



Article Detection of Non-Technical Losses in Irrigant Consumers through Artificial Intelligence: A Pilot Study

Vanessa Gindri Vieira ^{1,*,†}, Daniel Pinheiro Bernardon ^{1,*,†}, Vinícius André Uberti ^{2,†}, Rodrigo Marques de Figueiredo ^{2,†}, Lucas Melo de Chiara ^{3,†} and Juliano Andrade Silva ^{3,†}

- ¹ Graduate Program in Electrical Engineering-PPGEE, Federal University of Santa Maria—UFSM, Santa Maria 97105-900, RS, Brazil
- ² Polytechnic School, University of Rio dos Sinos Valley, São Leopoldo 93022-750, RS, Brazil; vuberti@unisinos.br (V.A.U.); marquesf@unisinos.br (R.M.d.F.)
- ³ Energy CPFL, Campinas 13088-900, SP, Brazil
- * Correspondence: vanessa.vieira@ufsm.br (V.G.V.); dpbernardon@ufsm.br (D.P.B.); Tel.: +55-55-3220-8344 (V.G.V.)
- ⁺ These authors contributed equally to this work.

Abstract: Non-technical losses (NTLs) verified in the power distribution grids cause great financial losses to power utilities. In rural distribution grids, fraudulent consumers contribute to technical problems. The Southern region in Brazil contains more than 70% of the total rice production and power irrigation systems. These systems operate seasonally in distribution grids with high NTL conditions. This work aimed to present an artificial intelligence-based system to help power distribution companies detect potential consumers causing NTLs. This minimizes the challenge of maintaining compliance with current regulations and ensuring the quality of services and products. In the proposed methodology, historical energy consumption information, meteorological data, satellite images, and data from energy suppliers are processed by artificial intelligence, indicating the suspicious consumer units of NTL. This work presents every step developed in the proposed methodology and the tool application in a pilot area. We detected a high number of consumers responsible for NTLs, with an accuracy of 63% and an average reduction of 78% in the search area. These results corroborated the effectiveness of the tool and instigated the research team to expand the application to other rice production areas.

Keywords: non-technical losses; rural electrical grids; artificial intelligence; pilot study; energy for agricultural processes; energy efficiency

1. Introduction

Commercial loss is one of the main problems in power distribution companies, contributing to financial losses. Several studies on energy loss detection have reported this problem. Glauner et al. [1] presented the most advanced research efforts in a detailed and up-to-date analysis of the algorithms, features, and datasets. They identified the most important scientific and technical challenges in detecting non-technical losses (NTLs) and suggested how these challenges can be addressed in the future. Shah et al. [2] proposed an algorithm to estimate actual energy losses when smart meter measurements are incorrect and to determine energy consumed by other sources of NTLs. They tested the algorithm using simulations to verify its effectiveness in the accurate identification of NTLs by determining technical losses.

To increase efficiency, power utility companies should minimize energy losses resulting from distribution system outages, anomalies in distribution grids, and fraudulent consumers. This is particularly crucial in rural distribution grids, which are usually large-sized and unautomated and operate on radial systems.



Citation: Vieira, V.G.; Bernardon, D.P.; Uberti, V.A.; de Figueiredo, R.M.; de Chiara, L.M.; Silva, J.A. Detection of Non-Technical Losses in Irrigant Consumers through Artificial Intelligence: A Pilot Study. *Energies* 2023, *16*, 6832. https://doi.org/ 10.3390/en16196832

Academic Editor: Simeon Olatayo Jekayinfa

Received: 5 July 2023 Revised: 29 August 2023 Accepted: 30 August 2023 Published: 26 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Our research question was: Is it possible to use artificial intelligence (AI) to identify irregular irrigation users? Irregular irrigation users are those committing fraudulent energy consumption. Identifying these irregularities allows inspection field teams to focus on the most suspicious sources of NTLs. To answer the research question, we defined a region to analyze consumers using irrigation systems for rice crops, which delimits the pilot study area. This work did not aim to prevent NTLs but to identify them in consumer units and help the power distribution company detect potential consumers causing NTLs.

Brazil is one of the top ten rice producers in the world, and rice production strongly depends on irrigation systems. In Brazil, crop production using different irrigation systems corresponds to 8.2 million hectares (Mha), and rice production occupies the second largest irrigation area, after sugarcane production, corresponding to 1.2 million hectares [3]. Thus, the analysis of energy consumption in irrigation systems becomes important.

In the pilot study area, the volume of water required by the flood-irrigated rice crops is approximately 6500 m³·ha⁻¹ during an average irrigation period of 90 days [4]. Thus, the large volume of water leads to high energy consumption in irrigation systems, becoming important to find economic losses.

Information on energy losses in Brazil and the rural distribution grids for irrigation systems will be presented. We also briefly analyzed some studies on NTLs.

1.1. Energy Losses in Brazil

In Brazil, the total energy losses in power distribution accounted for approximately 14% of the low-voltage market in 2021. This represented 43.9% of the total energy injected [5]. Figure 1 presents the energy losses comparing the total losses, technical losses, and NTLs over the years. In this period, total energy losses remained between 13% and 15% of the total energy injected into the customer. NTL is the difference between the total loss and the technical loss. In 2021, NTL was 6.5%. The total amount of energy losses corresponding to NTLs in power distribution was about 34.66 TWh, while technical losses were 40.40 TWh. Therefore, these energy losses significantly affect power utility companies and consumers.

ENERGY LOSSES IN BRAZIL



Figure 1. Energy losses compared to the energy injected over the years in Brazil.

The lack of equipment efficiency, NTLs, and power distribution grid overload are pertinent concerns in rural distribution grids. NTLs (commercial losses) refer to the amount of unbilled energy. Non-metered energy and fraudulent consumers lead to NTLs [6]. The cost of NTLs is paid by the power utility companies and/or legitimate consumers. NTLs occur both in urban and rural power distribution grids.

