Advances in Computational Intelligence (2022) 2:30
https://doi.org/10.1007/s43674-022-00042-y

ORIGINAL ARTICLE

Detection of cyber attacks on smart grids

Aditi Kar Gangopadhyay[1] · Tanay Sheth[2] · Tanmoy Kanti Das[3] · Sneha Chauhan[4,5]

Received: 18 November 2021 / Revised: 1 August 2022 / Accepted: 14 August 2022 / Published online: 31 August 2022

Abstract

The paper analyzes observations using a logic-based numerical methodology in Python. The Logical Analysis of Data (LAD) specializes in selecting a minimal number of features and finding unique patterns within it to distinguish 'positive' from 'negative' observations. The Python implementation of the classification model is further improved by introducing adaptations to pattern generation techniques. Finally, a case study of the Power Attack Systems Dataset used to improvise Smart Grid technology is performed to explore real-life applications of the classification model and analyze its performance against commonly used techniques.

Keywords Logical analysis of numerical data (LAD) · Power attack systems · ML · Python · Classification

1 Introduction

Smart grids are electrical grids that use information and communication technology (ICT) to provide efficient, reliable distribution and transmission. The importance of security and trust cannot be overstated. Among the several developing security vulnerabilities, the fake data injection (FDI) attack is one of the most serious, with the potential to increase energy distribution costs drastically (Ahmed and Pathan 2020). The reliable operation of any power system depends mainly on appropriate protection schemes developed for line faults and emergencies. The reliable protection scheme enables faster fault detection to restore the power supply as soon as possible after a failure. In recent years, with the dazzling assimilation of the physical energy transmission system in the smart grid with the cybernetic information and communication tools, the possibility of cyber-attacks poses a severe challenge to the development and implementation of the reliable protection mechanism. Fail protection components play an essential role in the overall operation and control of the power system. Increased pressure on rapid fault detection and a reduction in fault levels are emerging because the penetration of renewable energy has caused a shift from a classic protection scheme using local measures to a "wide-area measurement-based protection scheme" (Phadke et al. 2008). The protection scheme's effective performance based on the wide-area measurement is highly dependent on the information from the sensor transmitted to the control center over the network. The power system is overly dependent on the public communication network for reliable monitoring and operation, making it vulnerable to network attacks (Sridhar et al. 2011). False Data Injection Attack (FDIA) is considered the most effective network attack, in which the hacker can block the entire power grid with minimal effort. During the FDIA, the attacker destroys the integrity of a set of measurements used in the protection algorithm by altering

A. K. Gangopadhyay, T. Sheth, T. Kanti and S. Chauhan have contributed equally to this work.

✉ Aditi Kar Gangopadhyay
aditi.gangopadhyay@ma.iitr.ac.in

Tanay Sheth
tanay.v.sheth@gmail.com

Tanmoy Kanti Das
tkdas.mca@nitrr.ac.in

Sneha Chauhan
schauhan1@cs.iitr.ac.in

[1] Department of Mathematics, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India

[2] BITS Pilani Goa Campus, Goa 403726, Goa, India

[3] Department of Computer Application, National Institute of Technology Raipur, Raipur 492010, Chhattisgarh, India

[4] Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India

[5] Department of Computer Science and Engineering, National Institute of Technology Uttarakhand, Srinagar 246174, Uttarakhand, India

the meter/sensor measurements (Liang et al. 2016; Liu and Li 2017). The protection algorithm is part of the backup protection strategy, and the control center operates it. Transmission of erroneous data to the control center can cause unnecessary control actions, leading to unexpected events or even power outages. Therefore, the current scenario requires a protection scheme that is immune to data falsification or/and includes components for the preventive detention of the injection of false data. Conventional flawed data detection methods that are part of the state estimator should detect any malicious manipulation of sensor information. However, Liu et al. (2011) have proven that hackers with sufficient knowledge of system dynamics can bypass the lousy data detection technology and inject random errors into state variables using FDIA to inject information from malicious sensors. Thus, manipulation of the sensor information during an attack can provide a misleading picture of the dynamics and operation of the system, causing the relay to malfunction during the fault or causing the relay to trip and then isolate itself. The malfunction of the protection relay and the delay in detecting this type of attack can cause enormous economic losses, damage to assets, and the collapse of the subsystems and control mechanisms related to power systems.

