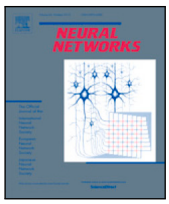
Neural Networks 174 (2024) 106245

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



Full Length Article

Modeling failures in smart grids by a bilinear logistic regression approach

Enrico De Santis *, Antonello Rizzi

Department of Information Engineering, Electronics and Telecommunications, University of Rome "La Sapienza", Via Eudossiana 18, Rome, 00184, Italy

ARTICLE INFO

Keywords:
Fault recognition
Fault classification
Smart grids
Bilinear model
Complex systems

ABSTRACT

Modeling and recognizing events in complex systems through machine learning techniques is a challenging task. Especially if the model is constrained to be explainable and interpretable, while ensuring high levels of accuracy. In this paper, we adopt a bilinear logistic regression model in which the parameters are trained in a data-driven fashion on a real-world dataset of power grid failure data. The bilinear white-box model – grounded on a specific neural architecture – has been proven effective in classifying faulty states with a performance comparable to several classifiers in technical literature. Additionally, the low computational complexity of the bilinear model, in terms of the number of free parameters, allows gaining insights into the fault phenomenon correlating the events that impact the power grid (exogenous causes) with its constitutive characteristics, thence eliciting the relational information hidden in the data. The proposed model is also able to estimate a vulnerability vector that can be associated, as a suitable characteristic "label", to power grid components, opening the way, as will be deeply demonstrated in the following, not only to predictive maintenance programs or condition monitoring tasks but also to risk assessment and scenario analyses in line with the explainable AI paradigm.

1. Introduction

With the advent of modern machine learning techniques, within the Artificial Intelligence umbrella, the possibility of modeling and controlling complex systems increased considerably. At the same time, complex systems theory can be a useful tool for providing new insights into machine learning approaches (Tang, Kurths, Lin, Ott, & Kocarev, 2020). Nowadays machine learning techniques are adopted not only in computer science but also across a range of industries concerned with data-intensive issues, such as consumer services, the diagnosis of faults in complex systems, and the control of logistics chains (Jordan & Mitchell, 2015). In any case, standard machine learning techniques are based on the construction of learning datasets made up of patterns which in most cases are a set of variables or measures on a process or system. Often these measurements, after the usual data preprocessing phases, consist of n-tuples of real numbers, each of which reflects a specific characteristic considered important, therefore correlated with the output of the system or process to be modeled (Bishop & Nasrabadi, 2006). Moreover, it is known that a classification algorithm, whether it is based on the connectionist paradigm or on explicit geometrical properties such as a suitable similarity measure between patterns, is a data-driven procedure to learn an input–output mapping with a suitable algorithmic architecture or model, the choice of which depends not only on performance but also on other properties such as the so-called "explainability". In fact, in recent research trends, a desirable property of a learning model is the one to be "interpretable" and "comprehensible" (Martino, De Santis, & Rizzi, 2020). In other words, several modern research programs, such as the one promoted by DARPA (Gunning & Aha, 2019), endeavor to create AI systems whose learned models and decisions can be understood and appropriately trusted by end users. According to DARPA "The XAI program's goal is to create a suite of new or modified machine learning techniques that produce explainable models that, when combined with effective explanation techniques, enable end users to understand, appropriately trust, and effectively manage the emerging generation of AI systems" (Gunning & Aha, 2019). If on the one hand, the endeavor in synthesizing explainable models is not so recent if we consider decision trees, on the other, the need for explainability is proportional to the power of modern deep learning architectures, which remain in most cases black-box or gray-box models. Moreover, the explainability of a machine learning model is usually inverse to its prediction accuracy, that is the higher the prediction accuracy, the lower the model explainability (Gunning & Aha, 2019; Xu et al., 2019).

In complex systems modeling through machine learning techniques and specifically in describing the state of the system, in some cases, there is the possibility of collecting a set of measures that can be grouped according to some semantic specification. This is the case of modeling the state of a system in predictive maintenance and condition

* Corresponding author.
E-mail addresses: enrico.desantis@unirma1.it (E. De Santis), antonello.rizzi@unirma1.it (A. Rizzi).

https://doi.org/10.1016/j.neunet.2024.106245
Received 18 September 2023; Received in revised form 30 December 2023; Accepted 13 March 2024
Available online 15 March 2024

monitoring (but also in biomedical applications (Wang, Wang, Hu, & Sorrentino, 2015)), where in determining the probability of failure or outage it is possible to individuate "constitutive" slow-varying variables that are related to the material state of the system or equipment and "exogenous" (or also "endogenous") fast-changing variables, which are considered as "forces" that impact the state of the system itself. For example, a failure event in a power grid can depend on the constitution of the component subject to a fault and on several events impacting its state, for example, the climatic or weather conditions (Guikema, Davidson, & Liu, 2006) (think to flooding or lightning strikes) or the electric load determined by human behavior in using electric devices. Hence, in synthesizing a condition monitoring system but also in constructing a model capable of generating scenario analyses, having available smart sensors spread on the network can be useful to semantically distinguish measures related to components (*constitutive parameters*) and measures on variables that impact to the system (*exogenous variables*). In separating semantic variables it is possible to build a predictive *relational* model that can be both lightweight and in line with the explainable AI paradigm. In this work, we propose a framework based on a bilinear model formally similar to the well-known logistic regression classifier for recognizing failures in a real-world MV power grid with the following characteristics that allow:

- to suitably estimate the vulnerability of components – characterized by well-suited constitutive parameters – in relation to given exogenous causes, i.e., the weather conditions and the electric load measured by smart sensors;
- to estimate the probability of failure;
- to provide a clear interpretation of the obtained model;
- to open toward scenario analyses for network long-term planning strategies gaining insights from the model.

Hence, the following study has also a methodological aim, that is constructing a learning model maintaining the trade-off between a very good level of interpretability and high accuracy with the possibility of extracting useful information from the model to gain insight into the fault phenomenon. It is emphasized that the approach is sufficiently general and can be applied in all cases where the causes can be disambiguated from the constitutive parameters. For example, in the field of medicine, the use of a drug in various clinical conditions can be modeled to evaluate its effects. The approach is based on defining the vulnerability vector as a linear function of the constitutive parameters (through a suitable parameter matrix) and estimating the probability of faults as a non-linear function of the dot product of the aforementioned vulnerability vector and the vector of exogenous causes. Given a dataset of failures where data patterns can be semantically grouped in constitutive parameters and exogenous causes the primary goal is estimating the (trainable) parameters of the linear combination (that define the vulnerability) that better align with exogenous causes. In other words, net to the output non-linearity – that will be a sigmoid function – the probability of a failure reaches its maximum value if the inner product is maximum, that is the vulnerability vector and the exogenous causes are aligned. This simple model allows estimating a (rectangular) matrix, which we call "correlation matrix", that relates constitutive parameters and exogenous causes and can be directly interpretable. Moreover, the matrix (along with a bias vector) allows for estimating the vulnerability related to a component given a set of constitutive parameters. Furthermore, a spectral decomposition of the matrix, performed by SVD, can be exploited to extract the main modes of the relation between constitutive parameters and exogenous cause in determining the vulnerability and, finally, the probability of fault. Once learned the parameters of the model, they can be used to visualize probability landscapes varying entries of the constitutive parameters or of the exogenous causes, with the interesting possibility of building a framework not only for detecting failures but also for carrying out data-driven risk analysis and network planning programs through a suitable knowledge discovery paradigm. In fact, the possibility of estimating the vulnerability of the components and the probability of failure allows us to carry on a risk analysis through the product of the probability of failure and the impact of the causes on the components, given their vulnerability. Moreover, utilizing machine learning for risk analysis in smart grids leads to more precise predictive maintenance, significantly reducing equipment failures and extending their operational life. It enables efficient resource allocation by identifying high-risk components, thereby saving costs and enhancing grid reliability. This approach bolsters the grid's resilience, particularly against extreme conditions and load changes, through early vulnerability identification and mitigation. The methodology's data-driven nature ensures informed, effective management decisions. Additionally, its capability for continuous adaptation to new data and changing grid conditions guarantees the smart grid's long-term stability and operational efficiency. We would like to point out that in this investigation an in-depth risk assessment analysis of the network and network planning in economic and programmatic terms is considered out of scope, while a scenario analysis will be carried out using various case studies.

