IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 17, 2021, accepted February 27, 2021, date of publication March 9, 2021, date of current version March 18, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064858

Non-Technical Loss Detection in Power Grid Using Information Retrieval Approaches: A Comparative Study

AGNALDO APARECIDO ESMAEL[1,2], HUGO HELITO DA SILVA[2], TUO JI[2], AND RICARDO DA SILVA TORRES[3], (Member, IEEE)

[1]Institute of Computing, University of Campinas, Campinas 13083-852, Brazil
[2]CPFL Energia, Rodovia Engenheiro Miguel Noel Nascentes Burnier, Campinas 13088-900, Brazil
[3]Department of ICT and Natural Sciences, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology (NTNU), 6009 Ålesund, Norway

Corresponding author: Agnaldo Aparecido Esmael (aesmael@cpfl.com.br)

This work was supported in part by the Companhia Paulista de Força e Luz (CPFL Energia), within the Agência Nacional de Energia Elétrica (ANEEL's) Research and Development Program, under Grant PD-00063-3037/2018, in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)-Brazil under Grant 140463/2818-6, and in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

ABSTRACT Non-technical loss (NTL) detection is a persistent challenge for Distribution System Operators. Data-driven solutions have been widely used nowadays to analyze customers' energy consumption and to identify suspicious fraud patterns for a posterior on-field inspection. However, the usage of such techniques, in particular the current deep learning methods, is not trivial and requires special attention to tackle imbalanced-class and overfitting issues. In this paper, we propose a new non-technical loss detection framework, which combines the effectiveness of convolutional neural network feature extractors with the efficiency of the Information Retrieval paradigm. In our solution, state-of-the-art pre-trained convolution neural networks (CNNs) extract deep features from electricity consumption time series represented as images. Next, these deep features are encoded into textual signatures and indexed using off-the-shelf solutions for posterior fraud searching. With this framework, the user can search for a specific fraud pattern in the utility database without having to train any classifier. The experiments performed in a real dataset provided by CPFL Energia, one of the largest electric utilities in Brazil, presented promising results both in terms of effectiveness and efficiency for the detection of fraudulent customers. In the conducted comparative study, we evaluate different time series image representations and CNN feature extraction approaches with regard to NTL detection results. Experimental results demonstrate that the combination of the Recurrence Plot image representation with the VGG16 CNN presented the best performance in terms of both effectiveness and efficiency.

INDEX TERMS Content-based retrieval, deep learning, feature extraction, information retrieval, machine learning, non-technical loss detection, pattern analysis, power grids, and time series retrieval.

I. INTRODUCTION

The reduction of electrical energy losses represents a specific issue for each Distribution System Operator (DSO) [1]. These losses result from technical and non-technical sources. Technical losses are the energy dissipation that occurs naturally in the electric grid mostly in consequence of Joule's effect [2]. Non-technical losses (NTLs) refer to the amount of energy that is delivered but not accounted for [1], which usually occurs due to non-legitimate behavior of DSO's customers that perform some kind of illegal interference in the network (theft) or in the meters (fraud).

On-site inspection of the customer's meters is the major action taken by DSOs for the detection and mitigation of NTLs. However, it is economically unfeasible to inspect all consumers. A typical strategy relies on shortlisting candidates for field inspection. In order to improve the assertiveness in the definition of candidates, data-driven solutions, such as machine learning techniques, have been widely used by DSOs. In the literature, fraud detection methods based on Artificial Neural Networks, Decision Tree, Support Vector Machines, Random Forest, and Optimum Path Forest, for example, are very popular (the reader may refer to [2], [3]

The associate editor coordinating the review of this manuscript and approving it for publication was Farhana Jabeen Jabeen.

to have access to comprehensive descriptions of existing initiatives).

Despite of good results achieved by machine learning approaches in some studies, the practical usage of such techniques in the NTL domain is not trivial and presents some drawbacks. First, there is no single solution capable of solving all cases of energy theft and fraud [2]. Second, most of the current proposed techniques were designed to smart meters, which are not a reality in many countries yet [4]. And third, the good performance of these methods is directly related to the quality of the features provided as input. These features are extracted from raw data in a pre-processing step that takes into account the DSO operators' expertise to define which information is most relevant to be used. However, there is no consensus in the literature about which features should be chosen [5].

Since there is no trivial way to characterize relevant patterns contained in the raw data, a potential solution is to explore ways to easily accommodate different (and eventually new) kinds of frauds by taking advantage of data-driven strategies. Current Deep Learning (DL) architectures [6] have been demonstrated to be a promising alternative to realize this task, given their remarkable success in several applications, such as medical image analysis [7], financial forecasting [8], salient object detection [9], and even for NTL detection [10]–[14]. A DL method is basically a multi-layered artificial neural network that learns hierarchically ways to represent the input data during its training step. The simplest representations are learned in the first layers and passed to the subsequent layers, which gradually generate more abstract features. In this way, the network itself discovers the most relevant information in the raw data with the goal of increasing the effectiveness performance in the target task. Then, the need of performing complex feature engineering is eliminated. However, the use of DL for the NTL detection is not trivial as well. Deep Learning methods often require huge labeled and balanced training sets to achieve high accuracy, which makes its practical use challenging for DSOs, since they commonly deal with imbalanced datasets due to the fraud sample scarcity.

How to explore the power description capability of deep learning solutions in an efficient and effective way, without dealing with the challenge of training models based on imbalanced class samples? This is the key challenge addressed in our work.

The aim of this paper is to introduce a novel NTL detection framework that takes advantages of the deep learning feature extraction power without facing the need of handling the associated heavy computational burden related to training it from scratch. Unlike other solutions, our framework models the problem of identifying fraudulent customers as an Information Retrieval (IR) task. Rather than training a classifier, we use pre-trained state-of-the-art Convolution Neural Networks (CNNs) to extract relevant features from customer electricity consumption data and then we encode this deep features as textual signatures. In this way, it is possible to index and search such signatures by using any available full-text search engine. With this framework, a user can search for a specific fraud type by using as query a time series of a known fraudulent consumer. In our method, there is no need to train any classifier, which makes it robust to the imbalanced data problem. Recall that time series associated with uncommon frauds could be used as query input in our system, which would return other collection time series with a similar pattern, i.e., time series associated with *candidate* frauds.

Our framework addresses both effectiveness (quality of fraud detection) and efficiency (short time for fraud search) aspects at the same time. In our formulation, consumption time series patterns are encoded into bi-dimensional representations (e.g., images). CNN-based features are then extracted from these images and transformed into textual signatures, which are then indexed using off-the-shelf IR technologies. Our solution is scalable to handle constantly evolving collections, and flexible to support the fraud detection problem even in scenarios where new fraud types are developed over time.

In summary, this paper has three main contributions:

1) We introduce a new formulation for the NTL detection problem based on the Information Retrieval paradigm;
2) We introduce a new framework that integrates *effective* image-based feature extractors with *efficient* and widely consolidated text-based search engines;
3) We perform a comparative study involving state-of-the-art deep learning based feature extractors and time series image-based representations.

We designed an evaluation protocol that considers real consumption data and fraud information associated with two collections associated with two different Brazilian cities. We then perform a comparative study with the goal of investigating which time series image representations (e.g., recurrence plot, spectogram, markov transition field, and Grammian angular field) and CNN feature extraction approaches (e.g., DenseNet121, InceptionResNewV2, InceptionV3, MobileNet, ResNet50, VGG16, and VGG19) would lead to effective NTL detection results. Experimental results demonstrate that the combination of the Recurrence Plot image representation with the VGG16 CNN presents the best performance in terms of both effectiveness and efficiency.

The remainder of this paper is organized as follows. In Section II, we discuss related work. Section III presents a brief overview of the information retrieval approach. We introduce our proposed framework in Section IV. Section V presents the protocol used in the experiments, and discusses the experimental results. Finally, our conclusions and directions for future work are discussed in Section VI.

II. RELATED WORK

Motivated by the fact that deep learning can learn new features from raw data, the use of deep learning for NTLs detection has been successfully explored in the literature recently.

TABLE 1. Comparison of the proposed approach with previous works.

Method	Data source used	No. of customers	Type of data	ML Algorithms
[11]	real electricity consumption	42372	Daily EC	Wide+CNN
[12]	real on-field inspections	106539	Weekly EC	LSTM+MLP
[13]	real electricity consumption	9655	Daily EC	LSTM+CNN
[14]	real electricity consumption	42372	Daily EC	CNN+Dilated Convolution
Proposed	real on-field inspections	9512	Monthly EC	CNN+Information Retrieval

For example, a study comparing the performance of different deep learning architectures, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Stacked Autoencoder, was done in [10]. Experiments were carried out on a synthetic dataset and showed that the CNN outperformed other classifiers. In another path, Zheng *et al.* [11] introduced a Wide and Deep Convolutional Neural Networks (CNN) framework to identify electricity thieves on secure smart grids. The Wide component is a fully connected layer of neural networks that is responsible for learning the global features from the 1D time series of electricity consumption data. The Deep CNN component reveals if there is periodicity or not on the electricity consumption of the customers. According to the authors, non-periodic fluctuations are hints of fraud. In order to properly use the CNN architecture, they designed a 2D manner of data representation by splitting the time series of consumption in sets of seven days (weekly) and stacking them to form a sort of image. This framework was tested on a real dataset that was made available by the State Grid Corporation of China (SGCC). The results obtained are significantly superior to traditional methods, such as Random Forest, Support Vector Machine (SVM), and Linear Regression.