Figure 2 shows the irrigated crop area in Brazil. In 2017, the irrigated crop area was 6.7 million hectares, of which 25% were concentrated in the southern region [3]. More than 70% of the Brazilian rice is produced in the state of Rio Grande do Sul [7]. Sugarcane is the main crop in the first largest irrigated area, which accounts for 39.8% of irrigation.



Figure 2. Distribution of irrigation in the Brazilian regions.

The irrigation land use over time in the Brazilian regions shows that irrigation has been widely used in several crops. Over the decades, the irrigated crop area in southern Brazil has decreased.

Figure 3 shows that the Brazilian rice crop area has decreased in recent years [8]. However, the average productivity has increased since the irrigated crop area is proportionally more productive than the rainfed area. The productivity of irrigated rice crops is more than three times higher than non-irrigated rice crops.



Figure 3. Variation of the rice crop area (irrigated and rainfed) compared with the total production in Brazil.

Rice production has been relatively stable due to greater efficiency in water use and advances in crop management technology. Although the crop area has decreased, the crop yield levels have increased. In the last decade, the area under rainfed cultivation declined to less than 500 thousand hectares, while the area under irrigated cultivation was between 1.3 and 1.4 million hectares [8].

Figure 4 shows the amount of energy consumed by irrigation systems in the state of Rio Grande do Sul (RS) [9]. This location concentrates the irrigated rice crops in Brazil. The highest energy consumption was in western municipalities, ranging from 30 to 210 million kWh. These are the main irrigated rice crop areas. The energy cost for irrigation systems corresponds to approximately 7% of the total production cost, after the cost of fertilizers (15%) and water (9%) [10]. In the rice crop period, between October and February, energy consumption increases by more than 500 GWh to meet the irrigation demand. Irrigation efficiency should also be considered, along with concerns about meeting the demand in rural activities. The irrigation systems use motors and pumps to provide water to crops [11]. Thus, improving irrigation efficiency increases productivity and producer profits.



Figure 4. Energy consumption for irrigation in the state of Rio Grande do Sul (RS).

Rice irrigation by flooding uses ditches in the cropped area to conduct the water to where irrigation is required. This irrigation system has lower implementation and maintenance costs. Rice farms in southern Brazil have extensive irrigated areas, which require pumping stations to maintain a water layer during the irrigation cycle between 80 and 100 days. The energy consumption in pumping stations receives a differential tariff system, Horo-Seasonal green, and adopts a restricted schedule.

Köpp et al. [12] suggested reference values for sizes or pumping station operations to define the performance index (*Pi*). The authors proposed an acceptable index corresponding to a "good" rating of energy consumption for each hectare of irrigated crop according to Equation (1).

$$Pi < 750 \text{ kWh} \cdot \text{ha}^{-1} \tag{1}$$

Due to the high volume of rice crops in the state, the pilot study area considers irrigation users from one of the municipalities with the largest irrigated rice crop area. Irrigation users are good candidates for NTLs due to the large amount of energy consumed.

We cannot access accurate information about crop areas since the concessionaire does not have adequate mapping of crops, nor the cropped area in the western border region of the state. The equipment in rural grids is not telemetered, making it difficult to determine monthly consumption. As a result, most data are collected over long intervals. The commercial sector database is responsible for billing and presents many inconsistencies, making it impossible to use data from some consumers.

1.2. Detection of Non-Technical Losses

Several solutions for the detection of NTLs have been proposed, such as Omar et al. [13], Park and Kim [14], and Buzau et al. [15]. This diversity is due to different contexts and problem specificities.

De Oliveira Ventura et al. [16] studied the impact of NTLs on power distribution companies in Latin America. The NTLs reduce companies' revenues, and the electricity tariffs paid by the final consumers include part of these revenue losses. The percentage of NTLs distributed to consumers varies according to the national regulator.

Some studies applied big data to power distribution grids. However, many studies focused on specific services in low- and medium-voltage distribution grids. Big data applications in power distribution grids can perform fault detection, predictive maintenance, transient stability, state estimation, power quality monitoring, topology identification, load and profile, load failure, and NTL detection [17].

Savian et al. [18] analyzed how NTLs affect countries, utilities, and society. They explained the main barriers and strategies for detecting NTLs and analyzed the most important regulations of NTLs from different countries. They demonstrated the impact of NTLs on the economy and society and presented strategies to mitigate electricity fraud.

Saeed et al. [19] classified the techniques for detecting NTLs as either hardware-based or non-hardware-based. Teles Faria et al. [20] linked NTLs to particular populations and locations considering social factors associated with electricity fraud.

Hardware-based NTL methods use meters that have specific devices installed on consumer units. This device enables power distribution companies (PDCs) to detect any malicious activities by consumers, according to Viegas et al. [6] and Xia et al. [21]. Installing these devices on the consumer premises requires significant new infrastructure.

Advances in communications and data processing on energy consumer behavior allowed for the development of non-hardware-based methods for detecting NTLs. Thus, researchers are investigating this type of NTLs solution method, whose main focus is to detect the presence of electricity theft from the energy consumption data, as reported by Cui et al. [22], Khan et al. [23], and Feng et al. [24].

Messinis and Hatziargyriou [25] and Saeed et al. [19] categorized non-hardware-based methods into three groups: data-driven, network-driven, and hybrid. These methods require energy consumption measurements. The data-driven method uses data related to the consumer, such as personal and spatial technical characteristics, social information, and financial information. The network-driven method uses data such as topology and measurements from remote terminal units (RTUs) and observer meters [26]. The hybrid method combines the two previous ones.