To the best of our knowledge, a neural architecture similar in structure (but not identical) to the one proposed by us is known by the name "Neural Tensor Network" and is treated in Socher, Chen, Manning, and Ng (2013) and Chen, Socher, Manning, and Ng (2013). Authors build a tensor layered bilinear neuron to capture relational information in knowledge bases, extracting association and common sense reasoning from large text corpora. The goal of the approach is to be able to state whether two entities $(e_1; e_2)$ are in a certain relationship R, where entities are embedding vectors for words in the database. Hence, at least in this approach, the involved relation matrix is square and the entries do not posses a clear semantical meaning as in our setting, in which exogenous causes and consecutive parameters are composed by clear and interpretable entries (whether conditions, loads, length of cables, constituent material, etc.). A bilinear approach with similar structure and intents is proposed in Wang et al. (2015) within the biomedical ambit. Authors claim to extend the classical logistic regression scheme, where a regression vector of weights associated to variables is estimated, to a bilinear scheme to put in relation (all) features describing patients and possible patient outcomes (cases, controls) in a binary classification settings. The learning scheme is grounded on a suitable decomposition of the regression square matrix. Interestingly, the study shows how the model can be adopted to perform patient score prediction, patient risk stratification and clinical context discovery. We can state that in a slightly different manner, we translate the same aims in the predictive maintenance and condition monitoring context, within the modern Smart Grids ambit. In Shi, Xu, and Baraniuk (2014) authors use the locution "sparse linear logistic regression" for extending the logistic regression scheme to data patterns having the form of data matrices (matrices-labels pairs). Despite the name, the computational scheme is quite different from the one adopted in the current study. We decided, however, to adopt the naming "bilinear logistic regression" for our scheme because it is formally identical to the classic logistic regression, net to the learning of a regression general matrix instead of a regression vector. Finally, with the claim to prove the effectiveness of the proposed method, we carry out our experiments on real-world data measured on the MV power grid that feeds the city of Rome in Italy and we compare classification results with other standard classifiers. Moreover, we propose two different learning schemes, the plain one – named plain B-LR – which learns a single correlation matrix with a low complex model and the boosted one – named Boosted-LR (B-LR) – that is capable of learning a set of correlation matrices in parallel, trying to capture different relations between exogenous causes and constitutive parameters. In order to investigate the complexity of the dataset and the behavior of the model in terms of sparsity a parameter regularization study is performed. Additionally, due to the application context, we study a calibration scheme for output (uncalibrated) probabilities in order to have an output in line with the observed frequencies (De Santis, Arnò, & Rizzi, 2022). We present also a denormalization scheme

of the models' parameters that will be useful, together with output calibrated probabilities, to carry out a set of knowledge discovery tasks, proving the effectiveness of the overall methodology in the context of further condition monitoring, risk analysis and mid-term plan strategies in network operation.

It is worth noting that this work follows in the wake of seminal studies published by the authors in the context of fault modeling in Smart Grids – within the same project – such as in De Santis, Livi, Sadeghian, and Rizzi (2015b), De Santis, Paschero, Rizzi, and Mascioli (2018), De Santis, Rizzi, and Sadeghian (2018). In De Santis et al. (2015b) a hybrid technique was used based on a one-class classifier obtained through a clustering procedure whose underlying weighted dissimilarity was learned through an evolutionary meta-heuristic while in De Santis, Rizzi, and Sadeghian (2018) the task has been re-framed in a binary classification perspective maintaining the same methodology, with slight modifications. In De Santis, Paschero, Rizzi, and Mascioli (2018) instead a first approach to the actual estimation of the probability of failure is presented using a completely bilinear scheme, therefore without layers that model the important non-linearities and proposing a piecewise linear approximation methodology precisely to mitigate the error due to the presence of these non-linearities intrinsic to the problem at hand. In addition, in the present work, the dataset used for experiments is not only more recent, as we will be able to say, but completely re-engineered, with a different layout and completely different features. While in previous investigations we worked with dissimilarity measures capable of operating with unstructured data of a heterogeneous nature, in the current study the data are in the form of n-tuples of real numbers, allowing us to implement a lightweight data-driven model.

The current paper is organized as follows. In Section 2 it is reported a brief literature overview in the context of failure modeling and prediction in Smart Grids through machine learning. Section 3 introduces the rationale behind the approach and illustrates the dataset. Section 4 deals with the problem formulation, defining the model and providing several insights on its structure. The set of experiments is described and discussed in Section 5 while conclusions are drawn in Section 6.

2. Related works on failure prediction and modeling in smart grids

As regards the classification of faults and the impact they can have on Smart Grids, useful insight can be gained from (Krivohlava, Chren, & Rossi, 2022; Rivas & Abrao, 2020). Artificial Intelligence in its many subdisciplines has been widely used in the context of Smart Grids and, specifically, to model fault states at various levels (Bose, 2017). In technical literature, deep learning techniques have been employed to monitor the states of important components such as insulators, transformers and transmission lines (Zhang, Han, & Deng, 2018). In Zhao, Xu, Qi, Liu, and Zhang (2016) authors propose to take advantage of high-level discriminative CNNs aiming to extract the features of the insulators and identify their defects achieving high accuracy. In Xi, Feilai, Yongchao, Zhiping, and Long (2017) the fault detection of the power line is faced with a sparse self-encoding neural network, in which normalized sub-band energy of wavelet decomposition is used as the characteristic parameters. In Bangalore and Tjernberg (2015) is studied a self-evolving maintenance scheduler framework for the maintenance management of wind turbines. In the context of condition monitoring authors adopt an artificial neural network (ANN)-based approach using data from supervisory control and data acquisition (SCADA). The investigation demonstrates that the ANN-based condition monitoring program is capable of indicating severe damage in the components being monitored in advance. An approach for fault detection and classification in power transmission lines based on convolutional sparse autoencoder is proposed in Chen, Hu, and He (2016). The study deals with a system that automatically learns features from a dataset of voltage and current signals, on the basis of which a framework for fault detection and classification is created. Performance is evaluated on different scenarios and signal types. An automatic procedure, based on a genetic algorithm capable of optimizing a diagnostic system for the recognition and identification of partial-discharge (PD) pulse patterns in the terminations and joints of solid dielectric extruded power distribution cables, is described in Rizzi, Mascioli, Baldini, Mazzetti, and Bartnikas (2009). The core of the diagnostic system is a fuzzy neural network, namely a Min–Max classifier in which the structural complexity is optimized through a Genetic Algorithm. An intelligent fault detection scheme for microgrids based on wavelet transform and deep neural networks is investigated in James, Hou, Lam, and Li (2017). The authors aim to design a system capable to provide fast fault type, phase, and location information for microgrid protection and service recovery. The study tries to demonstrate the efficacy of the proposed scheme in terms of detection accuracy, computation time, and robustness against measurement uncertainty. Two techniques for fault detection and classification in power transmission lines have been investigated in Shahid, Aleem, Naqvi, and Zaffar (2012). The approaches are based on One-Class Quarter-Sphere Support Vector Machine (QSSVM). The first technique exploits the temporal and attribute correlations of the data measured in a transmission line for fault detection during the transient stage. The second technique is, instead, grounded on a One-Class Support Vector Machine formulation, which exploits attribute correlations only for automatic fault classification. In Guikema et al. (2006) authors have established a relationship between environmental features and fault causes while a fault cause classifier based on the linear discriminant analysis (LDA) is proposed in Cai and Chow (2009). Specifically, in the last investigation, information regarding weather conditions, longitude–latitude information, and measurements of physical quantities (e.g., currents and voltages) related to the power grid have been taken into account. In any case, the literature on the subject under examination is vast and heterogeneous. This is because Smart Grids are complex technological systems and the types and causes of failure are multiple and definable at various semantic levels of the underlying layered architecture.

3. Background

In general, in the modeling of complex systems, it may be necessary to estimate the probability of a state starting from a series of measures, which can be grouped according to a certain semantic content. Often in data-driven modeling performed through machine learning techniques, there is a feature vector, whose components are heterogeneous measures and the output is a class label. The classifier then operates on the feature space and learns a model – a suitable mapping – to classify the state. In some cases, depending on the domain of the problem, it is possible to group the features into semantic units and study how these units interact in defining the state of the system. An example, which will form the ground of the case study presented in this investigation, is the synthesis of a predictive maintenance model, specifically for fault detection in a Smart Grid equipped with suitable measurement and processing systems. In fact, it is known that the electric network, whether it is for transport or distribution, is subjected to *exogenous* stresses which act as actual forces on the state of the system itself — see Fig. 1. Furthermore, the electricity grid is not a monolithic entity but is made up of many systems and subsystems, each characterized by a series of *constitutive parameters*. Among the exogenous variables, there are the meteorological conditions and the electric load or related seasonal information such as the time of year. Among the constitutive parameters, there are, instead, data regarding the structure of the network, such as the length and diameter of the cables, their composition in terms of material, the number of joints, the position, the age, etc. The exogenous variables, in this case, can be considered as agents causing a state of failure given certain constitutive parameters. Therefore, it becomes necessary to synthesize a model that not only relates exogenous causes and constitutive parameters but

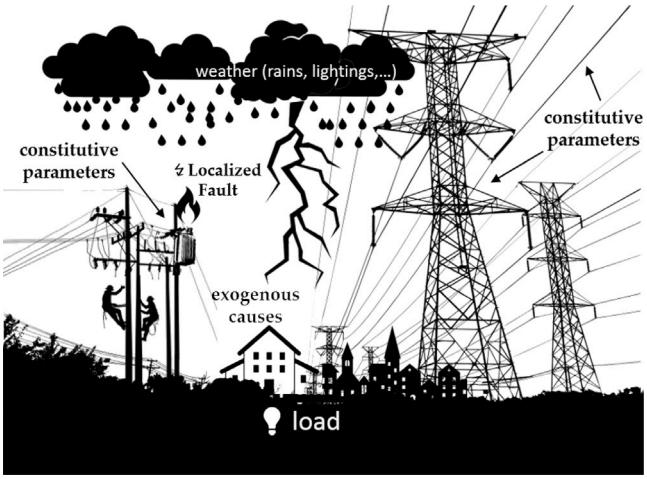


Fig. 1. Scenario relating to the impact of exogenous causes on the Smart Grid system made up of subsystems defined by suitable constitutive parameters.

which can allow estimating the vulnerability of the components from a risk assessment and network planning perspective. If the estimated model has minimal structural complexity and also allows a certain degree of interpretability from an ExplainableAI perspective, it can certainly be used for the intended purpose.