Several reasons have contributed to a significant increase in installation error-based data injection seizures. Some of the factors are continuous real-time online monitoring using sensors (CT, PT, PMU) and communication networks, using signal information of different locations or bus current or voltage. The recent work of FDIA in the power grid mainly focuses on FDIA modeling, attack detection, and defense measures (Liu et al. 2016). The possible impact of FDIA on the power system has been addressed in (Liang et al. 2016; Liu et al. 2011; Deng et al. 2016). Noteworthy solutions for FDIA detection reports in power grids are based on transmission line susceptance measurement (Deng and Liang 2018), reactance disturbance (Liu et al. 2018), joint transformation (Singh et al. 2017), extreme machine learning (Yang et al. 2017), optimized dispersion (Liu et al. 2014) and cumulative sum method (Li et al. 2014). Yang et al. (2013) proposed a countermeasure against FDIA, provided that the sensor measures the injected energy flow in the bus and connects to several other buses, which requires safety. The inaccessibility of these sensors will make it difficult for attackers to install FDIA. In Bi and Zhang (2014), Deng et al. (2015), a defense mechanism is proposed to protect a set of state variables.

Das et al. (2019) have proposed a logical analysis of numerical data (LAD) scheme of a simple, economically viable, and FDIA resilient for the attack on power systems, under the assumption that the adversary has complete knowledge of the system dynamics. The rule-based fault detection scheme identifies the limited set of sensors that would be secured using the cryptographic protocol, tamper-resistant hardware, and encryption-based data analysis by mapping the secure sensor information. This paper implements the LAD process with adaptations to simulate and optimize the results. The LAD process uses top-down and bottom-up approaches to produce pure patterns. The adaptations introduced in this produce combinatorial and impure patterns to maximize the performance of the LAD model. The adapted pattern generation generated 38 percent more true positive outcomes than the original model. The model uses a greedy algorithm to optimize the number of attributes based on conclusions drawn by paper (Almuallim and Dietterich 1994) which compares several feature minimization techniques and evaluates each one on the worst-case scenarios, time complexity, and average accuracy. The implementation analyzes power system FDIA scenarios - Mississippi State University and Oak Ridge to identify attack scenarios. Performances of various classifiers F1 Score of classification on the same dataset are discussed in the paper (Liu and Li 2017) and visualized below. In this paper, the proposed architecture yields an F1 Score of 0.86, outperforming the traditional classifiers.

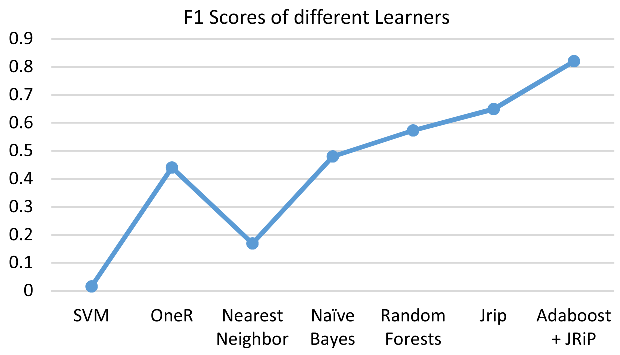


Fig. 1 F1 Scores of different Learners

Lastly, we explain the motivation behind employing Logical Analysis of Data (LAD) in smart grids. In data analytics, machine learning is widely used. According to academics and industry, practically all machine learning systems function statistically or optimize blindly. The results' causal logic remains a black box, restricting machine learning's usefulness. Our LAD research model employs machine learning to increase application accuracy and rules to ensure the interpretability of results, starting with a structured approach to causal reasoning. Simultaneously, we propose a new rule system based on a mathematically adapted model. It can handle potentially enormous datasets with limited computations.