Bazau *et al.* [12], in turn, proposed a novel Deep Learning approach, which combines both sequential and non-sequential data, to detect frauds in smart meters. Sequential data, which correspond to raw daily energy consumption of customers, are analysed through a long short-term memory network (LSTM). Non-sequential data are auxiliary information, i.e., geographical, contractual, and economic data, and they are passed through a multi layer perceptron (MLP) module. A hybrid module combines LSTM and MLP outputs to predict fraud cases. The experimental results showed that the incorporation of auxiliary data significantly improves the performance. Moreover, the model outperformed previous two deep learning methods used as baselines. In a similar approach, Hasan *et al.* [13] developed a CNN-based LSTM (CNN-LSTM) model for smart grid data classification. Recently, Finardi *et al.* [14] proposed a hybrid multi-head self-attention dilated convolution method to address the issues of the training in imbalanced data. The method takes advantage of both attention mechanisms and convolutional layers that are concatenated and unified through a convolution of kernel size 1. Besides it achieved higher accuracy score than the CNN baseline in performed experiments, their method presented a fast convergence time (just 20 epochs of training to converge against the 100 of baseline).

Table 1 summarizes the main features of related work that explores machine learning solutions based on convolutional neural networks. Similarly to the above initiatives, we also explore the power of CNNs in the characterization of times series consumption patterns. Different from those initiatives, however, we do **not** formulate the NTL detection as a classification problem. Instead, we utilize such CNNs as feature extractors that lead to textual signature representations that can be efficiently indexed by off-the-shelf Information Retrieval (IR) approaches. The formulation of the NTL detection problem as a search problem opens up new possibilities for handling a new variety of scenarios, such as:

- The proposed method can be employed to handle imbalanced distributions of fraud types.
- The proposed method can be used both with traditional and smart meters.
- The proposed approach is not only effective, but also is straightforward scalable to work efficiently with massive amount of data.
- Fraudsters are always looking for ways to bypass DSO's control mechanisms, so it takes some time to have enough samples of these new frauds to train models to identify them. With our framework, as soon as a new type of fraud is identified, it can be used as a query sample to search for similar misconducts based on time series patterns.

The integration of time series representations and image description approaches have been explored before [15]–[17]. Menini et *al.* [15], for example, investigated the use of texture descriptors based on LBP to characterize patterns of recurrence plot images constructed based of time series associated with vegetation indices. Dias et *al.* [16], [17] approached the same problem, now exploiting different CNN-based feature extractors. All those formulations, however, were validated in the context of classification problems involving remote sensing images. The studies of Santos *et al.* [18] and dos Santos *et al.* [19], [20] explored the possibility of extracting textual signatures from images, a strategy also used in our framework. None of those initiatives, however, explored CNNs in the feature extraction process. This literature gap was addressed by Amato et *al.* [21], which investigated state-of-the-art CNN-based image feature extractors for computing textual signatures. We adopt their formulation in the implementation of our IR-based NTL detection framework. Different from them, however, we integrate this textual signature extraction pipeline with image-based time series representations. In summary, to the best our knowledge, our work is

the first one to explore textual terms defined in terms of deep features associated with image representations of time series.

III. INFORMATION RETRIEVAL MODEL

Information retrieval aims at finding objects (e.g., textual documents) of an unstructured nature that satisfies an information need from within large collections stored on computers [22]. Due to advances in technologies of digital data acquisition and storage, huge and growing amount of data collections, ranging from financial to multimedia content, are available currently. In this context, it is essential to develop and use appropriate information systems to properly manage these massive collections [23]. The basic aspects of IR paradigm are briefly introduced as following.

A. FORMALIZATION

Let $\mathcal{C} = \{o_1, o_2, \ldots, o_m\}$ be a *collection* of m elements, where each element o_i of $\mathcal{C}$ is a digital object. Let $\mathcal{D}$ be a *descriptor*, which is defined [23] as a pair (ϵ, ρ), where:

- $\epsilon : o \rightarrow \mathbb{R}^n$ is a function to assign an *feature vector* $\boldsymbol{v}_o$ to object o.
- $\rho : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$ is a function that computes the distance between two *features vectors*.

The *similarity* of two objects o_i and o_j is obtained by using the descriptor $\mathcal{D}$ to compute the distance between them, which is given by $\rho(\boldsymbol{v}_{o_i}, \boldsymbol{v}_{o_j})$. The shorter the distance, the greater their similarity.

A *query* q is one sample of the digital objects that one wants to access from a collection and that is used as input in the retrieval systems. A *rank list* $\tau_q = \{o_1, o_2, \ldots, o_l\}$ is generated in response to query q, such that $\tau_q \subset \mathcal{C}$ has size l, with $l \ll m$, and its elements are sorted in decreasing order of similarity [24].

The *rank* (position) of an object o_i in the ranked list τ_q is denoted by $\tau_q(i)$. Therefore, if o_i is ranked before o_j in the ranked list, i.e. $\tau_q(i) < \tau_q(j)$, then $\rho(\boldsymbol{v}_q, \boldsymbol{v}_{o_i}) \leq \rho(\boldsymbol{v}_q, \boldsymbol{v}_{o_j})$, indicating that o_i is more similar to q than o_j [25].

IV. PROPOSED METHOD

This section introduces the proposed framework for non-technical loss detection based on information retrieval approaches.

A. FRAUD RETRIEVAL SYSTEM

Fig. 1 illustrates the main components of the proposed framework. The non-technical loss detection framework proposed is based on the typical information retrieval paradigm. The idea is to use the energy consumption data of a known fraudulent customer as a query, with the goal of selecting other suspicious cases of irregularity.

Basically, the proposed framework encompasses two stages, one off-line and another on-line. The off-line stage involves extracting features from the clients' energy consumption time series and encoding these features into textual documents, and then indexing them using state-of-the-art information retrieval approaches.

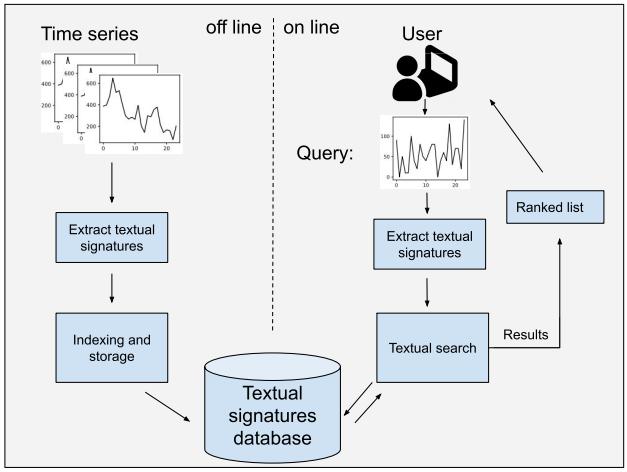


FIGURE 1. Overview of the proposed framework: A new non-technical loss detection system modeled according to an information retrieval paradigm.

In the on-line part, the user provides a time series that has a known fraud pattern as a query sample, the framework extracts features from this sample, generates its representation using textual terms, and uses a text search engine to retrieve from the time series collection the series that are similar to the query sample. The result is ranked in decreasing order of similarity before being displayed to the user. It is expected that other fraud cases are identified and presented at the initial positions (top ranks) of the resulting ranked list.

The steps of indexing, storing, and searching are expected to benefit from standard text search engines available nowadays, such as Lucine[1] or ElasticSearch.[2] These search engines have been successfully explored to efficiently manage large amounts of data in several applications. We then believe that their use in the domain considered in this study is feasible even considering that Distribution System Operators typically handle millions of customers, whose time series consumption patterns need to be analyzed regularly (e.g., monthly).

In the following, the main components of the proposed framework are described.

B. ON THE EXTRACTION OF TEXTUAL SIGNATURES FROM TIME SERIES

The generation of textual signatures is a key step of the proposed framework since their descriptive quality has a high impact on the effectiveness of the method in identifying consumption patterns that suggest the practice of fraud.

As pointed out before, very promising results have been obtained through the combination of bi-dimensional representations of times series (e.g., images) with state-of-the-art

[1] https://lucene.apache.org/core/ (As of Dec. 2020).
[2] https://www.elastic.co/ (As of Dec. 2020).

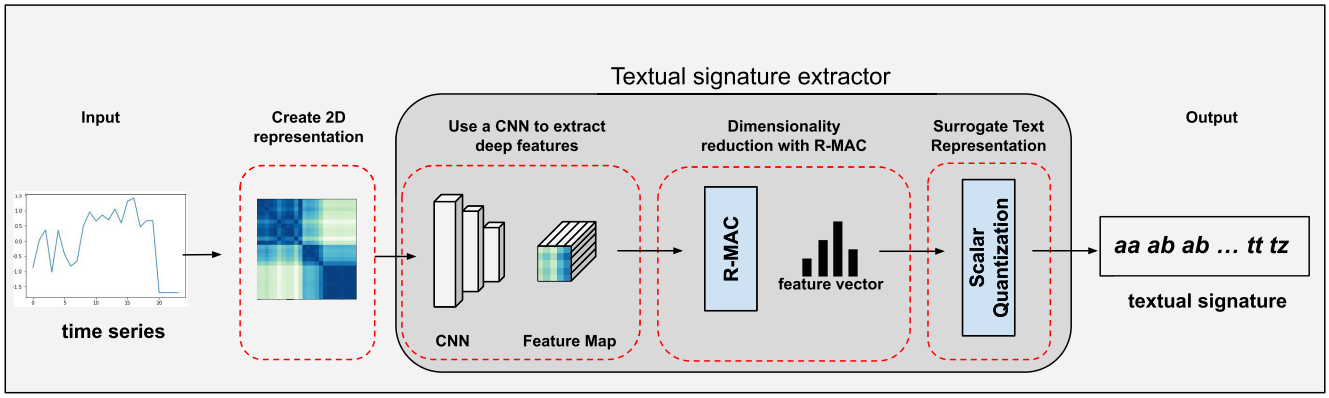


FIGURE 2. Steps for generating textual signatures from time series.

machine-learning-based image feature extractors. We follow this path in the construction of our feature extractor.