Data-oriented methods are solely based on data analysis and machine learning. The methods are categorized as either supervised or unsupervised [25]. Supervised methods use labeled data, which can be sorted into two classes: positive/fraud or negative/not-fraud, such as support vector machine (SVM), artificial neural network (ANN), optimum-path forest (OPF), decision trees (DT), and nearest neighbor (k-NN). Unsupervised methods do not use labeled data.

Network-based methods use information from smart meters and calculate various physical parameters of the distribution grid [27]. The methods are classified based on the main concept or algorithm used, namely, state estimation, load flow, or special sensors for detecting fraud [26].

Hybrid methods share characteristics of both data-oriented and network-oriented methods. Some combinations have been proposed by Messinis et al. [26], such as combining SVM with observer meters to verify the energy balance and combining SVM with decision trees and observer meters.

Ahmad et al. [28] proposed several approaches to detect unauthorized energy consumption and other methods used for electricity theft. They identified various setbacks and problems that arise in the implementation of measures to control unauthorized energy consumption.

The detection of NTLs is critical for power utility companies. This challenge is even more pronounced for customers in rural distribution grids. Power utility companies conduct inspections to detect NTLs at selected customer locations based on predictions. Rural inspections are expensive due to the long distances that should be covered by the technicians during on-site inspections for NTLs [1]. Inspection in rural areas is difficult, requiring the displacement of technicians and the inaccessibility of some consumers due to vegetation and environmental conditions. Thus, investment in the accuracy of prediction should be important.

Despite the several studies on detecting NTLs, there are limited studies on rural energy consumers or irrigated crops. This work proposes a methodology to identify areas of interest related to rural energy consumption for irrigation systems. We define a model that uses AI algorithms and applied it to selected data to detect NTLs. To validate the proposed methodology and data selection, a pilot study area is considered. We also analyze the accuracy of the proposed methodology.

The main contributions of this work are as follows:

- The identification of a dataset from the meteorological data of the study area, historical energy consumption data, and crop information.
- The analysis of the selected data for rural energy consumers served by non-automated distribution grid for irrigation systems.
- The definition of a methodology based on AI algorithms to detect NTLs in irrigated rice crops.
- A validation procedure for the proposed methodology.

2. Methodology

The proposed methodology describes the steps and techniques for detecting NTLs in crop irrigation systems. The tool development is part of a pilot project in partnership with the Federal University of Santa Maria and the power utility company CPLF Energy (Companhia Paulista de Força e Luz). This project aims to detect NTLs in the state of Rio Grande do Sul, southern Brazil. This region concentrates most rice producers using crop irrigation systems.

We divided the proposed methodology into three main phases: (1) the selection of variables based on their relevance to the energy consumption for irrigation systems, (2) the development of an AI model, and (3) the validation and adjustment of the model based on field inspections. Figure 5 shows the flowchart of the proposed methodology, presenting the study area, the detection system for NTLs, and the result analysis.



Figure 5. Flowchart of the proposed methodology.

This flowchart starts with the analysis region selection by searching the consumer database. Then, the algorithm is executed to select the input variables for the fuzzy logic algorithm. These variables are selected from those related to energy consumption for irrigation systems. The fuzzy logic algorithm identifies suspicious consumer units of NTLs. The system prioritizes the suspicious consumer units, and an inspection list is sent to the field inspection team. Then, the results from field inspections are compared to the inspection list generated by the system.

First, we analyzed the most significant variables that affect the energy consumption in rice irrigation systems. We considered the meteorological variables, the variables related to rice crops, and the data from the energy supplier. We adopted correlation analysis to understand the relationships among these variables and excluded redundant or unrelated variables to understand their respective relationships with energy consumption.

Next, we treated the selected variables using normalization, synchronization, and missing data removal. In the second step, the previously processed variables serve as inputs to the AI algorithms, as described in Figure 6.



Figure 6. Flowchart of the irrigation and expert systems.

A fuzzy logic algorithm receives this information and predicts if additional irrigation is required by the rice crops analyzed. Next, an expert system indicates the consumer units that significantly differ from the expected energy consumption for the current rice crop. These consumer units are identified as suspicious sources of NTLs. A list of suspicious consumer units is generated for field inspection.

Finally, field inspection teams inspect the suspicious consumer units. The data generated by the field inspection will be used to improve the algorithm. This will foresee the expansion of the current pilot project into a permanent one at the power utility company.

2.1. Variable Selection

A study on technological mapping identified the most significant factors that contribute to energy consumption in rice irrigation systems in southern Brazil. The study contains a literature review, interviews with experts, and field inspections of irrigated rice crops.

Data on water usage in rice crops is directly associated with the need for supplemental irrigation. These data are related to the energy consumption by the water pumping system. Meteorological and crop variables directly impact energy consumption. The energy consumption pattern of the consumer unit in previous harvests is also a significant factor. Furthermore, the suspicious consumer units of NTLs are reinforced by the total energy loss rate of the energy supplier.

Four groups of variables are available: meteorological data, crop data, historical energy consumption, and electricity data from the energy supplier.

Meteorological data: the National Institute of Meteorology (INMET) [29] provides sixteen variables from weather stations in Brazil, such as precipitation; wind speed and direction; and average, maximum, and minimum values for temperature, humidity, atmospheric pressure, and dew point. Data from weather stations closest to a given region are interpolated to provide more accurate values. Figure 7 shows the weather stations located in southern Brazil (dark ellipses).



Figure 7. Location of the weather stations in southern Brazil.

Crop data: crop area, soil type, and crop type variables are available from satellite images and specific algorithms. Figure 8 shows part of the satellite image data processing used to recognize the crop area and, subsequently, classify the crop type [30]. January is considered the optimal month for acquiring satellite images to identify rice crops. At the end of the satellite image processing, the crop area, in hectares, is identified and calculated for each segment of the rice crop.