The following methodological study can be considered a spin-off of a huge project, known as "ACEA Smart Grid Pilot Project" (Vitiello et al., 2015), aiming to develop an automated recognition tool for fault states in the power grid managed by Azienda Comunale Energia e Ambiente (ACEA), the distribution company that feeds the city of Rome (Italy). In addition, the tool is designed to offer also diagnostic features, allowing the characterization of the power grid status during fault events. The ACEA Smart Grid under consideration consists of uniform section backbones that radiate outwards and can counter-supply in case of branch failure. Each backbone has two distinct Primary Stations (PS) supplying it, with each half-line protected against faults through breakers. The MV power grid comprises lines (feeders) with a nominal voltage of 20 kV, with a few legacy lines operating at 8.4 kV. The MV section covers 10,490 km, while the LV section covers 11,120 km, with cables that can be on air or underground and vary in section and material (copper, aluminum, etc.) along the backbone. The MV section has 1565 lines in service and 76 PSs, while the LV section supplies 13,292 Secondary Stations (SS). Each MV line feeds a specific number of secondary stations, each equipped with two breakers to ensure that the substation is fed by only one PS in case of a fault, thereby maintaining the radiality condition of the network (De Santis et al., 2015b). The data collected to build a dataset capable of operating a snapshot of the state of the electricity grid – see Table 1 – comes from various company information systems, such as the SCADA system, the Petersen Alarm system, the Geographic Information System (GIS) and the Territorial Information System (TIS), the Remote Control System. The data was validated together with the ACEA field experts and the choice of features is the result of a deep and long analysis of the causes of failure and malfunctioning of the network. The faults analyzed are defined as "Localized Faults" (LF) as they last longer than three minutes. Most of the features come from smart sensors disseminated in the network through the aforementioned information systems, some features are calculated taking into account the operating diagrams and the electrical model of the network available to ACEA for standard planning operations. In Table 1 are reported the dataset features, the information regarding if a given feature is semantically considered a constitutive parameter or an exogenous cause and a brief description.

The dataset, after a careful data transformation procedure corroborated by field experts and data cleaning (De Santis et al., 2015b), consists of 5342 patterns of which 4591 were of normal operations while 751 were of LF. Therefore the dataset, as in most predictive maintenance tasks, is unbalanced and this makes the problem of classification and estimation of the probability of failure more challenging.

In Fig. 2 it is possible to appreciate the monthly distribution of faults during the year 2018 (panel (a)), the maximum value of the electric current measured in Ampere on a daily basis (panel (b)) and the distribution of rain during the year (panel (c)).

Finally, it is worth specifying that the subdivision into exogenous causes and constitutive parameters belongs to the inherent semantics of the human field expert who frames the problem with a certain logic thanks to his knowledge of the problem itself. Therefore, the relationships that the B-LR model learns are correlative in nature and there is no real mechanism that learns real causation instances.

4. Problem formulation

Given a vector $\mu \in \mathbb{R}^m$ of constitutive parameters and a vector $x \in \mathbb{R}^n$ of exogenous causes, the objective is estimating the probability of fault as a particular function of the form:

$$p(\omega = 1|\mu, x, \Theta) = \sigma(\mu, x; \Theta) \tag{1}$$

where $\omega \in \{1, 0\}$ is the class variable (1, 0 are the fault state and standard functioning state, respectively) and Θ is a set of free parameters. Moreover, we define the vulnerability vector $v \in \mathbb{R}^n$ as a linear function of the constitutive parameters μ summed to a bias vector:

$$v(\mu) = A\mu + v_0, \tag{2}$$

where $A \in \mathbb{R}^{n\times m}$ is a matrix that we call "correlation matrix" and $v_0 \in \mathbb{R}^n$. Finally, we define the probability of fault as a function of the scalar product between the vulnerability vector $v(\mu)$ and the vector of exogenous causes, that is:

$$p(\omega = 1|\mu, x, \Theta) = \sigma(v(\mu)^T x) = \sigma(\mu^T A^T x + v_0^T x). \tag{3}$$

In order to obtain a probability the function $\sigma(\cdot)$ is defined as a non-linearity given by a sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-\beta z + b_s}}, \tag{4}$$

where β is the slope parameter and b_s a bias scalar, both possibly trainable.

Hence, given the Eq. (1), for the probability of fault p we define the following model:

$$p(\omega = 1|\mu, x, \Theta) = \frac{1}{1 + e^{-v(\mu)^T x}} = \frac{1}{1 + e^{-\mu^T A^T x - v_0^T x}}, \tag{5}$$

where $\Theta = (A, v_0)$.

Given a dataset $D = \{(\mu_1, x_1, \omega_1), (\mu_2, x_2, \omega_2), \ldots, (\mu_N, x_N, \omega_N)\}$ of N data points as triples (μ_j, x_j, ω_j), the objective is to training a classifier defined by the hard classification rule:

$$\hat{\omega} = \underset{\omega}{\operatorname{argmax}}\, p(\omega|\mu, x; \Theta). \tag{6}$$

A general classification problem instance is defined as a couple of disjoint sets, namely training set S_{tr}, and test set S_{ts}. The model parameters are estimated on S_{tr} while the generalization capabilities are evaluated on S_{ts}.

It is worth noting that, considering Eq. (3) and in particular the set of free parameters Θ, estimating p means estimating a vulnerability vector $v(\mu)$ that aligns with the vector of exogenous causes x in case p is equal to 1.

Taking the negative log-likelihood of the dataset under a Bernoulli distribution with parameter p_i we can define the loss as binary cross-entropy. Specifically, if we assume that the probability of the positive

Table 1
List of the considered features in power grid dataset records. "CP" is for "Constitutive Parameter" while "EC" is for "Exogenous Cause".

	Feature	Type	Description
1	Day	–	Day in which the LF was detected
2	Minute (Minute-Fault)	CP	minute of the day in which the LF was detected
3	Cable section (Cable-section)	CP	section of the cable
4	Cable length (Cable-length)	CP	length of the cable
5	Feeder voltage (MV-voltage)	CP	nominal voltage of the feeder
6	Copper percentage ('Copper-percentage')	CP	weighted percentage of copper based on the length of a copper trunk on the total length of the branch
7	Air percentage (Air-percentage)	CP	percentage of the feeder in the air
8	Level P ratio (Level-P-Ratio)	CP	ordinal number of the secondary substation upstream of the faulty section expressed as a real number in [0,1]
9	Semi-backbone-MV-line current (semi-b-MV-line-current)	EC	last current sample measured on the line riser before the fault event
10	Current (Current)	EC	estimate of the portion of line current about the faulty branch calculated based on "normal scheme" topological data
11	Minimum temperature (Min.-Temp.)	EC	Minimum registered temperature
12	Maximum temperature (Max.-Temp.)	EC	Maximum registered temperature
13	Rain (Mm-rain)	EC	millimeters of rain measured by the weather station geographically closest to the cabin downstream of the fault on the date the fault occurred
14	Max. current (Max.-Current)	EC	maximum current sample recorded on the riser daily
15	Min. current (Min.-Current)	EC	minimum current sample recorded on the riser daily
16	Mean Current (Mean.-Current)	EC	average current sample recorded on the averaged amount daily
17	Mean Current ratio (Mean-Current-ratio)	EC	percentage of fault current calculated concerning the maximum current level computed as an annual average in the days of fault

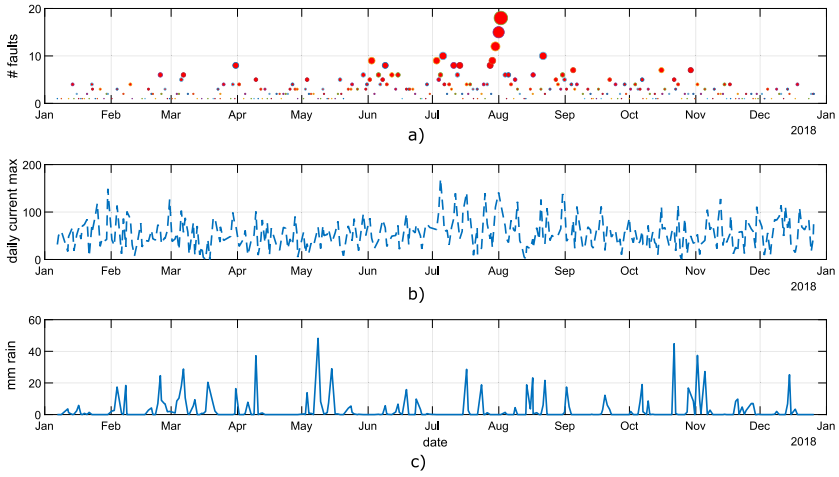


Fig. 2. The ACEA dataset. (a) Monthly distribution of faults during the year 2018. (b) Maximum value o the current [A] on daily bases. (b) distribution of rain during the year.

class ω_i is given by a Bernoulli distribution with parameter p_i, then the likelihood of the dataset can be written as:

$$\mathcal{L} = \prod_{i=1}^{N} p_i^{\omega_i} (1 - p_i)^{1-\omega_i} \tag{7}$$

Taking the logarithm of this likelihood gives us the log-likelihood:

$$\log \mathcal{L} = \sum_{i=1}^{N} \omega_i \log(p_i) + (1 - \omega_i) \log(1 - p_i), \tag{8}$$

which is equivalent to the binary cross-entropy loss up to a sign change and the normalization factor $1/N$:

$$L = -\frac{1}{N} \sum_{i=1}^{N} [\omega_i \log(p_i) + (1 - \omega_i) \log(1 - p_i)]. \tag{9}$$

The regularized form of the cross-entropy given in Eq. (9) can be written as:

$$L_{\text{reg}} = -\frac{1}{N} \sum_{i=1}^{N} [\omega_i \log(p_i) + (1 - \omega_i) \log(1 - p_i)] + \\ + \frac{\lambda}{2} |A|_F^2 + \frac{\lambda_b}{2} |b|^2. \tag{10}$$

where the term $\frac{\lambda}{2}|A|_F^2 = \frac{\lambda}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2$ is the penalty term that depends on a regularization parameter λ and the Frobenius norm of matrix A, and $\frac{\lambda_b}{2}|b|^2$ is the penalty term related to the bias $b = v_0^T x$.