To the best of our knowledge, this is the first work that integrates time series representations with image feature extractors in a fraud detection setting.

Figure 1 provides an overview of the adopted pipeline for time series feature extraction. The textual signature extraction comprises four steps. In the first step, a time series T, consisting of the monthly electricity consumption (1D signal), is encoded as an image (2D data). This encoding can be done by different ways, Section IV-B1 presents more details. Later, a textual signature extraction procedure is employed to create a textual representation that is expected to encode different patterns of the input time series.

1) TIME SERIES: IMAGE REPRESENTATION

Time series can be represented as images. Fig. 3 shows four different visual representations of time series that are used in this work. Bellow, we present details of each one of them.

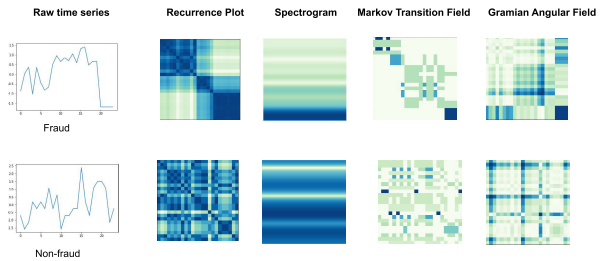


FIGURE 3. Four visual representations of time series.

a: RECURRENCE PLOT

In many applications, the analysis of recurrent behaviors is an important way to obtain an intuitive idea of the underlying dynamics of the complex systems [26]. However, these behaviors are usually difficult to be visualized in the time domain. Eckmann *et al.* [27] proposed a new tool, named Recurrence Plot (RP), that allows visualizing high-dimensional time series of complex systems as a square matrix, in which the matrix elements correspond to those times at which a state of a dynamical system roughly recurs. The visual patterns formed in a RP representation provide relevant information about the system, including its determinism and periodicity.

Mathematically, RP is defined according to Equation 1:

$$R_{i,j}(x(t)) = \begin{cases} 1, & \text{if } d(x(i)), x(j)) < r, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

where $R_{i,j}$ is the value of the position (i, j) in the RP matrix, which is 1 if the distance between the elements $x(i)$ and $x(j)$ of the time series $x(t)$ is smaller than a threshold r, or 0 otherwise.

From this definition, RP consists visually in a binary image in which only pixels whose intensities are bellow threshold r are encoded. However, is not intuitive to find an appropriate threshold value. Then, a variation of original RP, called distance plot (or unthresholded recurrence plot), is commonly used to eliminate the need of threshold setting. In this variation, all distances $d(x(i)), x(j))$ are plotted and RP looks like a gray-scale image.

b: SPECTROGRAM

A spectrogram is a visual representation of how the spectrum of frequencies of a signal is varying over time. The vertical axis of a spectrogram, in general, represents time, while the horizontal axis represents the discrete frequency steps. The amplitude of a particular frequency at a particular time is represented by the intensity or color of each point in the image [28]. Spectrograms have been widely used in the fields of music [29], speech processing [30], electroencephalography (EEG) analysis [31], and others.

For the generation of the spectrogram images (SI), the signal time series is segmented into fixed windows, which usually overlap, and then the Discrete Fourier Transform is applied to calculate the magnitude of the frequency spectrum

for each window. The transformed signal is defined as:

$$X_t(k) = \sum_{n=0}^{N-1} x(n)\omega(n)e^{-\frac{2\pi i}{N}kn}, \quad k = 0, \ldots, N-1 \tag{2}$$

where N is the length of window, $\omega(n)$ is Hamming window function, and k corresponds to the frequency $f(k) = kf_s/N$, where f_s is the sampling frequency in Hertz. The spectrogram of transformed signal $X_t(k)$ is defined by (3).

$$S(k, t) = |X_t(k)|^2 \tag{3}$$

The time-frequency matrix is then normalised into a grey-scale intensity image, with the range scaled between [0, 1], using (4).

$$SI(k, t) = \frac{S(k, t) - min(S)}{max(S) - min(S)} \tag{4}$$

c: GRAMIAN SUMMATION ANGULAR FIELD

Gramian Summation Angular Field (GASF) is a framework that encodes time series as an image based on polar coordinates [32]. First, the original time series is rescaled and smoothed, then it is transformed into a polar coordinate system by (5).

$$\begin{cases} \phi = \arccos(x_i), & -1 \leq x_i \leq 1, x_i \in \tilde{X}, \\ r = \dfrac{t_i}{N}, & t_i \in \mathbb{N}. \end{cases} \tag{5}$$

where t_i is the time stamp, N is a constant factor to regularize the span of the polar coordinate system, and $\tilde{X}$ is the rescaled time series [32].

Finally, the trigonometric sum between each point of the transformed time series is calculated to identify the temporal correlation within different time intervals, which is obtained with (6)-(7).

$$GASF = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} \tag{6}$$

$$GASF = \tilde{X}'.\tilde{X} - \sqrt{I - \tilde{X}^2}'.\sqrt{I - \tilde{X}^2} \tag{7}$$

I is the unit row vector $[1, 1, \ldots, 1]$.

d: MARKOV TRANSITION FIELD

A Markov Transition Field (MTF) is an image obtained from a process that computes dynamical transition statistics [32]. Given a time series $X = \{x_1, \ldots, x_n\}$, this the representation extraction approach identifies the Q quantile bins of X, assigns each x_i to the corresponding bins, calculates its Markov Transition Matrix W [33], and, finally, builds its MTF by (8).

$$MTF = \begin{bmatrix} w_{ij|x_1 \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_1 \in q_i, x_n \in q_j} \\ \vdots & \ddots & \vdots \\ w_{ij|x_n \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_n \in q_i, x_n \in q_j} \end{bmatrix} \tag{8}$$

where q_i and $q_j (q \in [1, Q])$ correspond to the quantile bins that contain the data at time stamp i and j, and w_{ij} denotes the transition probability of $q_i \rightarrow q_j$.

2) TEXTUAL SIGNATURE EXTRACTION

The textual signature extraction process explores successful formulations recently employed for supporting image searches [21] based on feature vectors computed by state-of-the-art machine learning feature extractors. Deep features are extracted from time series image representations using a pretrained convolutional neural networks (CNN). The reason for using pretrained CNN is to avoid the difficulties involved in the process of training a CNN from scratch, such as need of a large labeled dataset, defining a neural network architecture, and then prevent overfitting. The output of this step is a tensor (3-dimension data), also called "feature map" in the deep learning paradigm.

In the third step, the regional maximum activations of convolutions (R-MAC) descriptor [34] is applied to the feature map to reduce its dimensionality and preserve just the most informative features. R-MAC, which may be included in the top of pretrained convolutional layers of a CNN, computes a global representation of images independently of their size and without distorting their aspect ratio. This method extracts local features from several regions that are obtained from a rigid grid covering the image [35]. These local features are then max-pooled across several multi-scale overlapping regions. Next, these region-level features are independently $l2$-normalized, whitened with PCA, and $l2$-normalized again. Finally, the region descriptors are combined into a single vector by summing them and $l2$-normalized in the end [34]. The obtained image representation is a compact vector whose size ranged from 256 to 2k dimensions, depending on the CNN architecture [35]. Amato *et al.* [21] suggested the use of R-MAC among other methods because of its good effectiveness performance in image retrieval tasks.

Finally, in the last step, the numeric feature vector is transformed into a sequence of codewords resulting in a textual signature of the time series. This textual signature is obtained using the Scalar Quantization methodology recently proposed in [21].

It is worth to mention that the textual signature extraction framework proposed by Amato *et al.* [21] has not been explored before in the context of time series retrieval problems.

V. EXPERIMENTS AND RESULTS

This section describes the evaluation protocol adopted, as well as presents and discusses obtained results.

A. EXPERIMENTAL SETUP

The main steps of our experimental analysis were: identification of suitable datasets, clustering the fraud samples to identify different fraud types; data separation; textual signatures generation; dataset indexing; fraud searching; and performance assessment and comparative analysis. These steps are described next.

1) DATASETS

We carried out experiments in two datasets provided by CPFL Energia,[3] one of the largest private companies in the Brazilian electricity sector. These datasets consist of monthly consumption readings in kilowatt-hour (kWh) collected between Jan-2017 to Jun-2020 from 9512 residential customers of two medium-size cities of Brazil, Sorocaba-SP and Canoas-RS. All these customers were on-field inspected and labeled as fraud or no-fraud case. Table 2 shows the details of each dataset.

