PROCESS ANALYTICS AND MACHINE LEARNING TO PREDICT ARC LOSS IN AN ELECTRIC ARC FURNACE

*L. D. Rippon, I. Yousef, and R.B. Gopaluni
University of British Columbia
2332 Main Mall, Vancouver, British Columbia, Canada V6T 1Z4
*(*Corresponding author: leeripp@chbe.ubc.ca)*

B. Hosseini, J.F., Beaulieu, and C. Prévost
BBA Engineering Consultants
375 Sir Wilfrid Laurier Blvd., Mont-Saint-Hilaire, Quebec, Canada J3H 6C3

S. L. Shah
University of Alberta
9211 - 116 Street NW, Edmonton, Alberta, Canada T6G 1H9

ABSTRACT

Stable smelter operation is critical for successful production of base metals from particulate ore. This work studies the operation of an industrial direct current electric arc furnace that operates as a smelter in a large-scale metallurgical process. Specifically, unexpected loss of the plasma arc is an important unresolved problem with a significant impact on the production efficiency of the process. Moreover, given that electric arc furnaces are highly energy intensive units, even minimal improvements to the overall production efficiency represent meaningful reductions in the environmental footprint of the process over the lifetime of operation. A predictive inferential sensor is proposed to avoid high risk operating regimes and reduce the overall environmental footprint of the process. The alarm identifies high risk situations and instructs operators to take corrective actions to avoid the loss of arc. Large amounts of historical industrial process data have been collected, pre-processed and leveraged in a cross-validated supervised learning framework that trains the inferential sensor model. New data, previously unseen by the model, are drawn from the historical database and used to test the ability of the model to generalize. This work showcases our progress to date including the training, validation and testing of competing inferential sensor models and their ability to predict arc loss on industrial data.

KEYWORDS

Classification, deep learning, electric arc furnace, fault detection, machine learning, process analytics

INTRODUCTION

Throughout the last decade an immense amount of attention has been drawn to the field of artificial intelligence (AI). In addition to algorithmic and computational advances (e.g., graphical processing units), a significant portion of that attention stems from a series of breakthroughs in a domain of AI known as machine learning (ML). In fact, it is a sub-domain of ML, referred to as deep learning, that is responsible for a great deal of the recent buzz around AI. Deep learning is an emerging paradigm in ML that broadly involves using large amounts data to train neural networks with multiple hidden layers in order to obtain more expressive function approximations. While deep learning is commonly associated with pedagogical supervised learning examples such as classifying handwritten digits (c.f., MNIST) or images of cats and dogs, it is in fact consequential to unsupervised learning and reinforcement learning (Sutton & Barto, 2018). Deep learning techniques have demonstrated state of the art results on pattern recognition problems (LeCun, Benjio &

Hinton, 2015). In this work, we focus on using advanced techniques from supervised learning to increase the production efficiency and reduce the environmental footprint of a large-scale metallurgical process.

In supervised learning, matrices of feature vectors (or inputs) are presumed to provide contextual information for vectors of outputs (or targets). The term is inspired by viewing the labelled outputs as providing supervision to the ML algorithm (Goodfellow, Bengio & Courville, 2017). For instance, given tuples of inputs X and outputs Y, supervised learning algorithms aim to use large quantities of labeled training data to learn a mapping from $X \longrightarrow Y$ that accurately predicts Y given new observations of X. Outputs are continuous for regression problems and discrete (or categorical) for classification problems. In the context of metallurgical processing, using historical data to predict the amount of energy expended per unit of metal produced is an example of regression. Alternatively, a relevant example of supervised learning for classification is learning a mapping from the historical process data to labelled process upsets/faults in order to predict future process faults. In this classification example the output is discrete because it either indicates that there will be a fault (e.g., $Y = 1$) or that there will not be a fault (e.g., $Y = 0$). In fact, this predictive classification problem is the focus of this work and the fault we are trying to predict is the unexpected loss of plasma arc in a direct current (DC) electric arc furnace (EAF).