4.1. Boosting the bilinear logistic model

In the previous section, we model the probability of the relation between a vector μ and a vector x with a simple bilinear layer such as $p(\omega = 1|\mu, x, \Theta) = \sigma(v(\mu)^T x) = \sigma(\mu^T A^T x + v_0^T x)$. Among the various network architectures that can be designed, we propose here a "boosted" version, where are defined k correlation matrices A_i and k bias vectors v_{0_i} concatenated in k "heads". Indicating with h_i the output of the ith bilinear layer, the boosted bilinear logistic regression (boosted B-LR) has the following form:

$$p(\omega = 1|\mu, x, \Theta) = \sigma(W \cdot \text{concat}(h_i) + b), \tag{11}$$

where $h_i = \sigma(\mu^T A_i^T x + v_{0_i}^T x)$ and W and b are the weights of the last dense layer. This architecture allows learning different representations in terms of matrices A_i and biases v_{0_i},

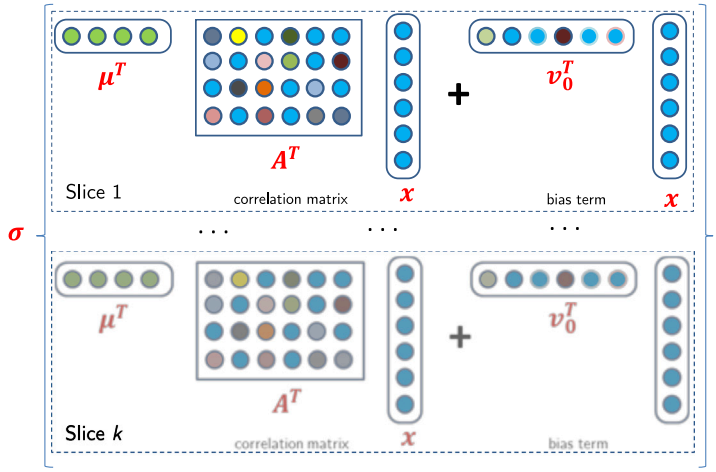


Fig. 3. Visualization of the Bilinear Neuron. Each dashed box represents one slice of the tensor; there are k slices where k, in this specific case, is the batch size.

4.2. The bilinear neuron

Rewriting the Eq. (3) in the form:

$$\begin{aligned}&\sigma\left[\mu^T A^T x + b\right] = \\ &= \sigma\left[[\mu_1 \mu_2 \dots \mu_m]\begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b\right] = \\ &= \sigma\left[b + \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\mu_i x_j\right],\end{aligned} \tag{12}$$

where $b = v_0^T x$ is a bias term, the elements a_{ij} of the matrix A can be interpreted as the weights associated with the pairs of variables μ_i and μ_j in the scalar product $\langle A\mu, x\rangle = (A\mu)^T x = \mu^T A^T x$. In other words, the elements a_{ij} indicate the relative importance of each pair of variables for the prediction of the dependent class variable (see Fig. 3).

4.3. On bilinearity

At this point, it may be interesting to sketch briefly some properties of the involved bilinear form, useful for gaining insight on the general interpretation of the vulnerability vector $v(\mu)$, the parameter matrix A and the bias vector v_0.

Let V and W be two vector spaces over the field $\mathbb{R}$ of scalars, with dimensions m and n, respectively. Then we can define a bilinear form $B : V \times W \to \mathbb{R}$ as follows. Let A be an $m \times n$ matrix. For any $x \in V$ and $y \in W$, we define the bilinear form $B(x, y)$ as:

$$B(x, y) = x^T A^T y \tag{13}$$

where a_{ij} is an element of A. Note that $x^T A^T y$ is a scalar, so $B(x, y)$ is a bilinear form, i.e., it is linear in both x and y. In other words, for any vectors $x_1, x_2 \in V$, $y_1, y_2 \in W$, and scalars $c_1, c_2 \in \mathbb{R}$, we have:

$$\begin{aligned}B(c_1x_1 + c_2x_2, y_1) &= (c_1x_1 + c_2x_2)^T A^T y_1 \\ &= c_1 x_1^T A^T y_1 + c_2 x_2^T A^T y_1 \\ &= c_1 B(x_1, y_1) + c_2 B(x_2, y_1)\end{aligned} \tag{14}$$

and

$$\begin{aligned}B(x_1, c_1y_1 + c_2y_2) &= x_1^T A^T (c_1y_1 + c_2y_2) \\ &= c_1 x_1^T A^T y_1 + c_2 x_1^T A^T y_2 \\ &= c_1 B(x_1, y_1) + c_2 B(x_1, y_2).\end{aligned} \tag{15}$$

So B is a bilinear form between V and W.

If V and W have the same dimension, say $m = n$, A is a square matrix.

Moreover, we can define a linear transformation $T_A : \mathbb{R}^m \to \mathbb{R}^n$ by $T_A(x) = Ax$. Then the dot product $x^T A^T y$ can be expressed as:

$$x^T A^T y = (Ax)^T y = (T_A(x))^T y \tag{16}$$

We can interpret this dot product as a scalar product between a linear transformation T_A and the vector y. In other words, we can define a bilinear form $B : \mathbb{R}^n \times \mathcal{M}(\mathbb{R}^m, \mathbb{R}^n) \to \mathbb{R}$, where $\mathcal{M}(\mathbb{R}^m, \mathbb{R}^n)$ is the set of all linear transformations (a vector space) from $\mathbb{R}^m$ to $\mathbb{R}^n$, by:

$$B(y, T_A) = (T_A(x))^T y = (Ax)^T y = x^T A^T y \tag{17}$$

So we can say that $(Ax)^T y$ is the scalar product between the linear transformation T_A and the vector y, expressed using the bilinear form $B(y, T_A)$.

4.4. Deeping on the correlation matrix trough SVD

The SVD decomposition (Golub & Van Loan, 2013; Strang, Strang, Strang, & Strang, 1993) is a very powerful tool for gaining insight into the structure of a system and its behavior described in terms of principal modes. In the case under examination, the interaction between the system's constituent parameters μ and exogenous causes x is described by the matrix A (and by the bias vector v_0) which captures, in the first instance, the reciprocal interrelationships in generating the probability p of an event (e.g., a failure), net of the non-linearity of the output sigmoidal function $\sigma(\cdot)$ – see Eq. (3).

Let apply the SVD to A^T expressing it as a product of three matrices:

$$A^T = U\Sigma V^T, \tag{18}$$

where U is an $m \times m$ orthogonal matrix, Σ is an $m \times n$ diagonal matrix with non-negative entries arranged in decreasing order along the diagonal, and V is an $n \times n$ orthogonal matrix ($U^T U = I$ and $V^T V = I$). The value $r = \min(m, n)$ is the rank of A^T (Kerschen & Golinval, 2002).

Substituting Eq. (18) into the previous expression we have:

$$\mu^T U \Sigma V^T x = \tag{19}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} \mu_i u_{ik} \sigma_k v_{jk}^T x_j = \tag{20}$$

$$= \sum_{k=1}^{r} \sigma_k \sum_{i=1}^{m}\sum_{j=1}^{n} \mu_i u_{ik} v_{jk}^T x_j = \tag{21}$$

$$= \sum_{k=1}^{r} \sigma_k (u_k^T x)(v_k^T y) \tag{22}$$

where the left singular vector u_i and right singular vector v_i are the ith columns of U and V, respectively, and the singular value σ_i is the ith diagonal entry of Σ.

In the context of Eq. (22), the singular values indicate how much each component of the constitutive parameters μ contributes to the output probability p, given a specific set of external causes x. Higher singular values correspond to greater contributions, and lower singular values correspond to smaller contributions. Moreover, the singular values can be used to measure the sensitivity of the output probability to changes in the input parameters. Specifically, the sensitivity of p to changes in μ is proportional to the ratio of the singular values. Thus, a larger ratio of singular values indicates greater sensitivity to changes in the input parameters.

Being the singular values σ_i the magnitudes of the projections of the matrix A onto the orthogonal basis vectors defined by the singular vectors u_i and v_i, they represent the amount of variation in A that is explained by each singular vector. Larger singular values correspond to directions in A that contain more information or are more important for explaining the output.