TABLE 2. Datasets considered in the evaluation protocol.

City	Number of samples	% of fraud samples
Sorocaba-SP	2243	19%
Canoas-RS	7269	10%

Since the date of on-the-field inspections also was availability, we constructed for each customer i the time series $t_i = \{x_1, x_2, \ldots, x_m\}$ by selecting the last m-consecutive energy consumption readings performed before the inspection (in this work, $m = 24$). Next, the time series were normalized by using the Z-Score method (9):

$$f(x) = \frac{x - \mu}{\sigma} \tag{9}$$

where x corresponds to the original data, μ, and σ, denote the mean, and the standard deviation, respectively. Next, the datasets were rescaled by appling the Logistic Sigmoid function (10):

$$g(x') = \frac{1}{1 + e^{-x'}} \tag{10}$$

where x' corresponds to the normalized data obtained from (9). These preprocessing transformations are important for two reasons: 1) the level of energy consumed by DSO's customers varies considerably from one to another, then it is essential to transform the time series to the same scale so they can be comparable; 2) CNN-based methods used in the feature extraction step expect that all features are centered around zero, have variance equal one, and are scaled to range [0,1] previously. All these requirements are achieved by employing (9) and (10), which are quite popular methods for normalization and scaling, respectively, in the Machine Learning literature.

2) CLUSTERING FOR IDENTIFYING FRAUD TYPES

Although our datasets were already labeled as fraud and no-fraud cases, we used the k-means algorithm to separate the fraud samples into groups of similar energy-consumption variance. The goal is to verify if the proposed framework is effective to identify fraudulent costumers that belong to the same clusters. We are making the assumption that an adopted procedure for committing a fraud will generate similar variances in the energy consumption time series of all customers who are committing a similar fraud. To define the number k of clusters, required by the k-means algorithm, we tried different values of k and evaluated the clustering performance with the silhouette coefficient. The $k > 2$ that presents the higher silhouette coefficient was chosen. Table 3 details the number of clusters chosen for each dataset and the number of fraud samples per cluster. Figures 4–5 depict examples of fraud patterns into each cluster for Sorocaba-SP and Canoas-RS datasets, respectively.

[3] `https://www.cpfl.com.br` (As of Dec. 2020).

TABLE 3. Samples per Clusters for each dataset.

Cluster #	Sorocaba-SP	Canoas-RS
1	35	112
2	73	163
3	55	107
4	39	104
5	65	-
TOTAL	267	486

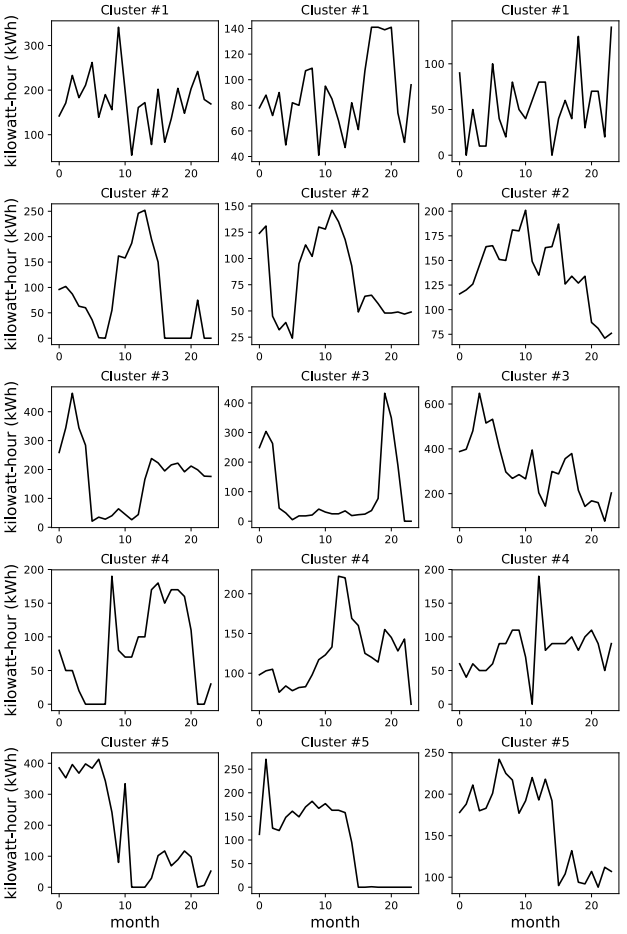


FIGURE 4. Examples of clusters related to the Sorocaba-SP dataset.

3) BASELINES

As baseline, we used the method proposed by Santos *et al.* in [18], which also extracts textual signature from time series

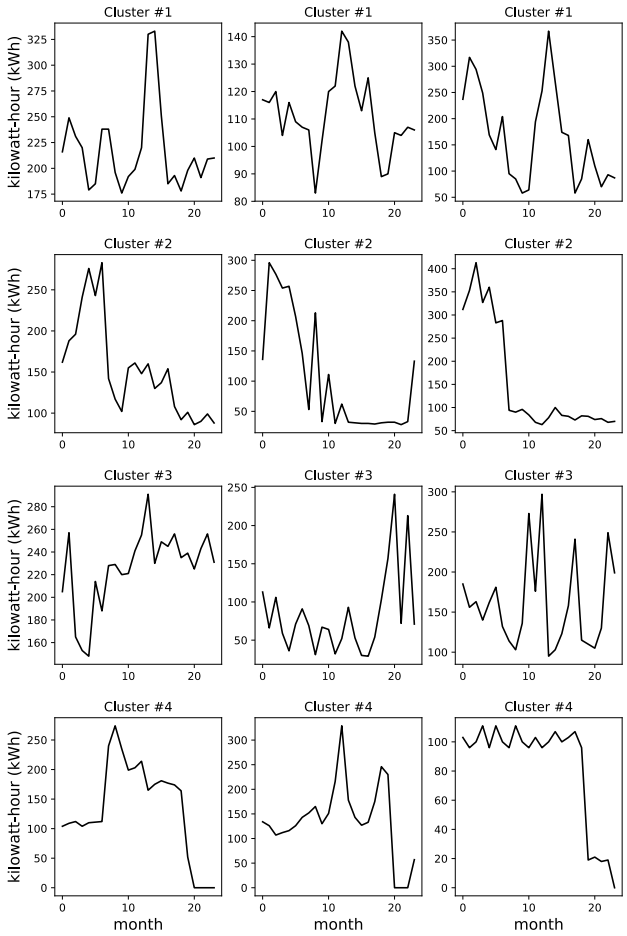


FIGURE 5. Examples of clusters related to the Canoas-RS dataset.

images to execute indexing and searching operations. However, their work does not explore deep learning approaches to extract features. Instead, the LBP texture descriptor [36] was employed in their study.

4) EVALUATION MEASUREMENTS

To evaluate the effectiveness of the proposed method, we considered two performance metrics widely used in the information retrieval domain: mean of precision at k ($mP@k$) and mean of recall at k ($mR@k$), which are briefly introduced as follows.

$mP@k$ is the mean of the precision scores obtained after a set of different queries has been performed, which is given as follows:

$$mP@k = \frac{\sum_{i=1}^{q} P@k}{q} \tag{11}$$

where, q is number of queries, and $P@k$ is the proportion of recommended items in the top-k set that are relevant (# of recommended items that are relevant divided by k).

$mR@k$ computes how many relevant objects were retrieved from the time series collections at the top-k positions:

$$mR@k = \frac{\sum_{i=1}^{q} Recall@k}{q} \tag{12}$$

where, q is number of queries, and $Recall@k$ is the proportion of relevant items found in the top-k results (# of recommended items that are relevant divided by the total # of relevant items).

5) EVALUATION PROTOCOL

We split the data into two folds: training set (1/3) and testing set (2/3). The training set is employed to fine tuning the parameters of the Scalar Quantization method, which is used in the last step of our pipeline. Textual signatures of the time series were then generated, using the methodology described in Section IV. Indexing and searching operations were carried out with Elasticsearch software, a very efficient and free search engine.

6) COMPARATIVE ANALYSIS

In the last step, we performed a comparative analysis to assess which pair (time series visual representation, Deep Neural Network feature extractor) is more suitable in terms of effectiveness and efficiency to be used in our proposed approach. There are different ways to represent a time series as an image, as seen in the Section IV-B1. In this work, we tested which of following four representation is the best for the proposed method: Recurrence Plot (RP), Gramian Summation Angular Field (GASF), Markov Transition Field (MTF), and Spectrogram (Spec). Since our method does not require the training of a deep learning model from scratch in its deep-feature extraction step, we also tested a pool of pre-trained models that were made available in Keras API[4] (version 2.2.4): DenseNet121, InceptionResNetV2, InceptionV3, MobileNet, VGG16, and VGG19. Recall that the proposed approach does not depend on any specific iterative procedure. Our solution benefits from *transfer learning*, a widely successful procedure adopted in several applications to explore pre-trained models [37]. Therefore, we do not perform any network training; we just use pre-trained models as feature extractor.

We adopted the effectiveness measurements described in Section V-A4. We run the experiments in an Intel Core i9-7900X CPU 3.30GHZ, 64 GB of RAM, and Ubuntu OS workstation.