2 LAD model structure and terminologies

The logical analysis of numerical data (LAD) is a combinatory and optimization-based data analysis method. In Boros et al. (1997) the authors develop the theoretical foundation of the binarization process. They also study the combinatorial optimization problems related to minimizing the number of binary variables. The paper establishes nineteen theorems and several lemmas along with proofs. It shows that any numerical dataset can be checked in polynomial time whether there is a binarization admitting an extension in the given class. A linear integer programming problem is formulated to provide an algorithmic framework for this minimization problem. Set covering problems and some heuristic algorithms are implemented and elaborated further for improvements. LAD detects structural information about datasets which can provide powerful means to solve various problems. Mainly it contributes to classification, automatic knowledge acquisition of expert systems, model-based decision support system development, database inconsistency detection, and feature selection.

The LAD methodology was first proposed for the case of binary data. LAD has applications in numerous disciplines, such as economics and business, seismology, oil exploration, and a few typical examples of binary classification problems. The papers (Boros et al. 2000) and (Hammer and Bonates 2006) are exemplary LAD use cases. It describes the implementation and wide applicability of LAD to Australian Credit Card, Boston Housing, Breast Cancer (Wisconsin), Congressional Voting datasets, and even pilot experiments such as Oil Exploration, Psychometric Testing and Labor Productivity in China, and an in-depth application of LAD as in the prognosis and diagnosis field of Medical Data Analysis with case studies of Ovarian Cancer Diagnosis using a Large Proteomic Dataset, Genome Data-based Breast Cancer Prognosis, respectively. These applications depict the robustness of LAD in any scenario. The dissertation paper (Bonates 2007) shows efficient ways of constructing LAD classification models having high accuracy and requiring minimal control parameters. It also extended the LAD methodology to deal with the critical class of regression problems that frequently appear in data analysis tasks. In this paper, the implementation in Python behaves as an ML classification framework and can adapt to changes with ease. The further sections describe the architecture of code, its adaptations, results, and the conclusions drawn.

2.1 Mathematical background

The essential mathematical foundation components of LAD (Alexe et al. 2007) are the following:

- To remove superfluous variables from the original dataset, we select a (usually minimal) subset S that can discriminate positive from negative observations. We work with the projections $+S$ and $-S$ on this group of variables in the following steps. While most data analysis methods include a “feature extraction” step, the LAD methodology uses it differently. Here, it emphasizes the interaction of variables and the importance of retaining those that can influence the positive or negative nature of observations individually and those whose “collective” or “combinatorial” effect is significant.
- We cover $+S$ with a family of (potentially overlapping) homogeneous subsets of the reduced real space, where each subset intersects $+S$ but is disjoint to $-S$. LAD only considers $\mathbb{R}^{|S|}$ intervals with faces parallel to the axes; these intervals are referred to as “positive patterns”. For finding “negative patterns”, a similar construction is used with −S.
- A subset of positive (respectively, negative) patterns is discovered whose union encompasses all the observations in +S (respectively, −S). A “model” is a collection of these two subsets of intervals.
- A classification approach defines each observation’s positive or negative character covered by the union of the two subsets of intervals of the model, leaving those observations uncovered by this union as “unclassified”.
- The resulting classification system’s correctness is verified using one of the standard validation methods.

The basic structure of LAD starts with a set of observations S. S consists of observations of two classes, positive and negative, respectively. Hence S is now categorized into $+S$ and $-S$ for the above two classes. Each observation carries an n number of attributes labeled $a_1, a_2, \ldots, a_n$. Each attribute is then analyzed to generate cut points labeled as $t_1, t_2, \ldots, t_n$. These cut-points generate binarized attributes $ba_1, ba_2, \ldots$. The set of all binarized attributes is labeled as V. Then the support set generation takes place as a set of the minimal number of binarized attributes labeled Q. Thereafter, Q is used to produce patterns $p_1, p_2, \ldots$. Finally, a classification model is built using all the generated patterns.