Now, we note that the left singular vectors $u_1, u_2, \dots, u_r$ form an orthonormal basis for the column space of A, while the right singular vectors $v_1, v_2, \dots, v_r$ form an orthonormal basis for the row space of A. Hence, in the context of the Eq. (22), the left singular vectors u_i represent the directions in the space of constitutive parameters that are most strongly correlated with the output probability p. Specifically, the ith left singular vector represents the direction in which a unit change in the constitutive parameter results in the largest change in the output probability p, given a specific set of external causes x. In other words, the left singular vectors u_i describe the variation of the constitutive parameters in terms of modes, and the corresponding singular values σ_i represent the importance of each mode in explaining the variation. The right singular vectors v_i describe the variation of the external causes in terms of modes, and the corresponding singular values σ_i represent the importance of each mode in explaining the variation. Finally, by evaluating the spread of the singular values σ_i we can truncate the SVD to the first k modes, capturing the most important variation in both μ and x, while the remaining modes capture the less important variation. Considering Eq. (22), the SVD can be used to decompose the original matrix A^T in a weighted sum of singular values σ_i of i rank-1 matrix generated by the outer product of left and right singular vectors u_i and v_i^T, respectively. Truncating the weighted sum it is possible to have a filtered version of the original A^T matrix.

We can also express the vulnerability vector $v(x)$ in term of the SVD of A as:

$$v(x) = Ax + v_0 = U\Sigma V^T x + v_0 = \sum_{i=1}^{r} \sigma_i (u_i)(v_i^T x) + v_0 \tag{23}$$

In this case, larger singular values correspond to the constitutive parameters that are more sensitive to changes in external causes, or in other words, the parameters that have a greater impact on the vulnerability of the system. The singular vectors of A^T give us information about the directions in the input space that are most important for explaining the vulnerability of the system, while the singular vectors of A give us information about the directions in the output space that are most important for explaining the effect of external causes on the system.

4.5. Data denormalization

In a real problem, the measurements related to a pattern that identifies a state of a system before being input to a learning system can be normalized. Therefore it may be necessary to calculate the learned quantities, specifically the A matrix and the v_0 vector in denormalized form for further analysis. One of the best-known normalization techniques is the "affine normalization" which is often performed on a feature basis.

In general, given the data matrices $U = [\mu_1, \mu_2, \dots, \mu_n] \in \mathbb{R}^{(N\times n)}$ and $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{(N\times m)}$, with n, m column vectors, respectively, the affine normalization formulation can be written as:

$$\begin{aligned} \hat{\mu}_i &= \frac{\mu_i - m_i^\mu}{M_i^\mu - m_i^\mu} = \frac{1}{M_i^\mu - m_i^\mu}\mu_i - \frac{m_i^\mu}{M_i^\mu - m_i^\mu} \\ \hat{x}_i &= \frac{x_i - m_i^x}{M_i^x - m_i} = \frac{1}{M_i^x - m_i^x}x_i - \frac{m_i^x}{M_i^x - m_i^x} \end{aligned} \tag{24}$$

where $M_i^{\mu,x}$ and $m_i^{\mu,x}$ are the maximum and minimum value for the ith feature of the vectors μ and x, respectively. Considering the vector of constituent parameters μ and the vector of exogenous causes x we can write the normalization relations in compact form as:

$$\begin{aligned} \hat{\mu} &= C_\mu \mu + b_\mu, \\ \hat{x} &= C_x \mu + b_x. \end{aligned} \tag{25}$$

The matrices $C_\mu \in \mathbb{R}^{(n\times n)}$ and $C_x \in \mathbb{R}^{(m\times m)}$ are diagonal and where the ith elements are the slopes $\frac{1}{M_i^\mu - m_i^\mu}$ and $\frac{1}{M_i^x - m_i^x}$ in Eq. (24), respectively. Furthermore, the ith elements of the bias vectors b_μ and b_x are $-\frac{m_i^\mu}{M_i^\mu - m_i^\mu}$ and $-\frac{m_i^x}{M_i^x - m_i^x}$, respectively.

Considering two normalized vectors $\hat{\mu}$ and $\hat{x}$, starting from the last term of Eq. (3) and using the normalizing relations in Eq. (25) we can obtain the following quantities:

$$\begin{aligned} &\hat{v}(\hat{\mu}) = \hat{A}\hat{\mu} + \hat{v}_0 \\ &p = \sigma(\hat{v}(\hat{\mu})^T \hat{x}) = \sigma(\hat{\mu}^T \hat{A}^T \hat{x} + \hat{v}_0^T \hat{x}) = \\ &= \sigma[(\mu^T C_\mu^T + b_\mu^T)\hat{A}^T (C_x x + b_x) + \hat{v}_0 (C_x x + b_x)] = \\ &= \sigma[\mu^T C_\mu^T \hat{A}^T C_x x + \mu^T C_\mu^T \hat{A}^T b_x + b_\mu^T \hat{A}^T C_x x + \\ &+ b_\mu^T \hat{A}^T b_x + \hat{v}_0^T C_x x + \hat{v}_0^T b_x]. \end{aligned} \tag{26}$$

In other words, starting from the normalized quantities we can recall the following relations for the non normalized one:

$$\begin{aligned} A &= C_\mu^T \hat{A}^T C_x \\ v_0 &= (b_\mu^T \hat{A}^T + \hat{v}_0^T) C_x \\ x_0 &= C_\mu^T \hat{A}^T b_x \\ p_0 &= (b_\mu^T \hat{A}^T + \hat{v}_0^T) b_x. \end{aligned} \tag{27}$$

with which, finally, we can express the probability of a state as:

$$p = \sigma(\mu^T A^T x + v_0^T x + \mu^T x_0 + p_0). \tag{28}$$

5. Simulation settings and results

The simulations were conducted using the TensorFlow framework (ver. 2.11) by designing a custom Keras layer that implements the bilinear form under study and the L_1 and L_2 regularization procedure on the trainable parameters (specifically the correlation matrix A and the bias vector v_0). To conduct a deep analysis of the behavior of the training system in estimating the probability of failure in case of unbalanced classes, it is possible to choose whether to make the sigmoid parameters β and b_s – inherent to the output layer of the network – trainable as well. The vector of the exogenous causes x has a dimension equal to 9 while that of the constituent parameters vector μ has 7 components, therefore the matrix A has 63 parameters while the bias vector has 9 parameters, for a total of 72 + 2 trainable parameters (considering also β and b_s – see Eq. (4)). The training was carried out in all the experiments by adopting a stochastic gradient descent using the Adam optimizer with a learning rate equal to 0.005 and a batch size equal to 32. During the experiments good stability was found in the classification results, therefore it was decided to fix the learning rate at the aforementioned value.

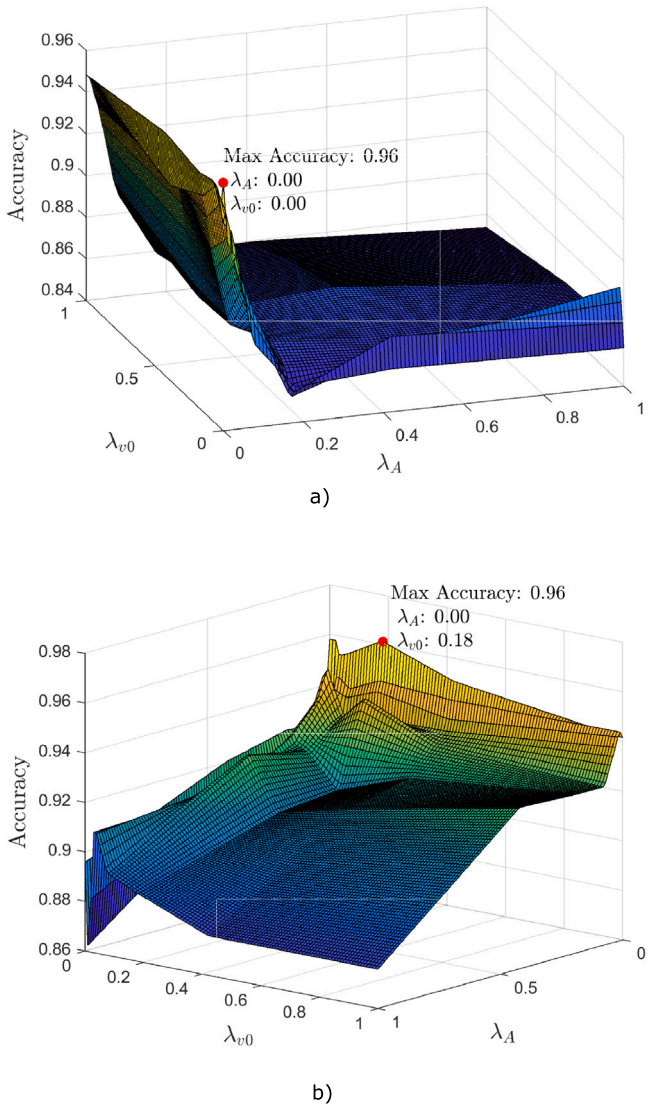


Fig. 4. Regularization surfaces obtained through a grid search with parameters λ_A and λ_{v0}. (a) L_1 regularization, (b) L_2 regularization.

5.1. Sensitivity analysis of the regularization parameters

In order to evaluate the effect of the regularization on the classification performance, we perform a grid search for the parameter λ_A and for the parameter λ_b – see Section 4. Experiments are conducted both in the case of L_1 regularization, for eliciting sparsity, and L_2. The parameters were evaluated for 20 logarithmically spaced values in the range $[0, 1]$. Fig. 4 shows the two surfaces measured in terms of Accuracy.

The maximum value has also been indicated in the figure. The experimental results show that at least for this dataset high values for λ_A and λ_b are not necessary. This indicates that the dataset under examination has a low complexity and the model does not need regularization.