It is important to highlight that, specifically for this work, we made a simplification in the pipeline shown in Fig. 1. Due to the small size of the time series used in the experiments, the R-MAC descriptor was not used. As explained in Section IV-B, the visual representations of time series have their dimensions determined by the size m of the time series. As we are using $m = 24$, the images generated in step 2 of our pipeline have a 24×24 dimension, which is below the minimum size accepted as input by Keras library. We resized our images to 48×48. As a result of these changes, the feature map generated as output by CNN's is already similar to a vector (1D), so there was no need to use the R-MAC.

[4] https://keras.io/api/applications/ (As of Dec. 2020).

B. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present the experimental results. The experiments were carried out in order to address the following research questions:

- RQ_1 Which time series image representation is more effective when handling the fraud detection problem as a information retrieval task?
- RQ_2 Which deep neural network is more effective to extract features from time series images?
- RQ_3 Are the proposed methods promising to handle fraud detection problems when compared with a baseline?
- RQ_4 Does the size of training set impact the effectiveness of the proposed method?

Table 4 shows the results of effectiveness for the city of Sorocaba-SP. As can be seen, the outcomes varied considerably according to the combination *Image Representation+Network* that is used in the proposed model. In terms of mP@5 and mRecall@5, the best performance scores were achieved with RP+VGG16, closely followed by RP+VGG19. In terms of mP@10 and mRecall@10, RP+VGG19 scored slightly higher than RP+VGG16. The baseline method, Santos *et al.*, presented the lowest scores for all types of images, indicating that it is less effective to properly characterize patterns from the time series visual representations than deep learning methods. The value of 25.4% achieved by our methodology significantly outperformed random predictions (19%). ANOVA with post-hoc Tukey test showed that there was no statistically significant difference between the RP+VGG16 and RP+VGG19, but this pair were significantly superior to all other configurations *Image+Network*.

Table 5 presents the results for the city of Canoas-RS. The results obtained are consistent with the Sorocaba-SP outcomes. RP+VGG16 and RP+VGG19 were the best, with a slight advantage for VGG16. The statistical test did not find any significant difference between these two combinations. The use of the baseline method again led to effectiveness scores of P@5, P@10, Recall@5, and Recall@10 that were lower than most of other combinations, corroborating the advantage that deep neural networks have over the classic methods of extracting features from images.

TABLE 4. Effectiveness Performance for pairs (image representation + feature extractor) for the Sorocaba-SP Dataset.

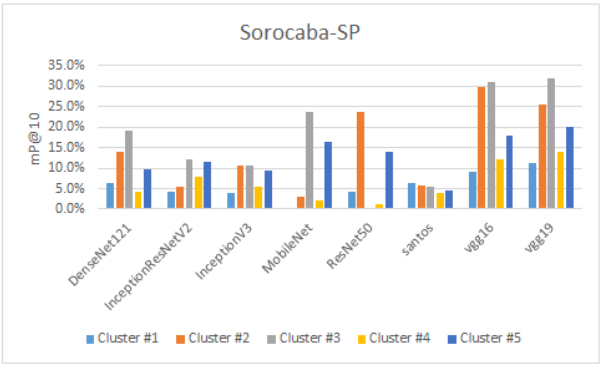
Setting (Image+Network)	mP@5	mP@10	mRecall@5	mRecall@10
GASF+DenseNet121	12.8%	11.3%	1.1%	1.9%
GASF+InceptionResNetV2	13.2%	11.0%	1.1%	1.9%
GASF+InceptionV3	10.3%	8.3%	0.9%	1.5%
GASF+MobileNet	10.9%	9.4%	1.0%	1.7%
GASF+ResNet50	17.2%	14.3%	1.6%	2.6%
GASF+VGG16	17.2%	15.7%	1.5%	2.8%
GASF+VGG19	16.9%	15.5%	1.6%	2.8%
MTF+DenseNet121	10.0%	10.0%	0.9%	1.7%
MTF+InceptionResNetV2	9.8%	9.2%	0.8%	1.6%
MTF+InceptionV3	9.0%	8.1%	0.8%	1.4%
MTF+MobileNet	8.9%	7.3%	0.8%	1.4%
MTF+ResNet50	6.3%	4.1%	0.6%	0.8%
MTF+VGG16	19.3%	15.4%	1.7%	2.6%
MTF+VGG19	20.9%	16.9%	1.8%	3.0%
RP+DenseNet121	12.8%	11.6%	1.1%	2.0%
RP+InceptionResNetV2	10.6%	8.6%	1.0%	1.6%
RP+InceptionV3	11.3%	8.7%	1.0%	1.5%
RP+MobileNet	11.1%	10.1%	0.9%	1.7%
RP+ResNet50	11.5%	10.7%	0.9%	1.6%
RP+VGG16	**25.4%**	**21.9%**	**2.2%**	3.7%
RP+VGG19	24.7%	**21.9%**	**2.2%**	**3.8%**
SPEC+DenseNet121	12.4%	12.8%	1.1%	2.2%
SPEC+InceptionResNetV2	12.0%	10.0%	1.1%	1.7%
SPEC+InceptionV3	13.1%	12.0%	1.1%	2.0%
SPEC+MobileNet	8.7%	9.4%	0.8%	1.7%
SPEC+ResNet50	1.4%	1.6%	0.1%	0.3%
SPEC+VGG16	19.0%	16.3%	1.7%	2.9%
SPEC+VGG19	13.9%	11.5%	1.3%	2.0%
Santos et al. (baseline)	7.0%	5.2%	0.7%	1.0%

TABLE 5. Effectiveness Performance for pairs (image representation + feature extractor) for the Canoas-RS Dataset.

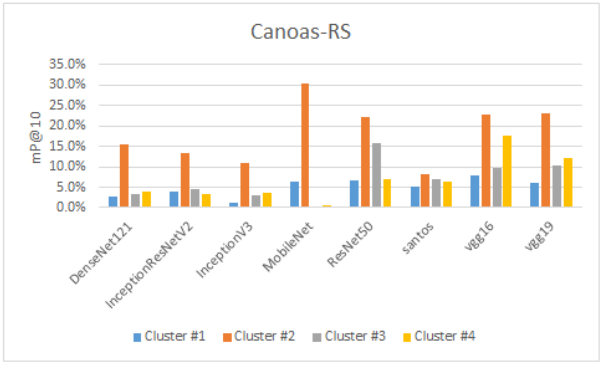
Setting (Image+Network)	mP@5	mP@10	mRecall@5	mRecall@10
GASF+DenseNet121	7.6%	7.1%	0.3%	0.5%
GASF+InceptionResNetV2	8.6%	7.7%	0.3%	0.6%
GASF+InceptionV3	5.3%	4.5%	0.2%	0.3%
GASF+MobileNet	11.2%	9.9%	0.4%	0.7%
GASF+ResNet50	16.1%	13.8%	0.6%	1.0%
GASF+VGG16	12.4%	11.0%	0.5%	0.8%
GASF+VGG19	15.9%	13.8%	0.6%	1.0%
MTF+DenseNet121	8.6%	9.0%	0.3%	0.6%
MTF+InceptionResNetV2	8.8%	8.5%	0.3%	0.6%
MTF+InceptionV3	10.2%	9.4%	0.4%	0.7%
MTF+MobileNet	7.9%	7.0%	0.3%	0.5%
MTF+ResNet50	2.7%	2.7%	0.1%	0.2%
MTF+VGG16	17.7%	16.3%	0.6%	1.2%
MTF+VGG19	16.8%	15.3%	0.6%	1.1%
RP+DenseNet121	7.5%	7.4%	0.3%	0.5%
RP+InceptionResNetV2	7.6%	7.2%	0.3%	0.5%
RP+InceptionV3	5.9%	5.3%	0.2%	0.4%
RP+MobileNet	11.0%	11.8%	0.4%	0.8%
RP+ResNet50	14.7%	14.0%	0.6%	1.1%
RP+VGG16	**18.1%**	**15.4%**	**0.7%**	**1.2%**
RP+VGG19	15.6%	14.0%	0.6%	1.1%
SPEC+DenseNet121	7.9%	7.1%	0.3%	0.5%
SPEC+InceptionResNetV2	6.8%	5.6%	0.2%	0.4%
SPEC+InceptionV3	6.8%	6.1%	0.3%	0.5%
SPEC+MobileNet	5.8%	5.4%	0.2%	0.4%
SPEC+ResNet50	0.9%	1.1%	0.0%	0.1%
SPEC+VGG16	13.8%	11.7%	0.5%	0.9%
SPEC+VGG19	12.9%	10.7%	0.5%	0.8%
Santos et al. (baseline)	9.5%	6.9%	0.4%	0.6%

The results of mP@10 per cluster when using RP to encode the time series are shown in Fig. 6. As can be seen, the use of the proposed framework was able to correctly identify cases of fraud in different clusters. In the Sorocaba-SP dataset, Fig 6-A, the samples from clusters 2, 3, and 5 were the ones with the highest mP@10 scores in each network. A possible explanation for this relies on the fact that the time series of these clusters have better defined consumption patterns (see Fig. 4). The superior performance of VGG16 and VGG19 compared to other networks was also evident in this figure. In the Canoas-RS dataset, Fig 6-B, all methods showed more modest results, which might have occurred due to the absence of consistent patterns in clusters 1 and 3 (see Fig. 5). Unlike what was seen in Sorocaba-SP dataset, ResNet50 presented a good outcomes together with VGG16 and VGG19.