Consider an industrial mining and metallurgy operation that continues production until a fixed deposit of saprolite ore has been extracted and processed. In this context, assuming a carbon intensive energy source, it is imperative to minimize the expended energy per unit of on-spec product in order to minimize the environmental footprint over the lifetime of the operation. Our work helps to support this objective by investigating the use of techniques from both traditional process analytics and advances in ML to predict the onset of arc loss such that operators can take remedial actions and maintain efficient operation. The remainder of this paper is organized by first providing a description of the process and a definition of the problem. The problem definition is followed by context into how the data was prepared for analysis which leads into a brief presentation of the selected algorithms. Finally, we present the results along with some conclusions and avenues for further investigation.

PROCESS DESCRIPTION AND PROBLEM DEFINITION

In this section we describe the metallurgical process in question and the accompanying fault that we aim to address with our proposed methods. Our intention is to be as clear as possible about the process and the problem but please note that some details are omitted due to confidentiality. A simplified depiction of the broader mining and metallurgical operation is illustrated in Figure 1. Drilling is performed to characterize the deposit and ore is extracted from an open pit mine using trucks and hydraulic shovels. Dump trucks transport the ore to the ore preparation plant where it is screened to remove waste rocks and crushed to prepare it for further processing. The crushed ore is then conveyed to the metallurgical plant (green) which is the process of interest for this work. Initial operations in the metallurgical plant are milling and drying which are conducted using hammer mill flash dryers. Dried ore is fed to a series of calcining cyclones where it is dehydrated before being sent to fluidized bed reducers to remove oxides and improve the electrical efficiency of the subsequent smelting operation in the DC EAF (Meihack, 1986; Jones, 2014; Keskinkilic, 2019). The DC EAF is the operation that is exhibiting unexpected faulty behavior that we aim to predict.

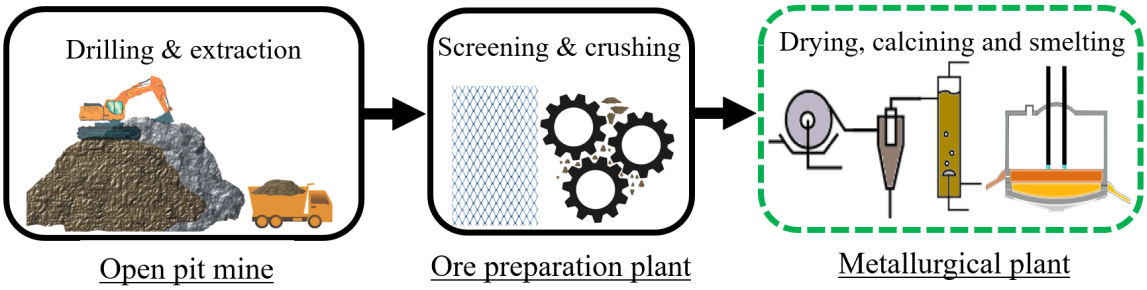

Figure 1. Simplified illustration of the relevant mining and metallurgical processes

Upstream processing stages convert the raw ore into a fine particulate feed that is ready to be fed to the DC EAF for smelting. As depicted in Figure 2, an open plasma arc spans from the cathode to the anode providing energy required to maintain temperatures of the slag and alloy above 1400°C, depending on the composition of the slag (Lagendijk & Jones, 1997). The cathode consists of two hollow graphite electrodes while the anode is the molten slag. The roof and side walls of the EAF are cooled with water to maintain a safe structural temperature (Hurd & Kollar, 1991). Multiple ports along the roof feed the furnace while the slag and alloy are tapped from the launders intermittently (Kotze, 2002). As an energy intensive unit, it is critical to operate the EAF in a stable manner to maximize production efficiency. Unexpected loss of the plasma arc is a recurring process fault that significantly impacts the electrical efficiency of the EAF.