5.2. Score calibration

Even if in Fig. 4 we referred to the estimate of the probability of an event (e.g., a failure), it is known that in most of the classifiers the probabilities can be uncalibrated. Therefore, the classifier's output can be defined more properly with the term "score". Formally speaking, the classifier under analysis is said to be well-calibrated if $P(\omega|s(\mu, x) = s)$ – where with "s" we indicated the uncalibrated probabilities, i.e., the output score. In other terms, the probability for a pattern x to belong to a label ω converges to the score $s(x) = s$ as the number of samples tends to infinity (Murphy & Winkler, 1977; Zadrozny & Elkan, 2001). In plain terms, the calibration of a classification system consists in mapping the scores (or not calibrated probability estimates) into proper probability estimates bounded in the range $[0, 1]$ by definition. Calibrating a classifier means learning suitable mappings between scores and calibrated probabilities (Martino, De Santis, Baldini, Rizzi, et al., 2019). There are many techniques for learning this mapping, such as the Platt Scaling (Smola, Bartlett, Schölkopf, & Schuurmans, 2000) and Isotonic Regression (Zadrozny & Elkan, 2001) techniques. The calibration performance can be assessed through the so-called reliability diagram, which is a graphical tool used to assess the calibration of a predictive model by comparing the predicted probabilities (i.e., scores) with the observed frequencies. On the x-axis we have the scores/probabilities while on the y-axis we have the empirical probabilities (observed frequencies), that are, namely, the ratio between the number of patterns in class ω with score s and the total number of patterns with score s. If the classifier is well-calibrated, then all points lie on the $y = x$ line (i.e., the scores are equal to the empirical probabilities). In our experiments, we used the Isotonic Regression calibration method but we also obtained good results with Plat scaling. Fig. 5 shows the reliability diagrams before (panel (a)) and after calibration (panel (b)). The learned model was used in subsequent experiments regarding failure analysis and modeling.

5.3. Classification performance comparison

In order to compare the classification performance of the bilinear logistic form on the ACEA dataset, 7 classifiers known in the literature were trained using the Matlab™Statistics and Machine Learning Toolbox™, which differ on the basis of the model structure and training methods. Specifically, the following classifiers were tested: Logistic Regression (LR), Linear Discriminant (LDA), Naive Bayes (NB), Gaussian Kernel SVM (G-SVM), weighted KNN, a Feedforward Neural Network (FFNN) and Random Forest (RF). For the hyperparameters tuning the Bayesian optimization scheme is adopted, which is a probabilistic model-based method that builds a probability model of the objective function and uses it to select the most promising hyperparameters to evaluate in the true objective function. As concerns the LR algorithm, the optimization regards the regularization strength (Lambda) and the regularization type (L1 or L2). For LDA the discriminant type (linear, quadratic and the respective diagonal forms) and the two regularizer parameters are optimized. Concerning the NB the distribution parameters (normal, kernel) are optimized. For the G-SVM the BoxConstraint, the KernelScale are optimized. For RF the number of trees, the maximum number of splits and the minimum leaf size are optimized, while for the FFNN the number of hidden layers and the number of neurons per layer are tuned.

Table 2 shows the performance in terms of Accuracy on the validation set, Accuracy on the test set, Area Under Curve (AUC) for the class $\omega = 1$ (target) and $\omega = 0$ (non-target).

Considering the results in general, the plain B-LR classifier achieves results comparable with the other classifiers analyzed, both in terms of Accuracy and AUC. The details of the results can also be appreciated in Fig. 6 where the Receiving Operating Characteristic curve (ROC) (Fawcett, 2006) for each class is reported for each investigated classifier. The work points on the ROC are also reported. The inserts instead show the normalized confusion matrices together with the Positive Predictive Value (PPV) and the False Discovery Rate (FDR). The classifiers proposed in this study achieve very low results in terms of false alarm probability making it useful in the context in which it is being investigated.

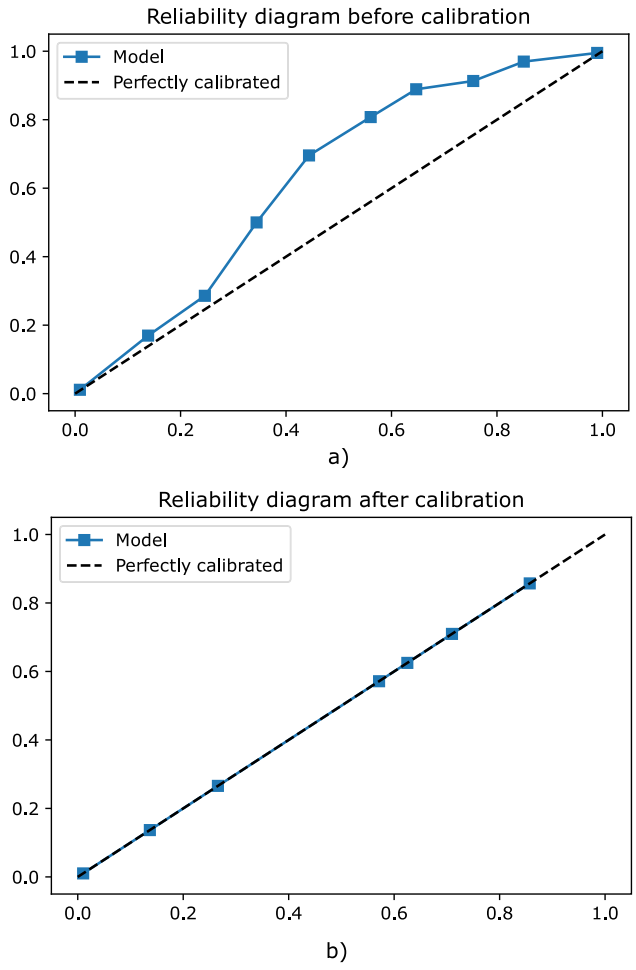


Fig. 5. (a) Reliability diagram before calibration. (b) Reliability diagram after calibration. Results from the ACEA dataset.

Table 2
Comparison of classification results with several classifiers known in technical literature.

Classifier	Val. accuracy	Test accuracy	AUC 1	AUC 0
Boosted B-LR	95.79%	95.88%	95.65%	95.65%
Plain B-LR	95.11%	96.18%	95.65%	95.65%
Logistic Regression	95.74%	95.42%	95.97%	95.97%
Lin. Discriminant	93.12%	93.26%	96.99%	96.99%
Naive Bayes	88.18%	89.5%	89.37%	89.37%
SVM	95.81%	95.51%	97.45%	97.45%
weighted KNN	93.99%	94.01%	94.00%	94.00%
FF Neural Network	95.67%	96.45%	94.28%	94.23%
Random Forest	94.62%	94.57%	91.60%	91.60%

This is particularly true for plain B-LR while the boosted version has comparable results in terms of accuracy. In any case, as expected, increasing the number of bilinear layers improves the convergence speed of the training phase. However, we have reported this variant here for merely experimental purposes, but we assume that the increase in the number of bilinear layers can be justified (and useful) by the presence of a dataset with a much higher sample size than the investigated ACEA dataset and greater underlying complexity.

5.4. Failure analysis and modeling

The main aim of the present section is to demonstrate the potential both of the bilinear model in scenario analyses and therefore of the representation learned through the estimation of the correlation matrix. First of all, in Fig. 7 (panel (a)) it is possible to appreciate the scatter plot — obtained through the t-distributed stochastic neighbor embedding (t-DSNE) dimensionality reduction algorithm (Van der Maaten & Hinton, 2008) – of the ACEA dataset considering the vectors at the same semantic level, i.e., by considering the composite vector $c = \text{concat}(\mu, x)$. Panel (b) within the same figure shows the projections of the vector of the constitutive parameters μ and of the exogenous causes x in the space of left and right singular values computed through the SVD. In particular, the patterns relating to the constituent parameters are the rows of the matrix obtained from the transformation $\mu^T U_{(1:k,:)} \sqrt{\Sigma_{(1:k,1:k)}}$ while the patterns relating to the exogenous causes are the rows of the matrix $\sqrt{\Sigma_{(1:k,1:k)}} V^T_{(1:k,:)} x$, for all the vectors μ and x belonging to the dataset and for the truncation parameter set to $k = 2$. The reason behind the choice of the value $k = 2$ can be found in Fig. 8 (panel b) where are reported the singular values of the matrix A. It is clear that the first two singular values have a high spread concerning the next ones. Panel (a), on the other hand, shows the heatmap of the A matrix estimated with the base B-LR model, where the exogenous causes are shown in rows and the constituent parameters in columns. Examining the correlation matrix it is noted that some pairs of exogenous cause-constitutive parameter features have high values indicating a strong correlation. This is the case of the percentage of the cable in the air (Air-percentage) and the electric current at the beginning of the backbone (semi-b-MVline-current) or the value of the Current feature before the failure and its position in terms of numbers of SS along the backbone (Level-P-Ratio). The sign of the entries in A tells us if the features are negatively or positively correlated. We can see for example that at least in this investigated dataset the rain is weakly correlated with the constitutive parameters except for Air-percentage. This feature is also strongly correlated with the features obtained from the electric current measured before the faults and the temperature measured in the fault location. Interestingly, but nevertheless an expected result, the cable length is correlated with the electric current features (specifically with Current). This is true also for the cable section which is strongly negatively correlated with the Current feature measured before the failure.