2.2 Binarization

The binarization procedure is as follows. The simplest non-binary attribute is the nominal (or descriptive) attribute. The typical nominal property is color, and its value can be red, green, yellow, etc. The binarization of the attribute is done directly by associating each value v_s of the attribute x against a Boolean variable. In the particular case of nominal attributes, which are binary, i.e., they take only two values, no additional binary variables are introduced. The values are renamed as 0 and 1 Boros et al. (2000).

$$b\left(x, v_s\right) = \begin{cases} 1 & \text{if } x = v_s \\ 0 & \text{otherwise} \end{cases}$$

The binarization of ordered attributes is common in many areas of human activity. For example, blood pressure, body temperature, pulse rate, and other medical parameters are called “normal” or “abnormal,” depending on whether they

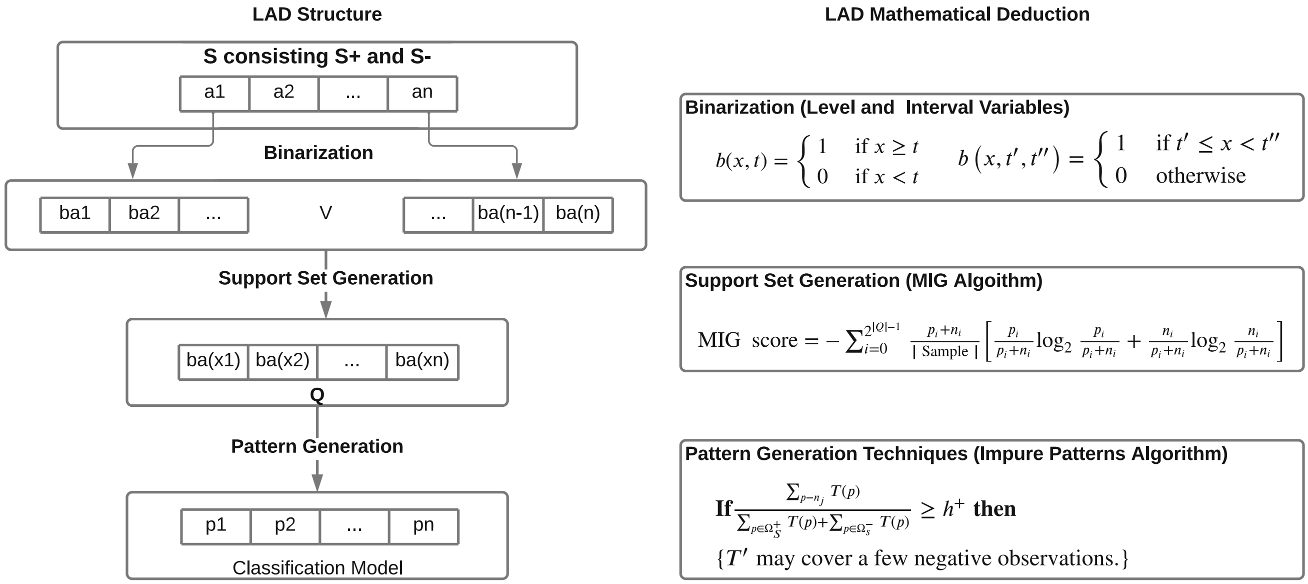

LAD Structure
S consisting S+ and S-
a1
a2
...
an
Binarization
ba1
ba2
...
V
...
ba(n-1)
ba(n)
Support Set Generation
ba(x1)
ba(x2)
...
ba(xn)
Q
Pattern Generation
p1
p2
...
pn
Classification Model
LAD Mathematical Deduction
Binarization (Level and Interval Variables)
$b(x,t) = \begin{cases} 1 & \text{if } x \geq t \\ 0 & \text{if } x < t \end{cases}$
$b\left(x,t',t''\right) = \begin{cases} 1 & \text{if } t' \leq x < t'' \\ 0 & \text{otherwise} \end{cases}$
Support Set Generation (MIG Algoithm)
$\text{MIG score} = -\sum_{i=0}^{2^{|Q|-1}} \frac{p_i+n_i}{|\text{ Sample }|}\left[\frac{p_i}{p_i+n_i}\log_2\frac{p_i}{p_i+n_i} + \frac{n_i}{p_i+n_i}\log_2\frac{n_i}{p_i+n_i}\right]$
Pattern Generation Techniques (Impure Patterns Algorithm)
If $\frac{\sum_{p-n_j} T(p)}{\sum_{p\in\Omega_S^+} T(p)+\sum_{p\in\Omega_S^-} T(p)} \geq h^+$ then
$\{T'$ may cover a few negative observations.$\}$