In Fig. 7, we show a successful example of searching for a non-trivial case of fraud using our framework in Canoas-RS dataset. In general, fraud detection methods easily identify irregularities when there is a structural break (abrupt change) at a point of the time series. However, smart fraudsters take care to gradually reduce energy consumption in a way to make it difficult for utilities to detect their misconduct by statistical analysis. The query sample in this figure is an example of a long period of low decreases in energy consumption. Our



(a)



(b)

FIGURE 6. Cluster Analysis using Recurrence Plot images for different deep neural network feature extractors.

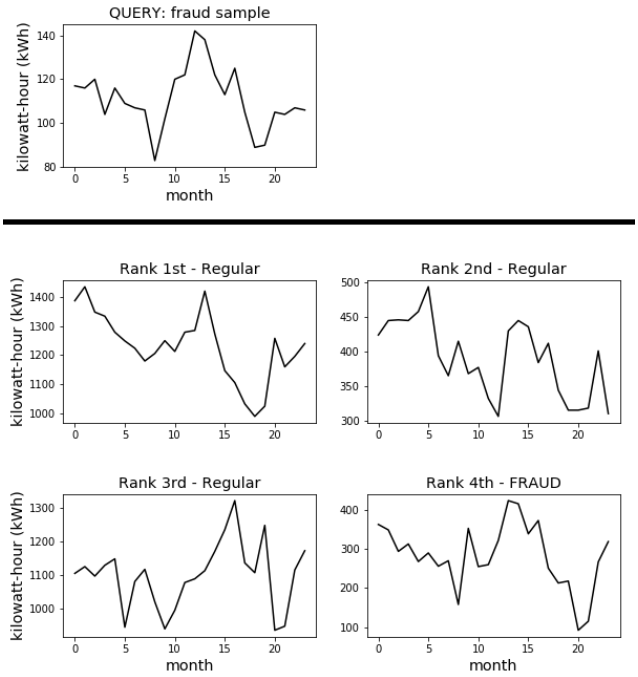


FIGURE 7. Fraud search example for Canoas-RS dataset. Top: monthly energy consumption time series of a actual fraudster consumer is used as a query sample. Bottom: first four positions of the ranked list generated by the retrieval system. A fraud case was detected in the fourth position. The ranked list was defined based on the use of RP+VGG16.

method was able to identify another similar fraud case and place it into the top-five positions of the ranked list. As can be seen, the identified irregularity has a mean consumption higher than the query sample, which shows the proposed framework is also robust to scaling factors.

Figure 8 shows the results of four queries performed in our experiments. Each query is defined according to the use of a different image representation computed from the same time series. The intention is to further analyze the effectiveness of the proposed method by also taking into consideration visual aspects of the generated ranked lists. As we can see, all queries returned cases of fraud. In special, the use of the Recurrence Plot representation (top row of the figure) produced a ranked list with five relevant cases at the 4th, 5th, 6th, 7th, and 10th positions. The use of the Spectrogram, Markov Trasition Field, and Gramian Summation Angular Field representations presented a bit worse performance; their respective ranked lists contain only three relevant cases. Except for the Spectrogram, most of the retrieved images exhibit little visual similarity with their respective query image. Moreover, there is a high variability in the visual patterns of the intra-ranking images. Note that just one case appears in many rankings (examples highlighted with a dashed red line). This suggests that different time series features have been encoded by each visualization method. Therefore, combining two or more of them may lead to better outcomes. We plan to investigate the use of fusion approaches, such as rank aggregation functions in future work. It is also worth to mention that, in terms of visual properties, Recurrence Plot was the method that better explored clear patterns for abrupt changes on the data, as it can be seen in the 5th, 9th, and 10th positions of the produced ranking. Since many frauds are associated with high variations on energy consumption, this fact may explain its better performance for the input time series. These findings corroborate the results presented in Tables 4 and 5 that already showed the superiority of the Recurrence Plot method for the NTL detection problem.

Table 6 shows the total searching time per dataset. As can be seen, the response time was low, showing the proposed method is also efficient in the execution of queries, which makes it a good choice for application in fraud detection problems.

Fig. 9 presents the execution time as a function of the number of queries for the Sorocaba-SP and Canoas-RS datasets. As the textual signatures used in the queries have different lengths, we expected that there could be a negative impact on the efficiency of the method. However, this did not happen. As can be seen, the time complexity always remained linear in both datasets for all CNNs tested, indicating the efficiency of the method was robust to changes in signature length.

In order to better evaluate the effectiveness of the proposed method regarding the size of the search space, we run experiments by partitioning the available data into different training/test sets proportions. Table 7 shows the results of the RP+VGG16 for the following three splits of the Canoas-RS dataset: 1) 1/3 training and 2/3 test, 2) 1/2 training and 1/2 test, 3) 2/3 training and 1/3 test. The percentage of

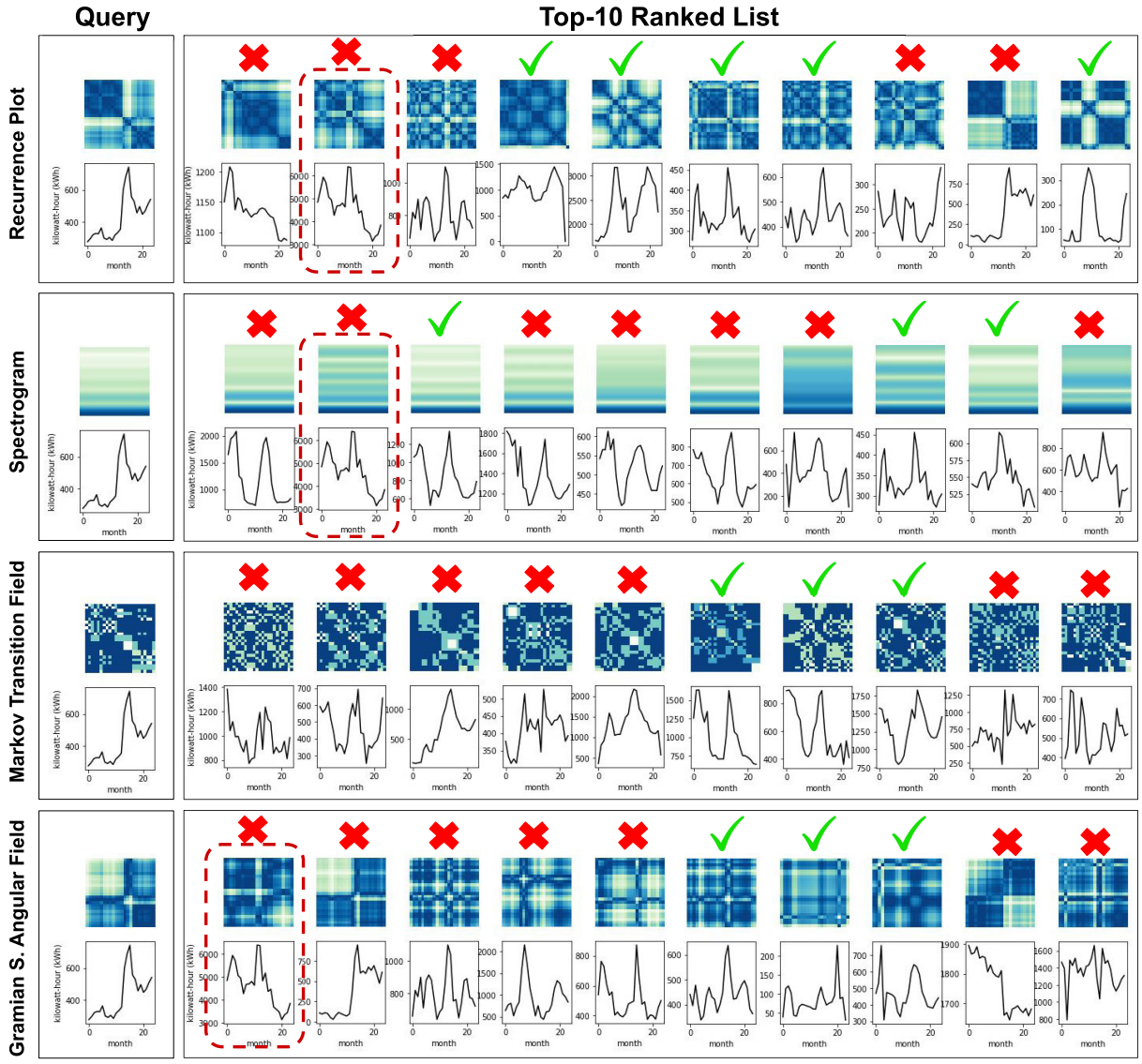


FIGURE 8. Examples of top-10 ranked lists with fraud candidates. Left side: Images used as query samples for the same time series. Searches are illustrated using different representations. From the top to the bottom, the Recurrence Plot, Spectrogram, Markov Trasition Field, and Gramian Summation Angular Field representations, respectively, were employed to perform searches on the Canoas-RS dataset. The corresponding raw time series plot is presented below each image. Right side: the first top-10 positions of the generated ranked lists. Green and red marks indicate whether the image on the list is a relevant or non-relevant case, respectively. We highlight in red one time series that was retrieved when three methods were used: Recurrence Plot, Spectogram, and Gramian Summation Angular Field. Given the query defined by the input time series, the VGG16 network was used to extract the deep features from images.

fraud cases within each set is approximately the same as the complete data. As the table shows, in all scenarios, we observe higher evaluation metrics scores as the training set size increases. We believe that such behavior happens in response to consequent increase of the number of training samples associated with fraud cases.