Figure 8. Satellite image processing for detecting flood-irrigated rice crops. (**a**) Sub-image using normalized difference vegetation index (NDVI), (**b**) Intermediate image with only crop areas, and (**c**) sub-image with limits of crop areas, resulting in crop segments.

Historical power consumption: historical energy consumption data are available from the power utility company database together with the installed capacity of the consumer unit. Electricity data: electricity data for rural consumer units can be accessed in the power utility company database, while data on energy losses are based on Reference [31].

This large number of data led us to select the most significant variables for the proposed methodology. Since some variables are mutually dependent, i.e., they carry redundant data, we can select only one of the correlated variables. Thus, we created a statistical model based on correlation analysis and direct selection.

For this step, we used a dataset comprising ten consumer units (irrigated rice crops) and three rice crop from 2019/20, 2020/21, and 2021/22. We defined a correlation value of 0.80. Figure 9 presents the results for the meteorological variables.

												32	4	Chill and	14	in.
		b		8	9		-	2 0	5	A		£		5 5	2 2	2
		200	B	9	1º	3 A	ie.	Hun	Hun	He.	40	a a	4	00	0º	00
	Rein	Min	Mar	Nun	400	Alle	tot a	200	MID	AND -	400	MID	to to	400	MIL	1 to
Rain																
Wind_Spd	-0.3654															
Wind_Burst	-0.3077	0.9707														
Wind_Dir	0.3135	-0.0730	0.1186													
Avg_Temp	-0.0357	-0.1929	-0.1618	-0.4007												
Min_Temp	0.0259	-0.2029	-0.1702	-0.3953	0.9997	а — — — — — — — — — — — — — — — — — — —										
Max_Temp	0.0342	-0.2033	-0.1732	-0.4041	0.9998	0.9995	Ĩ									
Avg_Hum	0.4066	-0.1572	-0.1223	0.5402	-0.6705	-0.6551	-0.6749									
Min_Hum	0.3901	-0.1209	-0.0845	0.5506	-0.6781	-0.6630	-0.6831	0.9989								
Max_Hum	0.4164	-0.1567	-0.1198	0.5439	-0.6651	-0.6499	-0.6697	0.9991	0.9976							
Avg_Atm_Press	-0.1191	0.0347	-0.0316	0.2084	-0.82.64	-0.8352	-0.8193	0.2370	0.2377	0.2281						
Min_Atm_Press	-0.1214	0.0297	-0.0377	0.2030	-0.8187	-0.8277	-0.8115	0.2270	0.2275	0.2182	0.9999					
Max_Press_Atm	-0.1211	0.0482	-0.0173	0.2124	-0.83 26	-0.8414	-0.8258	0.2419	0.2432	0.2328	0.9998	0.9995				
Avg_Dew_Point	0.1896	-0.3335	-0.2747	-0.2235	0.92.07	0.9287	0.9186	-0.3309	-0.3412	-0.3255	-0.9228	-0.9183	-0.92.84			
Min_Dew_Point	0.1963	-0.3332	-0.2752	-0.2216	0.9159	0.9241	0.9137	-0.3202	-0.3303	-0.3148	-0.9237	-0.9192	-0.92.92	0.9999		
Max_Dew_Point	0.1791	-0.3230	-0.2642	-0.2305	0.92.84	0.9360	0.9262	-0.3495	-0.3596	-0.3436	-0.9240	-0.9193	-0.92 95	0.9997	0.9993	

Figure 9. Results for the correlation analysis of the meteorological variables. The gray cells indicate high correlations.

From the sixteen variables available, we selected rainfall (precipitation), wind speed, wind direction, average temperature, average humidity, and minimum humidity. The remaining variables were excluded since they had a high correlation value with at least one of the selected variables.

The crop area, soil type, and crop type variables were included in the correlation analysis with the six meteorological variables to test their correlation with energy consumption. Although we stated the correlation value of 0.80 as a standard for variable selection, we also considered variables with higher correlation values. Table 1 presents the correlation analysis results for the meteorological variables. Rainfall (precipitation), wind speed, average temperature, and crop area were identified as the four variables correlated with energy consumption for supplemental irrigation of rice crops.

Table 1. Correlation between meteorological data and energy consumption data. The gray cells indicate high correlations.

	Cons. Unit 1	Rain	Wind_Spd	Wind_Dir	Avg_Temp	Avg_Hum	Min_Hum
Cons. Unit 1							
Rain	-0.6534603						
Wind_Spd	0.58037816	-0.3653789					
Wind_Dir	-0.1700813	0.31345217	-0.073045				
Avg_Temp	-0.0756694	-0.0356799	-0.1929023	-0.4006796			
Avg_Hum	-0.1504893	0.40660993	-0.1572344	0.54020871	-0.6705209		
Min_Hum	-0.1346516	0.39005796	-0.1209171	0.55059828	-0.6780857	0.99892129	

Energy loss data from energy suppliers is not directly related to energy consumption in irrigation systems. These data will be used later in the expert system step.

Correlation analysis plays a vital role and aims to improve the performance of the fuzzy logic algorithm. Reducing the number of variables decreases rule numbers and programming complexity of the algorithm. The number of inputs optimizes processing and computational costs during operation.

2.2. Fuzzy Algorithm

The fuzzy logic algorithm predicts the need for the supplemental irrigation of rice crops. The primary irrigation source for these crops comes from natural replenishment through precipitation. Evaporation, plant transpiration, and vertical water percolation through soil lead to water losses that are not naturally replaced. The main irrigation systems in southern Brazil use electric motors.

Irrigation demand is inversely proportional to precipitation but directly proportional to wind speed, average temperature, and crop area. The irrigated crop area variable was defined by analyzing data from satellite images of the region. The parameters of the meteorological variables were determined based on climate normals available by INMET [32].