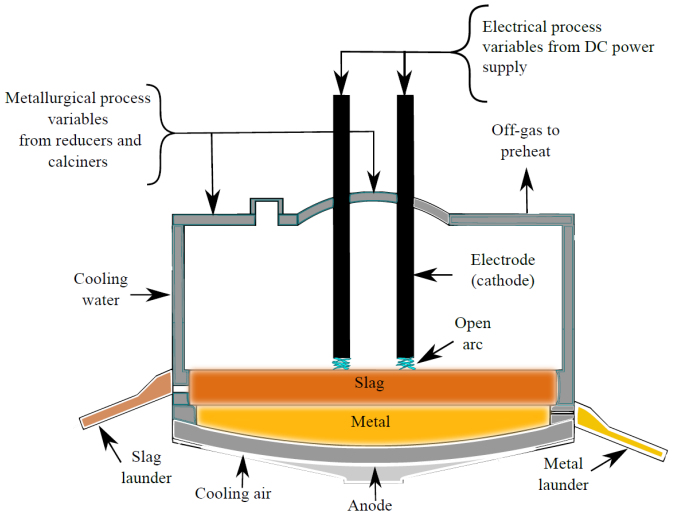

Figure 2. Schematic of a direct current electric arc furnace.

The open plasma arc is developed when the electric current between the graphite electrodes (i.e., cathode) and the surface of the molten slag (i.e., anode) is charged. During routine operation, the power sent to the furnace is relatively stable and the electrodes are held in a stationary position. However, due to process disturbances (e.g., varying slag composition, upstream disturbances in furnace feed, and electrical noise from the DC power supply) the power applied to each graphite electrode can vary significantly over time. The power fluctuations measured at each electrode are responsible for the loss of plasma arc. In fact, the arc loss problem is defined when all three of the following conditions are satisfied: i) prior to arc loss the power is stable (i.e., the standard deviation of the power applied to an electrode is less than 2 MW over a period of approximately 11.5 minutes), ii) a precipitous power drop (i.e., at least 10 MW within thirty six seconds), and iii) the power recovers (i.e., to within 5 MW of the original stable value within approximately 10 minutes). In this work those three conditions are rigorously applied to one year of operating data in order to generate a labelled data set that is amenable to supervised learning. More information pertaining to data preparation is provided in the following section.

Once the arc loss events are detected and labelled it is possible to filter and visualize periods of operation that are either smooth or faulty. In Figure 3, a visual comparison is provided with smooth operation in the top plot and faulty operation in the bottom plot. The faulty operation plot illustrates a cascade of arc losses that significantly impact the furnace feed rate, resulting in reduced energy efficiency and increased environmental footprint per unit of on-spec product. In this work, our objective is to be able to predict the onset of an arc loss event with an accuracy above 75%. Accuracy is defined as the ratio of true predictions (either negative or positive) to the total number of predictions. The predictive arc loss alarm needs to trigger five to ten minutes before the actual arc loss event such that operators have enough time to initiate corrective actions. A target of 75% is set for this preliminary analysis for two reasons. Firstly, predicting arc loss five

minutes before it occurs is no easy task and achieving an accuracy above 75% provides important indication that there is actionable information to be leveraged in the data. Secondly, depending on the asymmetry of the operating cost associated with acting upon false positives or ignoring true positives a 75% accuracy could represent an economically meaningful result for this operation.

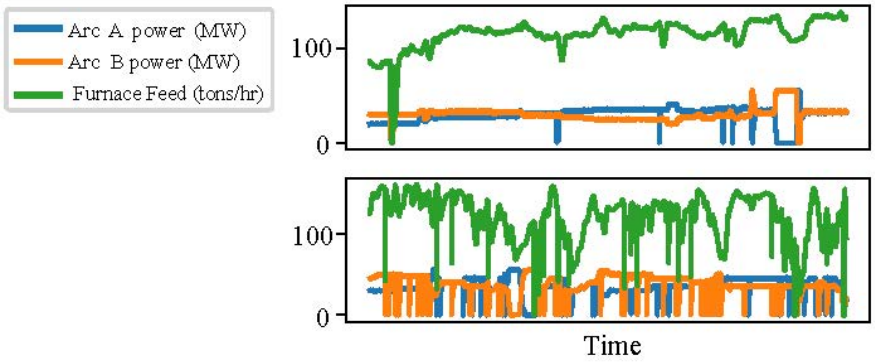

Figure 3. Visualizing the impact of the process fault on the operation.