Based on the data above, we can answer the questions raised at the beginning of the section.

RQ_1 The use of recurrence plot as time series representations is the best to encode time series patterns;

RQ_2 Among the deep learning networks used, VGG16 and VGG19 are the most suitable for the feature extraction stage of the proposed framework. Thus, RP+VGG16 or RP+VGG19 are the best choices for the proposed framework.

RQ_3 The proposed method clearly outperformed the baseline and showed to be promising to handle fraud detection problems.

RQ_4 The effectiveness of the proposed solution improves as the size of the training set increases.

C. LIMITATIONS

Despite the effective outcomes achieved by our framework, we identified some cases for which our method presented a

TABLE 6. Cumulative search time (in seconds) for each dataset using RP images.

Networks	Sorocaba-SP	Canoas-RS
DenseNet121	15.805	66.354
InceptionResNetV2	47.909	168.722
InceptionV3	14.186	54.136
MobileNet	0.974	2.145
ResNet50	11.248	37.067
VGG16	6.328	25.219
VGG19	9.728	27.224

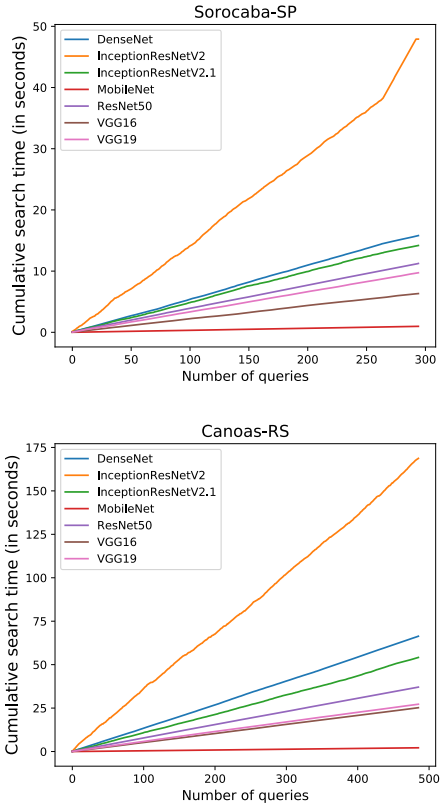


FIGURE 9. Efficiency analysis for different data-driven feature extractors. The figure presents the cumulative search time in seconds for different number of queries for the Sorocaba-SP and Canoas-RS datasets.

TABLE 7. Effectiveness performance for *RP+VGG16* method under different training set / test set proportions.

Dataset	Metrics	1/3 & 2/3	1/2 & 1/2	2/3 & 1/3
Sorocaba-SP	P@5	25.4%	27.3%	29.6%
	P@10	21.9%	22.5%	25.3%
Canoas-RS	P@5	18.1%	20.3%	21.0%
	P@10	15.4%	17.8%	18.5%

poor performance. In summary, we identified three types of failure that may occur, as follows:

- **Type #1:** The method correctly captures sudden and abrupt variations on the average level of energy consumption. However, it fails to recognize the direction of this variation, i.e., if it is a reduction or increase of consumption.

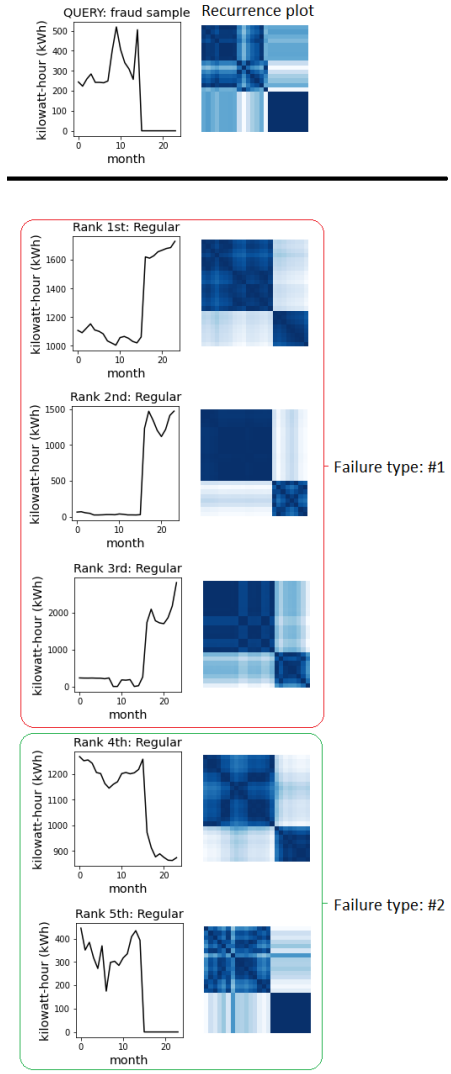


FIGURE 10. Examples of failure cases. Top: monthly energy consumption time series of a actual fraudster consumer is used as a query sample. Bottom: top-5 positions of the ranked list generated by the retrieval system and their respective recurrence plot images. No fraud case was detected for this query. Failure types #1 and #2 are highlighted. The ranked list was defined based on the use of RP+VGG16 for the Canoas-RS dataset.

- **Type #2:** Once the only information used by the method is energy consumption time series, all clients that significantly changed their consumption pattern during the analysis period are viewed as a potential fraudster. Nevertheless, many of these changes may be resulting from fair reasons, such as reduction of dwellers in a house, change of purpose concerning the use of the property (e.g., a commercial property that is changed to a residential one), renovation of the infrastructure of an industry consumer (e.g., replacement of non-efficient machinery).
- **Type #3:** The generated ranked list is never empty. This means that the model always outputs a set of suspicious clients even if there is no one real fraud case in the collection.

All the limitations presented above will produce false positives and, consequently, may lead to unnecessary on-the-field inspections, increasing operational costs.

Figure 10 illustrates failure cases of our method and sheds light on why these problems are occurring. This figure presents the top-5 positions of a ranked list that was generated in response to a fraud case used as query example. We show for each time series its visual representation based on Recurrence Plots. As it can be seen, no time series listed as fraud suspects is really a fraud and the failure types #1 and #2 are present. Regarding failure type #1, all the recurrence plots are visually very similar to the query one. This occurs because the recurrence plot computes the *module* of the difference of all the time series points. Then, independently of the fact that an increasing or decreasing of the consumption may have happened, the recurrence plot image will look like the same, which explains why our method is ''blind'' to the direction of the variations. However, this issue can be easily solved by computing, in a post-processing step, the linear regression of the time series and using the linear coefficient signal to filter false positives, i.e., cases in the ranked list, for which the signal of its linear coefficient disagree with the query one, must be ignored.

A post-processing step also can be used to address the failure type # 2. One possibility here is to analyze some customer profile information to refine the list, such as payment information, residence region, previous fraud occurrence, and so on. Failure type # 3 is more challenging. Maybe a direction to address it could be to define a similarity threshold. Only the cases that reached this threshold would be inserted into the ranked list. We plan to investigate such research direction in future work.

VI. CONCLUSION

In this paper, we proposed a novel framework to address the non-technical loss detection problem by modeling the identification of suspicious fraud cases as Information Retrieval tasks. The proposed approach relies on encoding the time series of monthly energy consumption of the utility customers into an image (2D data), whose content is characterized using state-of-the-art Convolution Neural Networks, which are powerful data-driven features extractor. Next, the content properties are converted into textual signatures in order to be later indexed and retrieved using full-text search engines. In this pipeline, our framework takes advantage of the transfer learning techniques to eliminate the need of training a machine learning classifier, which is a typical challenge in the NTL detection context because of the hard class imbalance in this domain.

We carried out experiments in a real electricity consumption dataset provided by CPFL Energia. Different time series image encoders and pre-trained CNN were tested to investigate which is the best setting for the proposed framework. Recurrence Plot image and VGG16 CNN presented the best performance in terms of both effectiveness and efficiency. The results obtained allowed us to conclude that this framework is a promising alternative to detect frauds in distribution grid especially in scenarios with very few fraud samples are available for training classifiers. Also, this solution is promising to deal with the evolution of fraud types over time.

In future work, we plan to investigate the use of techniques of re-ranking [38] and rank aggregation [24] to further improve the effectiveness results.We also plan to fine-tune the top layers of the VGG16 considering the target application.

Agnaldo Esmael is grateful to CNPq-Brazil (grant 140463/2818-6). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

REFERENCES

[1] (2017). *Reduction of Technical and Non-Technical Losses in Distribution Networks, CIRED WG CC-2015-2*. Accessed: Nov. 10, 2020. [Online]. Available: http://www.cired.net/files/download/188

[2] J. L. Viegas, P. R. Esteves, R. Melício, V. M. F. Mendes, and S. M. Vieira, ''Solutions for detection of non-technical losses in the electricity grid: A review,'' *Renew. Sustain. Energy Rev.*, vol. 80, pp. 1256–1268, Dec. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032117308328

[3] G. M. Messinis and N. D. Hatziargyriou, ''Review of non-technical loss detection methods,'' *Electr. Power Syst. Res.*, vol. 158, pp. 250–266, May 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378779618300051

[4] Y. Wang, Q. Chen, T. Hong, and C. Kang, ''Review of smart meter data analytics: Applications, methodologies, and challenges,'' *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[5] C. C. O. Ramos, J. P. Papa, A. N. Souza, G. Chiachia, and A. X. Falcao, ''What is the importance of selecting features for non-technical losses identification?'' in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2011, pp. 1045–1048.