Each variable was assigned to three membership functions (low, average, and high) based on the trapezoidal function. The parameters of each membership function were determined by combining statistical analysis and expert opinion. Figure 10 shows the fuzzification of the rainfall variable (precipitation).



Figure 10. Fuzzification of the rainfall variable.

The set of rules for the fuzzy logic algorithm is based on expert opinions about the impact of each variable on the need for supplemental irrigation. We consulted experts in meteorology, phytotechnics, and rice crops production. Then, historical data from the last three crop seasons were used to refine the rules.

The output of the fuzzy logic algorithm is defined as the necessary irrigation period within the crop, denoted by $period_f$ and measured in hours, proportional to the irrigation needs in that crop. Figure 11 illustrates the defuzzification and parameterization of the output variable.



Figure 11. Defuzzification of the output variable.

$$E = Pxperiod_f \tag{2}$$

$$CPA = E/A \tag{3}$$

This metric is compared to the statistical normal values for the analyzed region to identify consumer units that differ from the standard behavior. The expert system performs this step.

2.3. Expert System Algorithm and NTLs Suspicious Consumer Indication

According to Köpp et al. [12], the average energy consumption index per area (*CPA*) for rice irrigation in the pilot study area ranges from 550 to 750 kWh·ha⁻¹ per harvest. This approach aims to identify consumer units whose energy consumption is above or below the average range. We give particular attention to those consumer units whose energy consumption is below the average range, and they are tagged as suspicious consumer units of NTLs. The power utility company is also interested in consumer units whose energy consumption exceeds the average range since they indicate low efficiency in the irrigation system and may receive investments in energy efficiency projects.

The expert system considers the results of the energy consumption index per area of each consumer unit and intersects this information with the total energy loss index of the energy supplier in which the consumer unit is located.

The historical energy consumption behavior of the consumer unit is also compared to the nearby consumer units. For instance, in a particular crop, if the average behavior of the consumer group was to increase energy consumption and if the energy consumption varied in the opposite direction, it reinforces the suspicion of NTLs.

The expert system is developed using a fuzzy logic algorithm to incorporate the desired relationships between the three variables into the set of fuzzy rules. These results classify the suspicious consumer units of NTLs into very high suspicion, high suspicion, mean suspicion, low suspicion, or low efficiency.

Table 2 summarizes the standard inserted into the expert system to evaluate suspicious NTLs.

	Consumption Index per Area	Total Energy Losses	Behavior towards Neighbors
Very high suspicion	Below the average range	Above 25%	Variation in the opposite direction, greater than 35%
High sugnition	Below the average range	Above 25%	Variation in the opposite direction, between 1 and 35%
riigii suspicion	Below the average range	Below 25%	Variation in the opposite direction, greater than 35%
	Below the average range	Below 25%	Variation in the same direction
Mean suspicion	In the average range	Above 25%	Variation in the opposite direction
Low suspicion	In the average range	Below 25%	Variation in the same direction
Low efficiency	Above average range	Any situation	Variation in the same direction

Table 2. Guiding principles to fuzzy rules development.

For instance, a consumer unit classified as high suspicion of NTLs occurs when the energy consumption index per area is below the average range, while the total energy losses of the energy supplier is above 25% and the energy consumption behavior between the last harvests differs by more than 35%.

The expert system allowed us to generate a field inspection list, based on power utility company criteria.

2.4. Field Inspection

In the pilot project, we considered the total crop area of 75,800 hectares and about 475 consumer units for rice crop irrigation. Consumers were tested following the proposed methodology, to generate an inspection list of the suspicious consumer units associated with NTLs. Considering the total number of consumers, 90 were identified as the main suspects in the pilot study area. From these, a total of 60 were inspected by the inspection field teams to verify possible tampering and fraud in the power measurement systems. The next section describes the results of the field inspections.

3. Results and Discussion

The pilot study area comprised 5,702,098 km², as shown in Figure 12. This crop area is studied in two periods. The first one was between 2021 and 2022 and the second one between 2022 and 2023. The first crop period had 248 consumer units (2021/2022), while the second one had 227 (2022/2023).



Figure 12. Pilot study area.

Applying the proposed methodology to all consumer units selected, we can observe that 90 were pointed out as potential NTLs, as shown in Table 3. The suspicious consumer units of NTLs are classified from mean to very high suspicion. We also highlighted that the proposed methodology reduced the search space for NTLs by 73.79% in the first rice crop and 88.99% in the second one. In general, the search space for NTLs is reduced by 81.40%, as shown in Figure 13.



Table 3. Proposed methodology for consumer unit classification stratified by rice crop.

Figure 13. Reduction of the search space for suspicious and unsuspicious consumer units.

Field inspection lists are generated for all suspicious consumer units of NTLs detected. Since the number of unsuspicious consumer units is high, we selected a sampling of 10% of these consumer units to be inspected. Thus, for the first rice crop, 65 suspicious consumer units and 25 unsuspicious consumer units will be inspected. For the second rice crop, 25 suspicious consumer units and 20 unsuspicious consumer units will be inspected.

After field inspections, for the first rice crop, we confirmed 28 consumer units as responsible for NTLs of the total of 65 suspicious consumer units detected. All 25 unsuspicious consumer units sampled were confirmed unsuspicious. For the second crop, we confirmed 12 suspicious consumer units as responsible for NTLs of the total 25 detected, and all 20 unsuspicious consumer units sampled were confirmed.

By considering both crops together, there were 90 suspicious consumer units of NTL; 40 were confirmed as suspicious and 45 as unsuspicious. This confusion matrix analysis is shown in Table 4.