It is important to note that these results are obtained on a balanced dataset and thus the reported accuracies weigh true positives and true negatives equally (this will be discussed further in what follows). To address these challenges, we investigate both traditional linear methods for process analytics as well as propose some advanced nonlinear ML methods. Prior to building any models we must first transform the raw process historian data into a form that is amenable for further statistical analysis.

PREPARING THE DATA FOR STATISTICAL ANALYSIS AND SUPERVISED LEARNING

Raw industrial data is often riddled with problematic artefacts and the industrial EAF dataset is no exception. In fact, in all industrial applications, it is necessary to apply at least one data pre-processing technique (Famili, Shen, Weber & Simoudis, 1997). Data pre-processing aims to transform large amounts of raw, messy and noisy process data into a form that is suitable for different statistical ML algorithms. Data preparation is an onerous yet essential aspect in any ML workflow. In supervised learning, data pre-processing aids in eliminating data that is inadequate, irrelevant, extraneous, redundant or confounding in order to yield a better and faster learning process (Kotsiantis, Kanellopoulos & Pintelas, 2007). Moreover, the quality of data pre-processing is directly proportional to the quality of any ML model as the final output of pre-processing is the training data fed to the ML algorithm (Nawi, Atomi & Rehman, 2013). In what follows we describe some of the steps taken while preparing the data including data cleaning, data segmentation, and feature scaling.

Data cleaning

In this work, data cleaning includes structuring the data by unifying sampling rates, labelling the arc losses (discussed before), converting improper data types, removing outliers, and imputing missing values. Firstly, the sampling rate is inconsistent among measured variables (or features) as illustrated by the columns in the left-hand side of Figure 4. Hence, the most frequently sampled feature from each daily spreadsheet is identified and the corresponding time stamp is used as a unifying reference for all the other process features. The less densely sampled features are re-sampled based on the unified reference time by propagating the values forward to fill missing values (i.e. forward fill or zero order hold). This enables us to form a data-frame matrix with an equivalent number of rows per column and a unified temporal indexing of rows (i.e., only a single time stamp column needs to be kept). The next data cleaning task is to address the data type as the variables within each column must be of the same data type (i.e., either numerical or categorical). However, some columns exhibit inconsistent data types such as bad inputs or not a number (NaN) values (e.g., ‘tag not found’ or ‘bad quality’). To eliminate this improper data a dictionary of the

keywords associated with bad inputs is formed and every input that includes one of those keywords is identified and replaced by NaN values. Finally, the output label for every sample in the dataset was chosen based on the three conditions introduced in the previous section. If the sample of the unified time series satisfies all the three conditions, then it is labeled as constituting an arc loss fault.

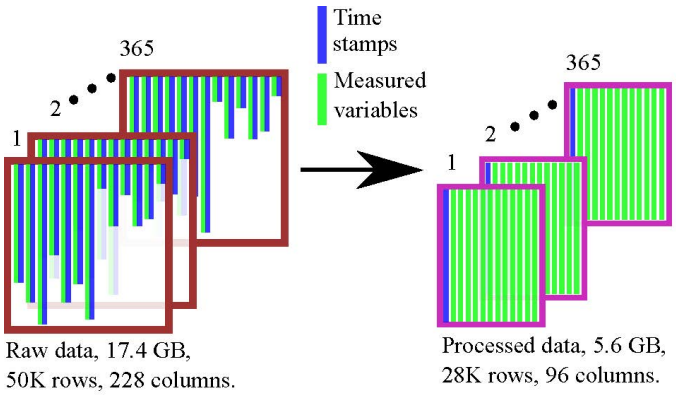

Figure 4. Processing the raw data into a form that is amenable for statistical analysis.