[6] Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' *Nature*, vol. 521, pp. 436–444, May 2015.

[7] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, ''A survey on deep learning in medical image analysis,'' *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841517301135

[8] E. Chong, C. Han, and F. C. Park, ''Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies,'' *Expert Syst. Appl.*, vol. 83, pp. 187–205, Oct. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417417302750

[9] Y. Ji, H. Zhang, Z. Zhang, and M. Liu, ''CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances,'' *Inf. Sci.*, vol. 546, pp. 835–857, Feb. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025520308926

[10] R. R. Bhat, R. D. Trevizan, R. Sengupta, X. Li, and A. Bretas, ''Identifying nontechnical power loss via spatial and temporal deep learning,'' in *Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2016, pp. 272–279.

[11] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, ''Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids,'' *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1606–1615, Apr. 2018.

[12] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gomez-Exposito, ''Hybrid deep neural networks for detection of non-technical losses in electricity smart meters,'' *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1254–1263, Mar. 2020.

[13] M. N. Hasan, R. N. Toma, A.-A. Nahid, M. M. M. Islam, and J.-M. Kim, ''Electricity theft detection in smart grid systems: A CNN-LSTM based approach,'' *Energies*, vol. 12, no. 17, p. 3310, Aug. 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/17/3310

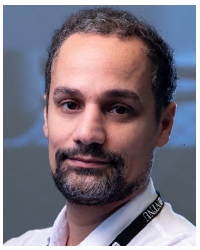
[14] P. Finardi, I. Campiotti, G. Plensack, R. D. D. Souza, R. Nogueira, G. Pinheiro, and R. Lotufo, ''Electricity theft detection with self-attention,'' 2020, *arXiv:2002.06219*. [Online]. Available: https://arxiv.org/abs/2002.06219

[15] N. Menini, A. E. Almeida, R. Lamparelli, G. L. Maire, J. A. D. Santos, H. Pedrini, M. Hirota, and R. D. S. Torres, "A soft computing framework for image classification based on recurrence plots," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 320–324, Feb. 2019.
[16] D. Dias, U. Dias, N. Menini, R. Lamparelli, G. L. Maire, and R. D. S. Torres, "Image-based time series representations for pixelwise eucalyptus region classification: A comparative study," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1450–1454, Aug. 2020.
[17] D. Dias, A. Pinto, U. Dias, R. Lamparelli, G. L. Maire, and R. D. S. Torres, "A multirepresentational fusion of time series for pixelwise classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4399–4409, 2020.
[18] E. S. Santos, B. Alberton, L. P. Morellato, and R. D. S. Torres, "An information retrieval approach for large-scale time series retrieval," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Yokohama, Japan, Jul. 2019, pp. 254–257.
[19] J. M. dos Santos, E. S. D. Moura, A. S. D. Silva, J. M. B. Cavalcanti, R. D. S. Torres, and M. L. A. Vidal, "A signature-based bag of visual words method for image indexing and search," *Pattern Recognit. Lett.*, vol. 65, pp. 1–7, Nov. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865515001956
[20] J. M. D. Santos, E. S. D. Moura, A. S. D. Silva, and R. D. S. Torres, "Color and texture applied to a signature-based bag of visual words method for image retrieval," *Multimedia Tools Appl.*, vol. 76, no. 15, pp. 16855–16872, Aug. 2017, doi: 10.1007/s11042-016-3955-4.
[21] G. Amato, F. Carrara, F. Falchi, C. Gennaro, and L. Vadicamo, "Large-scale instance-level image retrieval," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102100. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306457319301682
[22] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008. [Online]. Available: https://cds.cern.ch/record/2135372
[23] R. S. D. Torres and A. X. Falcao, "Content-based image retrieval: Theory and applications," *Revista de Informática Teórica e Aplicada*, vol. 13, no. 2, pp. 161–185, 2006.
[24] I. C. Dourado, D. C. G. Pedronette, and R. D. S. Torres, "Unsupervised graph-based rank aggregation for improved retrieval," *Inf. Process. Manage.*, vol. 56, no. 4, pp. 1260–1279, Jul. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306457318307647
[25] D. C. G. Pedronette, Y. Weng, A. Baldassin, and C. Hou, "Semi-supervised and active learning through manifold reciprocal kNN graph for image retrieval," *Neurocomputing*, vol. 340, pp. 19–31, May 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231219302309
[26] K. Iwayama, Y. Hirata, K. Takahashi, K. Watanabe, K. Aihara, and H. Suzuki, "Characterizing global evolutions of complex systems via intermediate network representations," *Sci. Rep.*, vol. 2, no. 1, p. 423, Dec. 2012.
[27] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhys. Lett.*, vol. 4, no. 9, pp. 973–977, Nov. 1987. [Online]. Available: http://stacks.iop.org/0295-5075/4/i=9/a=004?key=crossref.09fbeb6883f90a0adb050fbd7323bcd5
[28] T. A. Lampert and S. E. M. O'Keefe, "A survey of spectrogram track detection algorithms," *Appl. Acoust.*, vol. 71, no. 2, pp. 87–100, Feb. 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003682X09001959
[29] A. Ghosal, R. Chakraborty, B. C. Dhara, and S. K. Saha, "Song/instrumental classification using spectrogram based contextual features," in *Proc. CUBE Int. Inf. Technol. Conf. (CUBE)*. New York, NY, USA: Association Computing Machinery, 2012, pp. 21–25, doi: 10.1145/2381716.2381722.
[30] T. Özseven, "Investigation of the effect of spectrogram images and different texture analysis methods on speech emotion recognition," *Appl. Acoust.*, vol. 142, pp. 70–77, Dec. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003682X18300409
[31] M. Mustafa, M. N. Taib, Z. H. Murat, and N. H. A. Hamid, "GLCM texture classification for EEG spectrogram image," in *Proc. IEEE EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Nov. 2010, pp. 373–376.
[32] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Proc. 24th Int. Joint Conf. Artif. Intell. (IJCAI)*, Q. Yang and M. J. Wooldridge, Eds. Buenos Aires, Argentina: AAAI Press, 2015, pp. 3939–3945. [Online]. Available: http://ijcai.org/Abstract/15/553
[33] A. S. L. O. Campanharo, M. I. Sirer, R. D. Malmgren, F. M. Ramos, and L. A. N. Amaral, "Duality between time series and networks," *PLoS ONE*, vol. 6, no. 8, Aug. 2011, Art. no. e23378.
[34] G. Tolias, R. Sicre, and H. Jégou, "Particular object retrieval with integral max-pooling of CNN activations," 2015, *arXiv:1511.05879*. [Online]. Available: http://arxiv.org/abs/1511.05879
[35] A. Gordo, J. Almazán, J. Revaud, and D. Larlus, "End-to-end learning of deep visual representations for image retrieval," *Int. J. Comput. Vis.*, vol. 124, no. 2, pp. 237–254, Sep. 2017, doi: 10.1007/s11263-017-1016-8.
[36] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[37] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.
[38] F. Pisani, L. P. Valem, D. C. G. Pedronette, R. D. S. Torres, E. Borin, and M. Breternitz, "A unified model for accelerating unsupervised iterative re-ranking algorithms," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 14, p. e5702, Jul. 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5702

AGNALDO APARECIDO ESMAEL was born in Campinas, São Paulo, Brazil, in 1979. He received the B.Sc. and M.Sc. degrees in computer science from the University of Campinas (Unicamp), Brazil, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree in computer science. He is also a Data Scientist with CPFL Energia. His research interests include time series analysis, deep learning, and machine learning.

HUGO HELITO DA SILVA received the B.S. degree in statistics from the University of Campinas, Brazil, in 2005, and the M.B.A from Fundação Getúlio Vargas, Brazil, in 2011. He is currently the Head of Data Science with CPFL Energia, one of major energy company in Brazil. His research interests include machine learning, deep learning, and fraud detection models.

TUO JI received the B.S. degree in power system automation from the Huazhong University of Science and Technology, China, in 2008, and the Ph.D. degree from Washington State University, Pullman, WA, USA, in 2014. He is currently the Assistant Director of the Strategy and Innovation Department, CPFL, one of the major energy companies in Brazil. His research interests include machine learning and power system analysis.



RICARDO DA SILVA TORRES (Member, IEEE) received the B.Sc. degree in computer engineering and the Ph.D. degree in computer science from the University of Campinas (Unicamp), Brazil, in 2000 and 2004, respectively. He used to hold a position as a Professor at the Unicamp, from 2005 to 2019. He is currently a Professor in visual computing with the Norwegian University of Science and Technology (NTNU). He has been developing multidisciplinary eScience research projects involving multimedia analysis, multimedia retrieval, machine learning, databases, information visualization, and digital libraries. He has authored or coauthored more than 200 articles in refereed journals and conferences and serves as a PC member of several international and national conferences. He has been serving as a Senior Associate Editor for the IEEE Signal Processing Letters and an Associate Editor of the *Pattern Recognition Letters*.