Fig. 2 Phases of LAD

are within or outside a specific range. In many other examples, the parameter (for example, blood sugar level) is called "normal" or "abnormal" depending on whether it is above or below a certain threshold. In all of these examples, the binarization is done implicitly by comparing the value of a numeric attribute with some standard cut-off point (critical value). Following that, the same principle is applied for binarizing real numerical values. The variables are split into two types: level and interval. Level variables, as the name suggests, create levels, that is the binarization occurs based on the value being above or below every cut-off point(t), labeling it as 1 and 0, respectively.

$$b(x,t) = \begin{cases} 1 & \text{if } x \geq t \\ 0 & \text{if } x \ < t \end{cases}$$

Similarly, binarization for interval variables takes place if the value lies between two cut points.

$$b\left(x,t',t''\right) = \begin{cases} 1 & \text{if } t' \leq x \ < t'' \\ 0 & \text{otherwise} \end{cases}$$

While binarizing numerical attributes, the unique values are sorted in an array to calculate the cut points. In order to make this binarization procedure more robust concerning measurement errors (in the case of numerical attributes), we will use the cut-points as the range between midpoints of consecutive unique points $t_s = \frac{1}{2}\left(v_{s-1} + v_s\right)$.

While dealing with huge datasets, we set up a threshold of critical values generated from a particular attribute to minimize the generation of ambiguous cut points. For example, for an attribute containing 3000 cut points, critical points generated could be limited to 200. If it fails to generate less than 200 points, we try to improvise the attribute by rounding off each point's last digit and calculating until the number of cut points is less than the threshold.

2.3 Support set generation

After obtaining a binary dataset, the elimination of redundant attributes is prioritized. All LAD archives of observation points are partitioned into a set of true $(+S)$ and false $(-S)$ classes, i.e., we assume that no observation point is present in both simultaneously. This property is known as contradiction-free, a basic requirement to be maintained by any correct binarization technique, as clearly preserved by our process. A set of binary attributes is called a support set Q if the archive obtained by eliminating all the other attributes will remain contradiction-free. A support set is called irredundant if no proper subset of it is a support set Boros et al. (2000). The Support Set Generation method used here is Mutual Information Greedy (MIG) Algorithm Almuallim and Dietterich (1994). The paper concludes that the MIG algorithm maintained good average-case performance improving all the learning processes it was implemented on while exhibiting rather bad worst-case performance. The MIG algorithm has entropy, or score calculation function as follows:

$$\text{MIG score} = -\sum_{i=0}^{2^{|Q|-1}} \frac{p_i+n_i}{|\text{ Sample }|}\left[\frac{p_i}{p_i+n_i}\log_2\frac{p_i}{p_i+n_i} + \frac{n_i}{p_i+n_i}\log_2\frac{n_i}{p_i+n_i}\right].$$

In the Mutual-Information-Greedy Algorithm, the feature that leads to the minimum entropy when added to the current partial solution is selected as the best feature. The best feature is used to partition each group of training samples until each group is either solely positive or negative. An example of the execution without splitting the training sample into all 2^Q groups is elaborated below.

Algorithm 1 MIG Algorithm Almuallim and Dietterich (1994)

```
$Q = \phi$, $V = \{x_1, x_2, \ldots, x_n\}$, $\mathcal{S} =$\{ Sample \}
while $\mathcal{S}$ is not empty do
    Best-score $= \infty$
    for $x_i \in V$ do
        Score =0
        for $s \in S$ do
            $p_0 =$ # of positive examples in $s$ with $x_1 = 0$
            $n_0 =$ # of negative examples in $s$ with $x_i = 0$
            $p_1 =$ # of positive examples in $s$ with $x_i = 1$
            $n_1 =$ # of negative examples in $s$ with $x_i = 1$
            $e_0 = \frac{p_0}{p_0+n_0} \log_2 \frac{p_0}{p_0+n_0} + \frac{n_0}{p_0+n_0} \log_2 \frac{n_0}{p_0+n_0}$
            $e_1 = \frac{p_1}{p_1+n_1} \log_2 \frac{p_1}{p_1+n_1} + \frac{n_1}{p_1+n_1} \log_2 \frac{n_1}{p_1+n_1}$
            Score($x_i$) = Score($x_i$) - $\frac{p_0+n_0}{\text{Sample}}$ $[e_0 + e_1]$
        end for
        if Score($x_i$) < Best-score then
            Best-feature = $x_i$
            Best-score = Score
        end if
    end for
    for $s \in \mathcal{S}$ do
        Partition $s$ into $s_0$ and $s_1$, which are the sets of examples with
        Best-Feature =0 and 1, respectively.
        Replace $s$ in $\mathcal{S}$ by $s_0$ and $s_1$. However, if any of $s_0$ and $s_1$ is
        empty or contains only examples of the same class, then it should
        not be added to $\mathcal{S}$.
    end for
    Remove Best-Feature from $V$.
    Add Best-Feature to $Q$.
    Return $Q$.
end while
```