In panel (c) it is possible to retrieve the same information but in a multifaceted way. In fact, we report the values of the first, second and third left and right singular vectors related to the first three singular values. The bar plot expresses pictorially the directions of various "natural modes" – with power measured by singular values – determining a state of the system, measured by its probability and considering the interrelation of constitutive parameters and external causes. Similar considerations can be provided for the vulnerability vector $v(\mu)$ both exploiting the SVD decomposition or analyzing its entities directly, in the same way we did for the correlation matrix A. Fig. 9, instead, reports two specific types of representation that provide the user with a comprehensive pictorial view of the failure scenario. From panels (a) to (d) and (i) to (n) we report the average level lines of the probability surface related to failures computed parametrizing a pair of features (chosen among exogenous causes and constitutive parameters) within the range of variation calculated with the data normalization procedure. The average is computed by grouping the pattern pertaining to the ACEA dataset according to the season, meaning that we are learning a specific model (A and v_0) for each one of the four seasons. In panels (e) to (h) are reported the seasonal average of the vulnerability vectors $v(\mu)$ computed through the constitutive parameters. The correlation matrix A and the vulnerability vector v_0, after the training procedure, were denormalized by exploiting the relations obtained in Section 4.5.

Regarding the average probability diagrams, we let variate – within the precomputed normalization range – the (averaged) Min-Current and the Max-Current features measured before the failure. We can see a different behavior depending on the season. In all cases for high values of the Max-Current the probability of fault approaches 1. An interesting behavior can be found also investigating the constitutive parameters' pair Air-percentage-Cable length. While in winter there is an appreciable interrelation of both features, this behavior change in

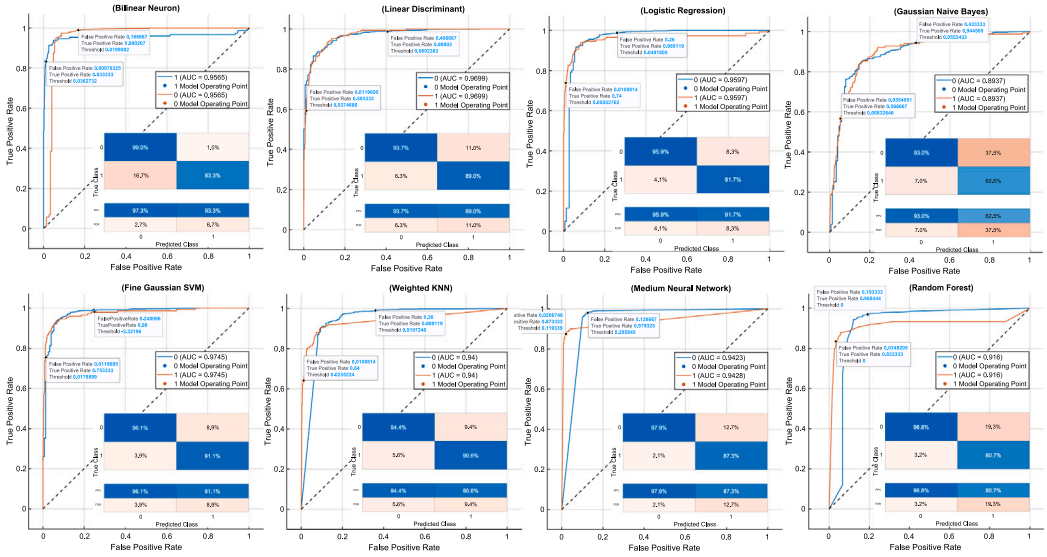


Fig. 6. Classification performance of the various tested models. It is reported for each algorithm the Receiving Operating Characteristic curve for each class and the AUC value. The inset panel reports the confusion matrix obtained from the test set with the Positive Predictive Value and the False Discovery Rate.

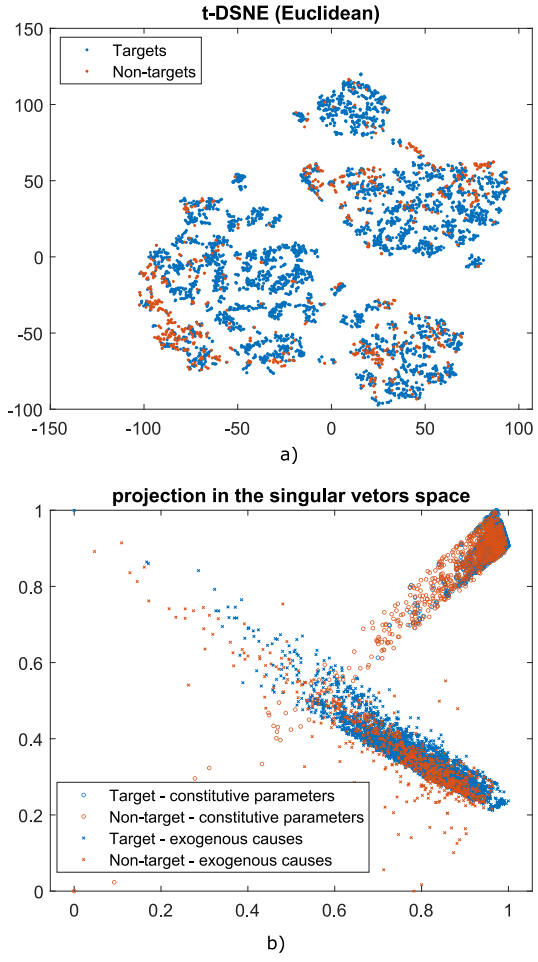


Fig. 7. Low dimensionality visualization of the ACEA datasets. (a) Scatter plot obtained with t-DNSE algorithm with Euclidean distance on the plain patterns (b) Scatter plots of the projection of constitutive parameters and external causes on the left and right space, respectively, generated by singular vectors.

spring and summer. Conversely, in the fall we found no dependence on the Air-percentage feature. The features were chosen by analyzing the picture provided by the analysis of the correlation matrix and by the inspection of the singular vectors. However, other interesting cases have not been reported for reasons of space.

6. Conclusion

In this paper it has been proven that the bilinear approach to fault modeling in a real-world power grid can be very useful. The effectiveness lies not only in having a low computational complexity algorithm with classification performance comparable to other known standard classifiers but it can also be considered as a good working point in the trade-off between generalization capacity and explainability, in line with the xAI paradigm. In fact, in terms of performance, the B-LR classifier achieves a low false alarm probability (with 96.18% of accuracy on test set) demonstrating its usability in the context in which it was designed, i.e., in predicting faults starting from heterogeneous measurements on the state of the power grid. In addition, after a suitable calibration procedure of the output scores, it can be used to estimate the empirical probability of failure given a pattern of measurements. Furthermore, the semantic disambiguation between exogenous causes and constitutive parameters of the features together with the data-driven learned correlation matrix opens the way to knowledge discovery procedures — These distinctive features tip the scale in favor of the proposed model over standard classifiers, assuming equal performance. The matrix can be used to relate the causes to the characteristics of the components and estimate their vulnerability, which, as we have demonstrated, can have a seasonal nature. This means that each component can be associated with a "label" indicating the vulnerability to certain conditions that may occur during operations, with the additional possibility of predicting even heterogeneous scenarios simply by querying the model with new data. Furthermore, the estimation of the probability of failure allows a stratified risk analysis for scenario analysis establishing priorities for intervention on the network in the short and medium term. As a future development, in addition to the application of this model to the estimation of impact metrics and risk analysis of the entire MV network feeding the city of Rome, the plan is to synthesize a model trained on a quantity of data spanning several years – when it will be available – placed in the pipeline to a Clustering system whose function is to estimate in an unsupervised way homogeneous subclasses of faults. Therefore, the aim is to

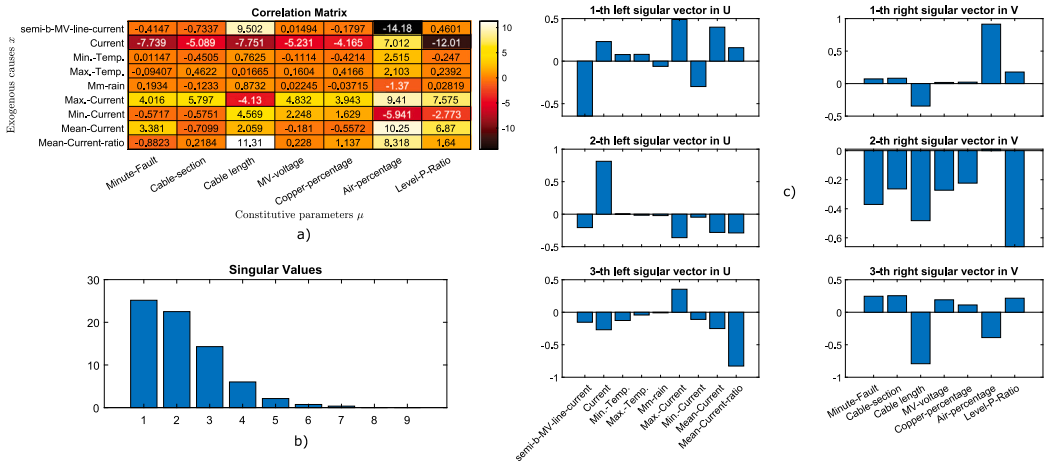


Fig. 8. Spectral analysis of the Correlation matrix computed on the entire ACEA dataset. (a) Correlation matrix. (b) Singular values related to the SVD of the matrix A. (c) First, second and third left and right singular vectors related to the SVD of the matrix A.