Data segmentation

Class imbalance (i.e., roughly the ratio of faulty samples to normal samples) is a well-known challenge in classification that occurs when the frequency at which classes are observed varies widely among classes. Specifically, in a binary classification problem, a common case of imbalanced classes arises when most samples fall into one class while only a small set of observations represent the other class (often the class of interest). Many ML and deep learning methods perform poorly under class imbalance. For instance, standard classification algorithms tend to under-value the minority class because their primary objective is to achieve the highest accuracy along the whole range of data (Krawczyk, 2016). In this work, over 99% of the data represents the positive class (i.e., normal operation) while less than 1% of the data corresponds to the negative class of interest (i.e., arc loss). Techniques such as re-sampling, feature engineering, and using alternative performance metrics are commonly suggested remedies to address class imbalance. In this work, because we are fortunate to have an abundance of historical arc loss data, simple under-sampling of the majority class is performed. Fifty-five-minute consecutive segments of data that capture observations within five to sixty minutes prior to every arc loss event are collected and labeled as a negative class. Subsequently, the positive segments are formed by sampling an equivalent number of segments that correspond to the fifty-five consecutive minutes observed during a smooth period of operation. The output of the data segmentation is a balanced dataset with 2136 data segments of 1100 samples (i.e., fifty-five minutes' worth of process data) and 96 process variables.

Feature scaling

The variables collected from industrial processes are usually measured with different units and different scales of magnitude that can result in unequal contributions to the analysis. Ultimately these inconsistent scales may induce significant bias in the learned models. Feature scaling is a form of data pre-processing that aims to transform the variables into a unified and common scale. The main objective of feature scaling is to reduce the bias with respect to discriminating the output class that is caused by variables with much larger magnitudes (Sola & Sevilla, 1997). In this work, the features are rescaled to form a Gaussian distribution with zero mean and unit standard deviation. The scaled values were calculated using the following equation for a given column of data in our dataframe matrix

$$x_{scaled} = \frac{x - \bar{x}}{\sigma}, \tag{1}$$

where x is a sample within a column, σ is the standard deviation along that column, and $\bar{x}$ is the mean value along that column. Equation 1 is applied along each column where the mean and standard deviation variables are different for each column.

During data preparation, there are many degrees of freedom with respect to the strategies employed. The outcome of the data analysis is sensitive to these processing decisions. Our philosophy for this study is to maintain as much of the raw information as possible while limiting the amount of external information inserted through preparatory operations. Employing this philosophy is especially beneficial for preliminary exploration of the data as it allows for a broad view to capture potential fault mechanisms throughout the process. This is apparent from our decisions such as mapping variables to the most densely sampled time-series as well as using a simple zero-order hold to fill missing entries as opposed to applying more sophisticated imputation techniques. Alternative data preparation strategies are of interest in future work.

METHODOLOGY

The binary classification nature of this study offers a variety of potential algorithms. As a preliminary study the modus operandi was comparative analysis to spotlight areas for further investigation. For contrast, our analysis broadly focused on two categories of methods, i.e., a) traditional linear process analytics tools and b) deep learning methods. These methods are compared with the cleaned, segmented and scaled data that is provided by the large scale industrial metallurgical plant previously described. Each data segment contains fifty-five minutes of data that occurs either five minutes before an arc loss event ($Y = 1$) or five minutes before an extended period of smooth operation ($Y = 0$). The goal of each classifier is to collect an unseen fifty-five-minute segment and predict whether there will be an arc loss event after five minutes. The chosen techniques are introduced widely in literature and their theoretical breakdown is beyond the scope of this report. Instead, a brief description of each method is provided along with some references for the interested reader.

The traditional linear methods for process data analytics that are selected for this study include logistic regression (LR), linear support vector machine (L-SVM), principal component analysis (PCA), and partial least squares (PLS). The sigmoid function, with outputs bounded between zero and one, is used for binary classification with LR with roots dating back to the early 19th century and applications in autocatalysis (Cramer, 2002). This work applies a soft margin SVM which was introduced in 1995 and aims to find a separating hyper-plane between the two classes of data (Cortes & Vapnik, 1995). In the context of fault detection and diagnosis SVMs have been applied for model-plant mismatch detection in paper machine control systems (Lu et al., 2017; Rippon, et al., 2019). In the early 20th century PCA was introduced with the objective of decomposing a multivariate dataset into a basis set of linearly uncorrelated principal components (Pearson, 1901). Finally, PLS is a slightly more modern method that was introduced in the 1970s and has been used extensively in process systems engineering to perform fault detection (Wold, 2001).