2.4 Pattern recognition

Patterns are combinations of Boolean attributes of specific orientation that help us classify between positive and negative classes. For example, a combination of binary attributes b1 = 1 and b2 = 0 are only found in positive classes and not in negative classes, which becomes a well-defined positive pattern. The symmetrical definition for negative patterns also holds. The simplest pattern generation method is based on the use of the combinatorial enumeration technique. Given that there are various possible quality metrics for any given pattern, it is important that the pattern generation process must not lose any best patterns. Here best patterns symbolize any pattern that classifies many data points of a specific class at once. Any pattern generation technique should follow two basic principles. The simplicity principle is that short patterns are preferred over longer ones. The second principle is about comprehensive patterns, i.e., all observations of a particular class are to be classified by one of the patterns. We followed a bottom-up approach up to third-degree positive patterns for our model as higher degree pattern generation was not computationally feasible. The pattern generation process is explained through the Algorithm 2 (Das et al. 2020).

Algorithm 2 Pattern Recognition Algorithm

```
Input:   $\Omega_s^+, \Omega_s^- \subset \{0, 1\}^\pi$, – Sets of positive and negative obser-
    vations.
$\bar{d}$ - Maximum degree of generated patterns.
$k^+$- Minimum number of positive observations covered by a pattern.
$h^+$- Required homogeneity of a pattern.
Output: P    - Set of prime patterns.
$P = q$.
$C_0 = \{n\}$.
for $d = 1, \ldots, d$ do
    if $d < d$ then
        $C_d = \emptyset. \{C_d$ is not required.$\}$
    end if
    for $T \in C_{d-1}$ do
        $p =$ maximum index of the literal in $T$.
        for $s = p + 1, \ldots, n$ do
            for lnew $\in \left\{l_s, \bar{l}_s\right\}$ do
                $T' = T \| l_{\text{aew}}$ .
                for $i = 1$ to $d - 1$ do
                    $T'' =$ remove $t^{\text{th}}$ literal from $T'$.
                    if $T^n \& C_{\text{d}-1}$ then
                        break
                    end if
                end for
                if $k^+ \leq \sum_{y \in \Omega_\xi} T'(y)$ then
                    $\{T'$ covers at least $k$ many positive observations.$\}$
                    if $\frac{\sum_{p-n_j} T(p)}{\sum_{p \in \Omega_S^+} T(p) + \sum_{p \in \Omega_s^-} T(p)} \geq h^+$ then
                        $\{T'$ may cover a few negative observations.$\}$
                        $\text{P} = \text{P} \cup \left\{T'\right\}$.
                        Remove the points (or observations) covered by
    T$'$ from $\Omega_3^t$.
                    end if
                else if $d < \bar{d}$ then
                    $C_d = C_d \cup \{T'\}$
                end if
            end for
        end for
    end for
end for
```

While generating patterns with limited computational resources, we were able to build classifiers that were highly accurate but not comprehensive enough. We introduce two adaptations in our model inspired from Das et al. (2020). First, we introduce imperfect patterns in our model. Imperfect patterns are those patterns that have incorrect classified observations but are below a certain threshold. We set that threshold as 10 percent for our model, which helped us overcome the set covering problem extensively. The second

Table 1 Sample data

	a_1	a_2	a_3
$+S$	Square	12	True
	Circle	29	False
	Triangle	6	True
	Triangle	22	True
$-S$	circle	12	False
	Square	33	True
	Square	1	False