Actual Suspicious Unsuspicious Unsuspicious 40 0			Predicted				
Actual Suspicious 40 0			Suspicious	Unsuspicious			
Unsuspicious 50 45	Actual	Suspicious Unsuspicious	40 50	$\begin{array}{c} 0\\ 45 \end{array}$			

Table 4. Confusion matrix analysis of two crop seasons using the proposed methodology.

The confusion matrix analysis can determine the accuracy and error rate of the proposed methodology. The accuracy is the ratio between the true and false positive results and the total number of observations, as given by Equation (4). The error rate is the ratio between the true and false negative results and the total number of observations, as presented in Equation (5). We also determined the specificity and precision of the model. Specificity is the ratio between the true negatives and the sum of false positives and true negatives, as stated in Equation (6). Precision is the ratio between the true positives and the sum of true positives and false positives, as in Equation (7).

$$accuracy = \frac{TruePositive + TrueNegative}{Total}$$
(4)

$$errorrate = \frac{FalsePositive + FalseNegative}{Total}$$
(5)

$$Specificity = \frac{TrueNegative}{FalsePositive + TrueNegative}$$
(6)

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$
(7)

The accuracy of the proposed methodology was 63%, while the error rate was 37%. The accuracy was slightly below that of Dominguez et al. [33], who applied a machinelearning model and reached 74% accuracy. Viegas et al. [34] used a fuzzy-based model and achieved similar accuracy (63.6%). Some studies reported precision results above 90%, such as Salman Saeed et al. [35] who used decision tree, Messinis et al. [26] who applied support vector machines, and Saeed et al. [36] who used ensemble bagged tree models. It may seem that the proposed methodology was inferior when compared to those last studies. However, for high accuracies in which the dataset is unbalanced, this measure is not suitable. For example, if a dataset has 90% of negative cases and the model classifies all of them as negative, an accuracy of 90% will be reached. In addition, the aforementioned studies did not focus on rural areas, accounting for several consumers due to the customer density in urban grids. In this work, for the pilot study area, the number of consumer units is quite small compared to the other studies. In this sense, specificity and precision measurements are used as an additional analysis. For both indicators, the proposed methodology reached 100%. Salman Saeed et al. [35] reached 98.2% and 93.2% for specificity and precision, respectively. Saeed et al. [36] reached 98.2% for specificity, and Messinis et al. [26] did not present such information.

The power distribution company currently uses correlation analysis and direct selection to identify the most significant variables in the meteorological and cultural data and energy consumption records. The sampling and data collection are completely manual, and the detection of NTLs has an accuracy of 57%. This work proposed a methodology that uses automation to obtain relevant data from the database, makes a total analysis of the areas of interest, and presents greater precision. Thus, the fuzzy logic algorithm can bring greater flexibility and speed for application in different areas of interest.

For the pilot study area, the results were promising and proved the effectiveness of the proposed methodology. In addition, this work is restricted to one culture and one area due to the difficulty of accessing commercial data from the power utility company. However, we could evaluate the effectiveness of the proposed methodology.

We can observe that the error rate is linked only to false positives, i.e., the detection of suspicious consumer units of NTLs has proved to be energy-inefficient consumer units. This is due to the selection criterion for the suspicious consumer units of NTLs since this classification encompasses both the inefficiency of the consumer unit and the presence of NTLs.

By changing the criteria to consider only high and very high suspicious consumer units, there is a high chance that consumer units with NTLs would not be included in the list, which would be disastrous for a system seeking this fact. However, even considering a more adjusted suspicion criterion, we can guarantee that unsuspicious consumer units of NTLs do not have this characteristic. Thus, this criterion can be understood as the most interesting one. Some adjustments should be made to the modeling to include other factors that are not currently accessible, such as a long-term history for each consumer unit.

The results from field inspections classified the inconsistencies as follows: 3.5% as irregular with billing impact and 10.6% as irregular without billing impact. The first category includes fraud, equipment problems, and process errors. In practice, the field inspection teams should register the fraud detected as equipment failures and recovery as part of the loss since the legislation allows for the concessionaire to review the values of the last three months from inspection registration.

4. Conclusions

Power utility companies have faced the challenge of identifying NTLs in rural areas. Effectively, field inspections are difficult in the rural distribution grids in the state of Rio Grande do Sul, in southern Brazil, due to their characteristics.

The pilot study area was effective at reducing the search area for NTLs in consumer units. This work established a multi-parametric and multi-criteria model that allowed us to classify consumer units based on the suspicion of NTLs.

In this pilot study, we selected rice crop irrigation to analyze energy consumption. Our proposed methodology achieved promising results and could be robust enough to establish standards for other crops, such as sugarcane and coffee, achieving the same accuracy as rice.

This work presents a comprehensive methodology to establish the criteria for suspicious consumer units of NTLs in irrigation systems. The contribution is significant since on-site inspection is costly due to large areas usually located far from urban centers. Reducing the search space for these consumer units helps to reduce inspection costs and increases the effectiveness of actions to prevent NTLs in irrigation systems.

Reducing commercial losses enables power utility companies to improve load forecasts, providing compensation during more critical moments. As a result, the quality of the energy delivered to all consumers served by the rural grids improves.