Models for PCA and PLS were trained and validated over a full spectrum of component numbers. The highest validation accuracies were obtained with 9 components for PLS, 80 components for PCA with L-SVM, and 86 components for PCA with LR. The high number of PCA components raises concerns about over-fitting but nonetheless, to be consistent, the models with the highest validation accuracies were selected for testing. For the LR and L-SVM classifiers, L2 regularization was applied but kernel based SVM methods were not considered here. Determining the relative importance of the 96 process variables with respect to predicting arc loss is of interest for diagnosis. Preliminary investigation into feature importance used a random forest classifier and tracking the mean decrease in accuracy associated with removing specific features. Notable features of importance include power measurements on the electrodes and temperature measurements on the furnace feed ports. Comprehensive follow-up studies are necessary to further quantify feature importance and move beyond arc loss prediction into arc loss diagnosis and ideally prevention.

Two of the techniques analyzed in this study fit within the context of deep learning methods. The first deep learning technique is a deep fully connected artificial neural network (ANN) while the second

method relies on a deep convolutional neural network (CNN). Biological neurons in the human brain serve as the inspiration for ANNs which have proven useful as versatile classification tools that can accurately approximate high-dimensional nonlinear functions (Jain, Mao, & Mohiuddin, 1996; Tu, 1996). In this work the ANN implemented is a multi-layered fully connected perceptron model with five hidden layers, 50 neurons per layer, rectified linear unit (ReLU) activation functions and a batch size of 100. During training a non-exhaustive set of additional network architectures was studied (e.g., 100 neurons per layer and ten hidden layers) with the prior configuration selected based on the validation accuracy. The ANN takes a 55-minute segment of data as an input and uses a softmax activation function in the output layer to produce classification predictions.

In the mid 1990s CNNs rose to prominence by achieving state of the art results in image classification tasks such as character recognition (Le Cun & Bengio, 1994). To apply CNNs the segments of input data are fed as images to the CNN architecture which contains two convolution layers and two pooling layers with ReLUs as activation functions. After the convolution and dropout layers the CNN classifier has a single dense layer with 256 neurons. As before, different configurations were considered during training by coarsely varying the number and type of layers, the number of neurons per layer and the size of the kernel in each layer. The CNN method utilizes the very popular adaptive moment estimation (Adam) optimizer. Preliminary studies with recurrent neural networks (RNNs) led to significant over-fitting but nonetheless these remain promising avenues that warrant further investigation. A more exhaustive account of network architectures is desirable for future studies such as the use of neural architecture search.

The performance metric used to compare each method is the classification accuracy of that method exhibited on the testing dataset. The classification accuracy is simply the number of correct classifications divided by the total number of classification estimates. Recall that for this study the classes are approximately balanced, i.e., an equivalent number of faulty and smooth segments exist. This is only approximate because the data is only evaluated on the testing dataset. The entire prepared dataset is taken and split into three categories with 70% of the data for training the classifiers, 20% of the data for validating the classifier hyper-parameters and 10% of the data for testing the classifiers. A straightforward holdout strategy is employed for cross-validation which implies that each configuration of each method is trained on the 70% training data and the configuration with the highest performance on the 20% validation data is selected for final testing on the 10% testing set. The results presented in the following section are reported on the final testing dataset.

RESULTS AND DISCUSSION

The classification accuracy of each method on the testing set is illustrated in Figure 5. The results presented in this section corresponds to the performance of each classification algorithm with respect to predicting unseen testing set samples. Relatively consistent values were observed for the training, validation and testing accuracies of the linear methods. The deep learning methods, however, report a much higher training accuracy relative to their testing accuracy as a result of their expressivity and capacity to overfit. This is monitored during training using a validation loss. Initial analysis indicates that the PLS alarm sensitivity leans slightly towards missing alarms (i.e., false negative) as opposed to alarming during non-events (i.e., false positives). Again, it is important to state that these results are obtained on a balanced dataset (i.e., an equal number of events with and without arc loss). An elaborative quantification of alarm selectivity will be undertaken to tune candidate methods in future studies.

