paper is novel in the HV field and will be considered in future work for experiments on a larger dataset with varied HV faults. The two proposed feature extraction methods are simple and low in computation making them ideal to implement in an instrument for industrial application. However, in future work more feature extraction techniques will be investigated with the aim to improve the classification performance. A possible limitation to this work is the imbalance in data segments. This issue will also be investigated in the future by looking at the cost function or weighing the unique labels.

REFERENCES

- [1] J. E. Timperley, D. Buchanan, and J. M. Vallejo, "Electric generation condition assessment with electromagnetic interference analysis," *IEEE Transactions on Industry Applications*, vol. PP, no. 99, pp. 1–1, 2017.
- [2] F. Kreuger, *Industrial High Voltage: 4. Coordinating, 5. Testing, 6. Measuring*. Delft, The Netherlands: Delft University Press, 1992.
- [3] J. E. Timperley and J. M. Vallejo, "Condition assessment of electrical apparatus with emi diagnostics," *IEEE Transactions on Industry Applications*, vol. 53, no. 1, pp. 693–699, Jan 2017.
- [4] I. Mitiche, G. Morison, A. Nesbitt, M. Hughes-Narborough, B. G. Stewart, and P. Boreham, "Classification of emi discharge sources using time-frequency features and multi-class support vector machine," *Electric Power Systems Research*, vol. 163, pp. 261 – 269, 2018.
- [5] I. Mitiche, G. Morison, A. Nesbitt, M. H-N., B. G. Stewart, and P. Boreham, "Classification of partial discharge signals by combining adaptive local iterative filtering and entropy features," in *IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP)*, Oct 2017, pp. 335 – 338.

- [6] I. Mitiche, G. Morison, A. Nesbitt, P. Boreham, and B. G. Stewart, "Classification of partial discharge emi conditions using permutation entropy-based features," in *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug 2017, pp. 1375–1379.
- [7] I. Mitiche, G. Morison, A. Nesbitt, M. Hughes-Narborough, B. G. Stewart, and P. Boreham, "Classification of partial discharge signals by combining adaptive local iterative filtering and entropy features," *Sensors*, vol. 18, no. 2, 2018.
- [8] M.-L. Zhang, J. M. Pea, and V. Robles, "Feature selection for multi-label naive bayes classification," *Information Sciences*, vol. 179, no. 19, pp. 3218 – 3229, 2009.
- [9] J. Zhao and L. Itti, "Classifying time series using local descriptors with hybrid sampling," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 623 – 637, 2016.
- [10] N. Chatlani and J. J. Soraghan, "Local binary patterns for 1-d signal processing," in *2010 18th European Signal Processing Conference*, Aug 2010, pp. 95–99.
- [11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, June 2005, pp. 886–893.
- [12] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul 2002.
- [13] *CISPR/CIS/A - Radio-interference measurements and statistical methods*, IEC EMC 33.100.10 - Emission, 2015.