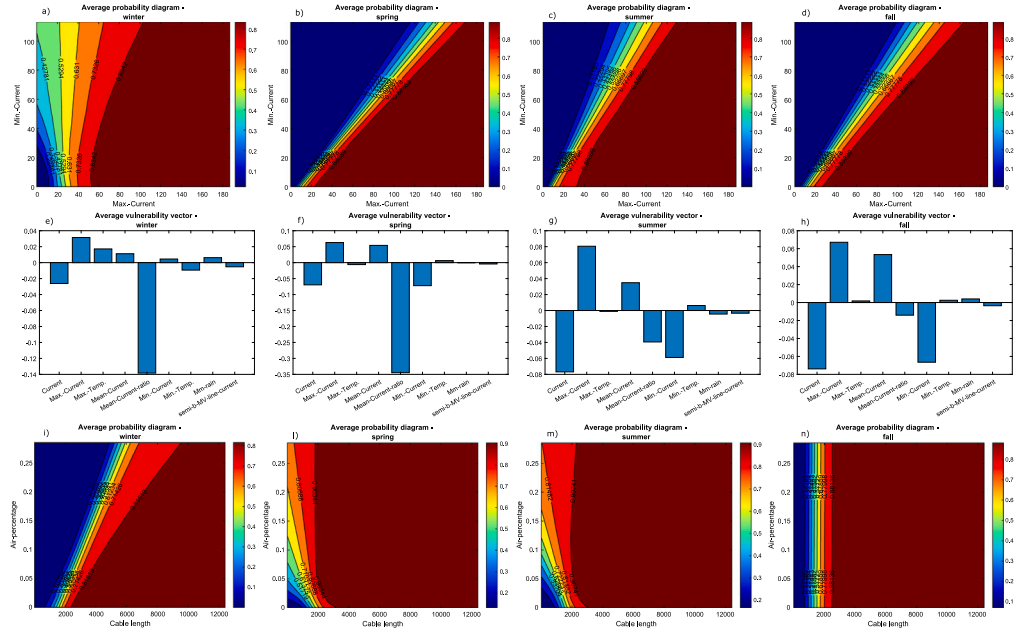


Fig. 9. Seasonal Analysis. Average probability diagrams for some exogenous causes (upper panels) and some constitutive parameters (lower panels). Average vulnerability diagrams (middle panels).

provide the user with a bilinear model suited for the specific subclass of failure, hence allowing the extraction of useful information from each discovered cluster. Finally, it is important to stress that the proposed approach can be generalized to other predictive maintenance problems, as well as to other diagnostic problems, not necessarily constrained to pertain to technological networks and devices (e.g., clinical diagnostic problems). As a general criterion, this approach can be applied every time a set of features can be conveniently partitioned on a semantic basis, between external causes and constituent parts characteristics, opening the way to an easy and effective knowledge discovery tool for a wide range of possible applications.

CRediT authorship contribution statement

Enrico De Santis: Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Antonello Rizzi:** Funding acquisition, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Antonello Rizzi reports financial support was provided by MOST Sustainable Mobility Center.

Data availability

Data will be made available on request.

Acknowledgments

This study was carried out within the MOST – Sustainable Mobility Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1033 17/06/2022, CN00000023). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

Bangalore, P., & Tjernberg, L. B. (2015). An artificial neural network approach for early fault detection of gearbox bearings. *IEEE Transactions on Smart Grid*, *6*(2), 980–987.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*: *vol. 4*, Springer.

Bose, B. K. (2017). Artificial intelligence techniques in smart grid and renewable energy systems—Some example applications. *Proceedings of the IEEE*, *105*(11), 2262–2273.

Cai, Y., & Chow, M.-Y. (2009). Exploratory analysis of massive data for distribution fault diagnosis in smart grids. In *2009 IEEE power & energy society general meeting* (pp. 1–6). IEEE.

Chen, K., Hu, J., & He, J. (2016). Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder. *IEEE Transactions on Smart Grid*, *9*(3), 1748–1758.

Chen, D., Socher, R., Manning, C. D., & Ng, A. Y. (2013). Learning new facts from knowledge bases with neural tensor networks and semantic word vectors. arXiv preprint arXiv:1301.3618.

De Santis, E., Arnò, F., & Rizzi, A. (2022). Estimation of fault probability in medium voltage feeders through calibration techniques in classification models. *Soft Computing*, *26*(15), 7175–7193.

De Santis, E., Livi, L., Sadeghian, A., & Rizzi, A. (2015b). Modeling and recognition of smart grid faults by a combined approach of dissimilarity learning and one-class classification. *Neurocomputing*, *170*, 368–383. http://dx.doi.org/10.1016/j.neucom.2015.05.112.

De Santis, E., Paschero, M., Rizzi, A., & Mascioli, F. M. F. (2018). Evolutionary optimization of an affine model for vulnerability characterization in smart grids. In *2018 international joint conference on neural networks* (pp. 1–8). IEEE.

De Santis, E., Rizzi, A., & Sadeghian, A. (2018). A cluster-based dissimilarity learning approach for localized fault classification in smart grids. *Swarm and Evolutionary Computation*, *39*, 267–278.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874.

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations*. JHU Press.

Guikema, S. D., Davidson, R. A., & Liu, H. (2006). Statistical models of the effects of tree trimming on power system outages. *IEEE Transactions on Power Delivery*, *21*(3), 1549–1557.

Gunning, D., & Aha, D. (2019). Darpa's explainable artificial intelligence (XAI) program. *AI Magazine*, *40*(2), 44–58.

James, J., Hou, Y., Lam, A. Y., & Li, V. O. (2017). Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks. *IEEE Transactions on Smart Grid*, *10*(2), 1694–1703.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Kerschen, G., & Golinval, J. (2002). Physical interpretation of the proper orthogonal modes using the singular value decomposition. *Journal of Sound and Vibration*, *249*(5), 849–865. http://dx.doi.org/10.1006/jsvi.2001.3930, URL https://www.sciencedirect.com/science/article/pii/S0022460X01939306.

Krivohlava, Z., Chren, S., & Rossi, B. (2022). Failure and fault classification for smart grids. *Energy Informatics*, *5*(1), 33.

Martino, A., De Santis, E., Baldini, L., Rizzi, A., et al. (2019). Calibration techniques for binary classification problems: A comparative analysis.. In *IJCCI* (pp. 487–495).

Martino, A., De Santis, E., & Rizzi, A. (2020). An ecology-based index for text embedding and classification. In *2020 international joint conference on neural networks* (pp. 1–8). IEEE.

Murphy, A. H., & Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *26*(1), 41–47.

Rivas, A. E. L., & Abrao, T. (2020). Faults in smart grid systems: Monitoring, detection and classification. *Electric Power Systems Research*, *189*, Article 106602.

Rizzi, A., Mascioli, F. M. F., Baldini, F., Mazzetti, C., & Bartnikas, R. (2009). Genetic optimization of a PD diagnostic system for cable accessories. *IEEE Transactions on Power Delivery*, *24*(3), 1728–1738.

Shahid, N., Aleem, S. A., Naqvi, I. H., & Zaffar, N. (2012). Support vector machine based fault detection amp; classification in smart grids. In *IEEE globecom workshops* (pp. 1526–1531). http://dx.doi.org/10.1109/GLOCOMW.2012.6477812.

Shi, J. V., Xu, Y., & Baraniuk, R. G. (2014). Sparse bilinear logistic regression. arXiv preprint arXiv:1404.4104.

Smola, A. J., Bartlett, P., Schölkopf, B., & Schuurmans, D. (2000). Probabilities for SV machines. In *Advances in large-margin classifiers* (pp. 61–73). MIT Press.

Socher, R., Chen, D., Manning, C. D., & Ng, A. (2013). Reasoning with neural tensor networks for knowledge base completion. *Advances in Neural Information Processing Systems*, *26*.

Strang, G., Strang, G., Strang, G., & Strang, G. (1993). *Introduction to linear algebra*: *vol. 3*, Wellesley-Cambridge Press Wellesley, MA.

Tang, Y., Kurths, J., Lin, W., Ott, E., & Kocarev, L. (2020). Introduction to focus issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics. *Chaos. An Interdisciplinary Journal of Nonlinear Science*, *30*(6), Article 063151.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne.. *Journal of Machine Learning Research*, *9*(11).

Vitiello, S., Flego, G., Setti, A., Fulli, G., Liotta, S., Alessandroni, S., et al. (2015). A smart grid for the city of rome: A cost benefit analysis. http://dx.doi.org/10.2790/092657, ISBN: 9789279468575 9789279468582 ISSN: 1018-5593, 1831-9424, URL https://publicationstest.jrc.cec.eu.int/repository/handle/JRC95273.

Wang, X., Wang, F., Hu, J., & Sorrentino, R. (2015). Towards actionable risk stratification: A bilinear approach. *Journal of Biomedical Informatics*, *53*, 147–155.

Xi, P., Feilai, P., Yongchao, L., Zhiping, L., & Long, L. (2017). Fault detection algorithm for power distribution network based on sparse self-encoding neural network. In *2017 international conference on smart grid and electrical automation* (pp. 9–12). IEEE.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A brief survey on history, research areas, approaches and challenges. In *Natural language processing and Chinese computing: 8th cCF international conference, NLPCC 2019, dunhuang, China, October 9–14, 2019, proceedings, part II 8* (pp. 563–574). Springer.

Zadrozny, B., & Elkan, C. (2001). Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. *1*, In *Icml* (pp. 609–616).

Zhang, D., Han, X., & Deng, C. (2018). Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE Journal of Power and Energy Systems*, *4*(3), 362–370.

Zhao, Z., Xu, G., Qi, Y., Liu, N., & Zhang, T. (2016). Multi-patch deep features for power line insulator status classification from aerial images. In *2016 international joint conference on neural networks* (pp. 3187–3194). IEEE.