Most of the methods, applied in this work, require tuning a combination of multiple hyper-parameter, which is challenging; therefore, we do not claim to have exhausted all possible combinations, but we believe that the selected hyper-parameters provide an insightful high level comparison of the chosen methods. Blue bars on the left side of Figure 5 show the testing accuracy of the traditional linear methods for process analytics whereas the red bars on the right side of Figure 5 show the testing accuracy of the deep learning techniques. Note that PCA is paired with either LR or L-SVM because PCA is simply an unsupervised dimensionality reduction technique that still requires a supervised classifier.

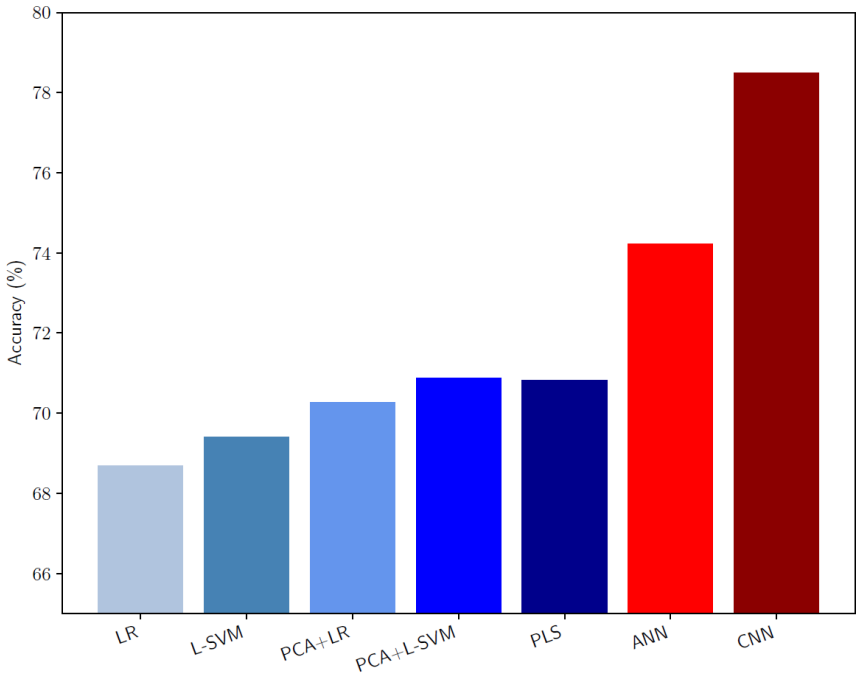

Figure 5. Summary of the predictive classification results on the testing data set

Figure 5 illustrates the superior performance of the deep learning techniques relative to the traditional linear classification methods in predicting the class of unseen fifty-five-minute data segments from the testing set. However, since this is a binary classification application, a classification accuracy of 50% is the baseline for comparison; therefore, all the classification methods deliver meaningful predictive performance. It can be concluded that our dataset has notable non-linearity (i.e. hidden non-linear relationships between the process variables) which explains the better classification accuracies of the deep learning techniques that deal with non-linearity. Moreover, it is apparent from these results that the use of PCA for dimensionality reduction improves the performance of the LR and L-SVM classifiers. This is presumably due to a reduction in over-fitting on the training data that leads to better generalization during testing by performing the classification on the principal components.

Ultimately, the superior performance of the deep learning methods and the benefits of using dimensionality reduction are consistent with expectations drawn from literature. Apart from deep learning, there are nonlinear extensions to the traditional methods (e.g., kernel based SVM) that could be part of a future comprehensive comparison. Appeal to scalability and empirical performance motivated our selection of deep learning (specifically CNNs) as a nonlinear representation learning technique to compare to the traditional linear methods. Model interpretability is an open problem for advanced ML techniques such as deep learning. Admittedly it is a problem that we do not address in this study as we are fortunate that operators can take actions to stabilize the furnace without knowing the exact cause of arc loss. Developing and validating a robust arc loss risk indicator is an important first step that justifies further exploration of the data to determine underlying causal mechanisms which may yield rigorous control policies to avoid arc loss.