Author Contributions: Conceptualization, V.G.V. and D.P.B.; methodology, V.G.V., V.A.U. and R.M.d.F.; software, V.G.V. and V.A.U.; validation, V.G.V. and V.A.U.; formal analysis, R.M.d.F. and V.A.U.; investigation, V.G.V., V.A.U. and R.M.d.F.; resources, V.G.V. and J.A.S.; data curation, V.G.V., J.A.S. and R.M.d.F.; writing—original draft preparation, V.G.V., V.A.U. and R.M.d.F.; writing—review and editing, V.G.V., D.P.B., V.A.U., R.M.d.F., L.M.d.C. and J.A.S.; visualization, V.G.V., V.A.U. and R.M.d.F.; supervision, D.P.B.; project administration, D.P.B. and L.M.d.C.; and funding acquisition, D.P.B., J.A.S. and L.M.d.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by CPFL Energia through the project "Sistema de Detecção de Perdas não Técnicas em Áreas de Irrigação Empregando Técnicas de Inteligência Artificial" (developed under the ANEEL R&D Program PD-00063-3065/2020) and partially by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES/PROEX)—Finance Code 001.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful for the technical and financial support provided by CPFL Energia (Programa P&D/ANEEL) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES/PROEX—Finance Code 001) for the opportunity to develop this project.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

- NTLs Non-technical losses
- NDVI Normalized difference vegetation index
- PDCs Power distribution companies
- RTU Remote terminal units
- SVM Support vector machine
- ANN Artificial neural network
- OPF Optimum-path forest
- DT Decision trees
- K-NN Nearest neighbor

References

- 1. Glauner, P.; Meira, J.A.; Valtchev, P.; State, R.; Bettinger, F. The Challenge of Non-Technical Loss Detection Using Artificial Intelligence: A Survey. *Int. J. Comput. Intell. Syst.* **2017**, *10*, 760–775. [CrossRef]
- Shah, A.; Mesbah, W.; Al-Awami, A.T. An algorithm for detaching technical losses from non-technical losses in distribution systems. In Proceedings of the 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 16–18 February 2021; pp. 1–5. [CrossRef]
- 3. ANA. Atlas Irrigação: Uso da Água na Agricultura Irrigada; Technical Report; Agência Nacional de Águas: Brasília, Brazil, 2017.
- 4. SOSBAI. Arroz Irrigado: Recomendações Técnicas da Pesquisa para o Sul do Brasil; Technical Report; Sociedade Sul-Brasileira de Arroz Irrigado: Cachoeirinha, Brazil, 2018.
- ANEEL. Perdas de Energia eléTrica na Distribuição; Technical Report; Agência Nacional de Energia Elétrica: Brasília, Brazil, 2022. Available online: https://git.aneel.gov.br/publico/centralconteudo/-/raw/main/relatorioseindicadores/tarifaeconomico/ Relatorio_Perdas_Energia.pdf (accessed on 5 October 2022).
- Viegas, J.L.; Esteves, P.R.; Melício, R.; Mendes, V.M.F.; Vieira, S.M. Solutions for detection of non-technical losses in the electricity grid: A review. *Renew. Sustain. Energy Rev.* 2017, 80, 1256–1268. [CrossRef]
- Duarte Junior, A.J.; Streck, N.A.; Zanon, A.J.; Ribas, G.G.; da Silva, M.R.; Cera, J.C.; do Nascimento, M.d.F.; Pilecco, I.B.; Puntel, S. Rice yield potential as a function of sowing date in southern Brazil. *Agron. J.* 2021, *113*, 1523–1534. [CrossRef]
- 8. ANA; CONAB. *Mapeamento do Arroz Irrigado no Brasil*; Technical Report; Agência Nacional de Águas e Saneamento Básico, Companhia Nacional de Abastecimento: Brasília, Brazil, 2020.
- 9. ANA. *Atlas Irrigação: Uso da ÁGua na Agricultura Irrigada;* (Translation) Irrigation Atlas: Water Use in Irrigated Agriculture; Technical Report; Agência Nacional de Águas e Saneamento Básico: Brasília, Brazil, 2021.
- Instituto Rio Grandense do Arroz. Custo de Produção Médio Ponderado—Arroz Irrigado Rio Grande do Sul—Safra 2021/22; Technical Report; Instituto Rio Grandense do Arroz, Departamento Comercial e Industrial—Seção de Política Setorial: Porto Alegre, Brazil, 2022. Available online: https://admin.irga.rs.gov.br/upload/arquivos/202207/19145145-custo-de-producao-2021-22.pdf (accessed on 2 July 2023).
- 11. Köpp, L.M.; Peiter, M.X.; Robaina, A.D.; Toescher, C.F. Characterization of Pumping Stations in Irrigated Rice Crops from Rio Grande Do Sul, Brazil. *Eng. Agríc.* **2016**, *36*, 342–351. [CrossRef]
- 12. Köpp, L.M.; Peiter, M.X.; Robaina, A.D.; Girardi, L.B. Performance indices for pumping stations in irrigated rice fields. *Ciênc. Rural* **2016**, *46*, 1384–1389. [CrossRef]
- Omar, S.J.; Fred, K.; Swaib, K.K. A state-of-the-art review of machine learning techniques for fraud detection research. In Proceedings of the 2018 International Conference on Software Engineering in Africa (SEiA), Gothenburg, Sweden, 27 May–3 June 2018; pp. 11–19.
- 14. Park, C.H.; Kim, T. Energy Theft Detection in Advanced Metering Infrastructure Based on Anomaly Pattern Detection. *Energies* **2020**, *13*, 3832. [CrossRef]
- 15. Buzau, M.M.; Tejedor-Aguilera, J.; Cruz-Romero, P.; Gómez-Expósito, A. Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning. *IEEE Trans. Smart Grid* 2019, 10, 2661–2670. [CrossRef]
- 16. De Oliveira Ventura, L.; Melo, J.D.; Padilha-Feltrin, A.; Fernández-Gutiérrez, J.P.; Sánchez Zuleta, C.C.; Piedrahita Escobar, C.C. A new way for comparing solutions to non-technical electricity losses in South America. *Util. Policy* **2020**, *67*, 101113. [CrossRef]
- Barja-Martinez, S.; Aragüés-Peñalba, M.; Munné-Collado, I.; Lloret-Gallego, P.; Bullich-Massagué, E.; Villafafila-Robles, R. Artificial intelligence techniques for enabling Big Data services in distribution networks: A review. *Renew. Sustain. Energy Rev.* 2021, 150, 111459. [CrossRef]
- 18. Savian, F.d.S.; Siluk, J.C.M.; Garlet, T.B.; do Nascimento, F.M.; Pinheiro, J.R.; Vale, Z. Non-technical losses: A systematic contemporary article review. *Renew. Sustain. Energy Rev.* 2021, 147, 111205. [CrossRef]
- 19. Saeed, M.S.; Mustafa, M.W.; Hamadneh, N.N.; Alshammari, N.A.; Sheikh, U.U.; Jumani, T.A.; Khalid, S.B.A.; Khan, I. Detection of Non-Technical Losses in Power Utilities—A Comprehensive Systematic Review. *Energies* **2020**, *13*, 4727. [CrossRef]
- Teles Faria, L.; Melo, J.D.; Padilha-Feltrin, A. Spatial-Temporal Estimation for Nontechnical Losses. *IEEE Trans. Power Deliv.* 2016, 31, 362–369. [CrossRef]
- Xia, X.; Xiao, Y.; Liang, W. SAI: A Suspicion Assessment-Based Inspection Algorithm to Detect Malicious Users in Smart Grid. IEEE Trans. Inf. Forensics Secur. 2020, 15, 361–374. [CrossRef]
- 22. Cui, L.; Qu, Y.; Gao, L.; Xie, G.; Yu, S. Detecting false data attacks using machine learning techniques in smart grid: A survey. *J. Netw. Comput. Appl.* **2020**, *170*, 102808. [CrossRef]
- 23. Khan, Z.A.; Adil, M.; Javaid, N.; Saqib, M.N.; Shafiq, M.; Choi, J.G. Electricity Theft Detection Using Supervised Learning Techniques on Smart Meter Data. *Sustainability* **2020**, *12*, 8023. [CrossRef]
- 24. Feng, Z.; Huang, J.; Tang, W.; Shahidehpour, M. Data mining for abnormal power consumption pattern detection based on local matrix reconstruction. *Int. J. Electr. Power Energy Syst.* **2020**, 123, 106315. [CrossRef]
- Messinis, G.M.; Hatziargyriou, N.D. Review of non-technical loss detection methods. *Electr. Power Syst. Res.* 2018, 158, 250–266. [CrossRef]
- 26. Messinis, G.M.; Rigas, A.E.; Hatziargyriou, N.D. A hybrid method for non-technical loss detection in smart distribution grids. *IEEE Trans. Smart Grid* **2019**, *10*, 6080–6091. [CrossRef]