Table 2 Binarized form of sample data

b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
1	0	0	1	0	1	0	0	0	1
0	1	0	1	0	0	0	1	0	0
0	0	1	1	1	0	0	0	0	1
0	0	1	1	0	0	1	0	0	1
0	1	0	1	0	1	0	0	0	0
1	0	0	1	0	0	0	0	1	1
1	0	0	0	0	0	0	0	0	0

adaptation introduced in the patterns was combined patterns. We generated a hybrid fifth-degree pattern as a combination of a third-degree pattern and not a second-degree pattern. For example, (b1 = 1 and b2 = 1 and b3 = 1) and not (b4 = 1 and b5 = 1). This helps cover more observations while maintaining the accuracy and simultaneously avoiding the time complexity of generating 5-degree patterns.

3 LAD case study with a sample dataset

The steps followed to carry out the Logical Analysis of Data on the given sample dataset are as follows:

(a) Consider the dataset given in Table 1. We observe that the attributes a1, a2 and a3 are nominal, numerical and binary in nature, respectively.

(b) The first attribute denoting shapes, which is nominal in nature, can be converted into 3 binarized attributes. Each unique shape becomes a binary variable.

b_1	b_2	b_3
$a_1 = square$	$a_1 = circle$	$a_1 = triangle$

(c) The second attribute, is numerical in nature and hence cut points are to be calculated. The cut points formed are [3.5,9,17,25.5,31]. The binary variables formed are shown in Table 2.

b_4	b_5	b_6
$3.5 \leq a_2$	$3.5 \leq a_2 \leq 9$	$9 \leq a_2 \leq 17$

b_7	b_8	b_9
$17 \leq a_2 \leq 25.5$	$25.5 \leq a_2 \leq 31$	$31 \leq a_2$

(d) The last attribute has Boolean values which is itself binary in nature.

b_{10}
$a_3 = True$

(e) Putting it all together

The support set generation technique later minimizes the binarized dataset as it may contain redundant attributes. Patterns are recognized from the dataset obtained after getting the support set, and the classifier is modeled. For example, in the above example, we can directly observe one such pattern is b3= 1 and b4 =1 that is unique only to the +S set.

4 Cyber physical system survey

Cyber-physical systems (CPS) refer to a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities. The ability to interact with and expand the capabilities of the physical world through computation, communication, and control is a key enabler for future technology developments. A complete summary of anomaly detection strategies is provided by Chandola et al. (2009). They did not include deep learning-based approaches for CPS in an early effort to review anomaly detection methods. People's lives have been revolutionized by commodity IoT solutions. Smart home applications, for example, allow users to interact with house appliances automatically. Methods for analyzing programs to safeguard privacy and find vulnerabilities in these applications have been proposed in Celik et al. (2019). Meanwhile, Giraldo et al. (2018) looked into anomaly detection approaches based on CPS physical features (for example, the evolution of the physical system under control). The findings of studies on SCADA system network security are described, with a focus on risk assessment approaches in Cherdantseva et al. (2016). A review of anomaly detection methodologies in CPS was published by Mitchell and Chen (2014), Nazir et al. (2017), and Zacchia Lun et al. (2018). However, the approaches used do not incorporate deep learning methods and are more traditional, such as state estimation and intrusion detection. A study of deep learning-based anomaly detection systems was conducted in Chalapathy and Chawla (2019) apart from traditional CPS systems.

5 Power system attack case study using LAD model

The dataset used for the model is Power System Attack Datasets provided by Mississippi State University and Oak Ridge National Laboratory. The Natural and Attack States of the power systems are considered positive and negative observations. This dataset is used because the power system disturbances are complex in nature and can be attributed to a wide range of sources, including man-made and natural events. Currently, power system operators are heavily dependent on making decisions about the appropriate course of action for the cause and response of the interference experienced. In the case of cyber attacks on the power system, human judgment is less certain since there is an overt attempt to disguise the attack and deceive the operators as to the true state of the system. To enable the human decision-maker, we explore the viability of the LAD Model as a means for discriminating types of power system disturbances and focus specifically on detecting cyber-attacks where deception is a core tenet of the event. The five types of scenarios covered in the datasets are Short-circuit fault, Line maintenance, Remote tripping command injection (Attack), Relay setting change (Attack), and Data Injection (Attack). The scenarios are explained below (Borges et al. 2014).