Achieving an accuracy greater than 78% on the balanced testing dataset using CNNs is very promising for this preliminary study. Although these early results are promising it needs to be stated clearly that further development is necessary to validate these techniques in settings that better simulate a production environment. Real-time deployment can be simulated by streaming historical data to help prototype computationally optimized models that provide sufficiently high rates of inference. Fortunately, the computationally demanding aspect of training the representation and prediction models can be performed

offline. Finally, subsequent models will require validation and testing on larger datasets that exhibit realistic class imbalance.

CONCLUSIONS AND FUTURE WORK

The use of process data analytics and ML is an emerging research area that offers significant benefits to the modern process industry where an enormous amount of informative data is collected and archived. This work is a small part of a larger movement to migrate advanced data analytics techniques from statistics and computing sciences to process industries. The first step taken in this work was pre-processing the raw data collected and measured from the EAF process to create structured clean data in a form that an ML algorithm can efficiently parse. It is important to note that the incorporation of process domain knowledge is critical for defining the problem, labelling the arc loss events, preparing the data and interpreting the results.

Preliminary analysis of the experimental results indicates that developing improved CNN implementations that leverage dimensionality reduction is a promising direction for further studies. A comprehensive exposition of the experimental results including confusion matrices for each technique will be presented in the studies to follow. Moreover, future work aims to investigate sophisticated techniques for learning representative features such as the use of autoencoders. Formalizing the decisions and actions taken during data preparation as a hyperparameter selection problem is of interest in future work as well. Finally, as mentioned before it will be insightful to test these algorithms in a production setting which will certainly introduce additional challenges.

REFERENCES

Cramer, J. S. (2002). The origins of logistic regression.

Famili, A., Shen, W., Weber, R., & Simoudis, E. (1997). Data preprocessing and intelligent data analysis. *Intelligent Data Analysis*, *1*(1-4), 3-23.

Hurd, D., & Kollar, J. (1991). Direct current EAF-a review. *Steel Times*, *219*(5), 245.

Goodfellow, I., Bengio, Y., & Courville, A. (2017). *Deep learning*. Cambridge, Mass: The MIT Press.

Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, *29*(3), 31-44.

Jones, R. T. (2014). DC arc furnaces—past, present, and future. In Celebrating the Megascale (pp. 129-139). Springer, Cham.

Keskinkilic, E. (2019). Nickel laterite smelting processes and some examples of recent possible modifications to the conventional route. *Metals*, *9*(9), 974.

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2007). Data Preprocessing for Supervised Leaning. *International Journal of Computer, Electrical, Automation, Control And Information Engineering*, *1*(12).

Kotze, I. J. (2002). Pilot plant production of ferronickel from nickel oxide ores and dusts in a DC arc furnace. *Minerals Engineering*, *15*(11), 1017-1022.

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221-232.

Lagendijk, H., & Jones, R. T. (1997). Production of ferronickel from nickel laterites in a DC arc furnace. Mintek.

Le Cun, Y., & Bengio, Y. (1994, October). Word-level training of a handwritten word recognizer based on convolutional neural networks. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition, vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5)* (Vol. 2, pp. 88-92). IEEE

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436-444.

Lu, Q., Gopaluni, R. B., Forbes, M. G., Loewen, P. D., Backström, J. U., & Dumont, G. A. (2017). Model-Plant Mismatch Detection with Support Vector Machines. *IFAC-PapersOnLine*, *50*(1), 7993-7998.

Meihack, W. F. A. T. (1986). The potential role of fluidized beds in the metallurgical industry. *Journal of the Southern African Institute of Mining and Metallurgy*, *86*(5), 153-160.

Nawi, N., Atomi, W., & Rehman, M. (2013). The Effect of Data Pre-processing on Optimized Training of Artificial Neural Networks. *Procedia Technology*, *11*, 32-39.

Pearson, K. (1901). Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *6*(2), 559.

Rippon, L. D., Lu, Q., Forbes, M. G., Gopaluni, R. B., Loewen, P. D., & Backström, J. U. (2019). Machine Direction Adaptive Control on a Paper Machine. *Industrial & Engineering Chemistry Research*, *58*(26), 11452-11473.

Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, *44*(3), 1464-1468.

Sutton, R., & Barto, A. (2018). Reinforcement Learning: An Introduction. MIT press.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, *49*(11), 1225-1231.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, *58*(2), 109-130.