- 27. Otuoze, A.O.; Mustafa, M.W.; Mohammed, O.O.; Saeed, M.S.; Surajudeen-Bakinde, N.T.; Salisu, S. Electricity theft detection by sources of threats for smart city planning. *IET Smart Cities* 2019, *1*, 52–60. [CrossRef]
- Ahmad, T.; Chen, H.; Wang, J.; Guo, Y. Review of various modeling techniques for the detection of electricity theft in smart grid environment. *Renew. Sustain. Energy Rev.* 2018, 82, 2916–2933. [CrossRef]
- 29. INMET. *Banco de Dados Meteorológicos do INMET*; Technical Report; Instituto Nacional de Meteorologia: Brasília, Brazil, 2022. Available online: https://bdmepinmet.gov.br/ (accessed on 15 March 2022).
- Vieira, V.; Biazzi, R.; Bernardon, D.P.; de Sousa, N.B.; Eichkoff, H.S.; Pereira, P.R.S.; Lemes, D.L.; de Figueiredo, R.M.; Jacques, M.M.; Barriquello, C.H.; et al. Uma Metodologia Baseada em Inteligência Artificial para Detecção de Perdas Não Técnicas em Irrigantes. In Proceedings of the XXIV Congresso Brasileiro de Automática (CBA), Fortaleza, Brazil, 16–19 October 2022. Available online: https://www.sba.org.br/cba2022/wp-content/uploads/artigos_cba2022/paper_5094.pdf (accessed on 21 March 2022).
- Biazzi, R.R.; Bernardon, D.P.; Becker, E.; de Chiara, L.M.; Silva, J.A. Technical and Nontechnical Energy Loss Estimation Including Volt/Var Control for Active Distribution Systems. J. Control Autom. Electr. Syst. 2022, 33, 255–267. [CrossRef]
- INMET. Normais Climatológicas do Brasil 1991–2020; Technical Report; Instituto Nacional de Meteorologia: Brasília, Brazil, 2022. Available online: https://portal.inmet.gov.br/normais (accessed on 12 March 2023).
- Dominguez, J.; Rivera, A.; Botina, K.; Perdomo, G.; Montoya, O.; Campillo, J.; Delahoz, E. Data-driven framework for the detection of non-technical losses in distribution grids. In Proceedings of the 2020 IX International Congress of Mechatronics Engineering and Automation (CIIMA), Cartagena de Indias, Colombia, 4–6 November 2020; pp. 1–6. [CrossRef]
- Viegas, J.L.; Esteves, P.R.; Vieira, S.M. Clustering-based novelty detection for identification of non-technical losses. Int. J. Electr. Power Energy Syst. 2018, 101, 301–310. [CrossRef]
- 35. Salman Saeed, M.; Mustafa, M.W.; Sheikh, U.U.; Jumani, T.A.; Khan, I.; Atawneh, S.; Hamadneh, N.N. An efficient boosted C5.0 decision-tree-based classification approach for detecting non-technical losses in power utilities. *Energies* **2020**, *13*, 3242. [CrossRef]
- Saeed, M.S.; Mustafa, M.W.; Sheikh, U.U.; Jumani, T.A.; Mirjat, N.H. Ensemble bagged tree based classification for reducing non-technical losses in multan electric power company of Pakistan. *Electronics* 2019, *8*, 860. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.