Short circuit fault: It is a short that can occur at any point in the power line and the percentage range indicates the location.

Line maintenance: Remote relay trip instruction is given to open one or more breakers.

Remote tripping command injection (attack): It is an attack when the attacker sends a false command to relay to open the breaker.

Relay setting change (Attack): The attacker changes the relay configuration to prevent it from tripping when an actual fault occurs.

Data Injection (Attack): A genuine fault is imitated to induce a blackout by changing parameters like current, voltage, etc.

The data was drawn from 15 datasets containing 128 features and thousands of samples. Across the classification systems, an average of 3,711 attack instances and 1,221 normal instances were included in each file for the analysis.

5.1 Data pre-processing

All the columns with more than sixty percent missing values were eliminated, and then the rows with missing values were filtered out. The final clean dataset consisted of 31514 samples. 80-20 (Train-Test) random split was performed on each dataset using the sklearn library.

5.2 Classification metrics

For each classified datasets, a confusion matrix is calculated consisting true and false, negatives and positives respectively. Four classification metrics, Accuracy, Precision, Recall, and F1 Score are calculated as follows.

$$\begin{aligned} \text{Accuracy} &= (\text{TP} + \text{TN}\ ) \,/\, (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \\ \text{Precision} &= \text{TP}/\ (\text{TP} + \text{FP}) \\ \text{Recall} &= \text{TP}/\ (\text{TP} + \text{FN}) \\ \text{F1} &= (2 * \text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}) \end{aligned}$$

Here TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively (Hossin and Sulaiman 2015).

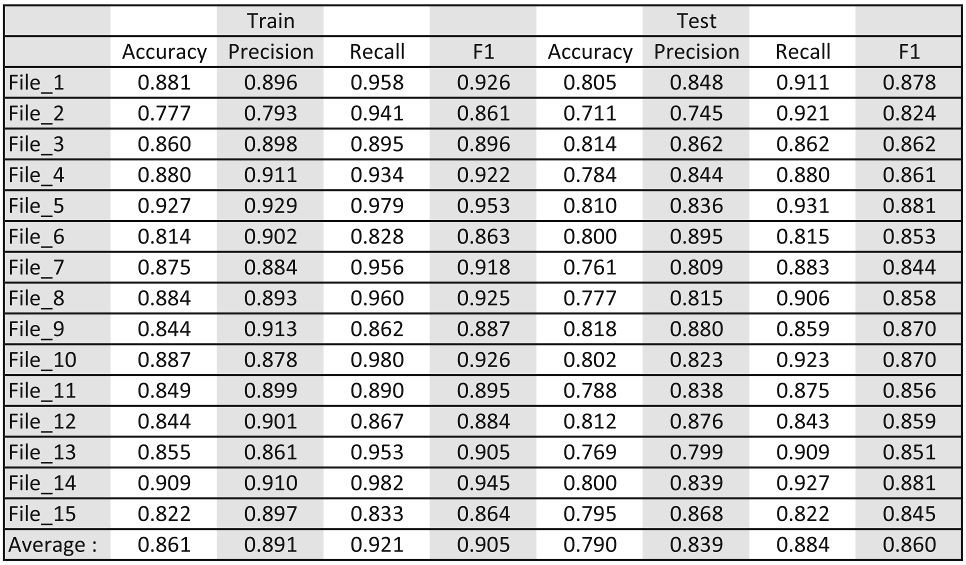
6 Results

Test and train results of detection of cyberattacks on smart grids using Power Attack Systems dataset and LAD model have been tabulated in Figure 3. The dataset is distributed in 15 files. Each dataset is split into train and test datasets by 80–20. Then LAD model is applied to record the performance of the train and test dataset in terms of accuracy, precision, recall, and F1 score. The average accuracy, precision, recall, and F1 score for the test datasets are 79%, 83.9%, 88.4%, and 86%, respectively. In the graphically represented Test dataset results of LAD (Figure 4), the blue, orange, grey, and yellow trend lines represent the accuracy, precision, recall, and F1 score, respectively. It portrays the consistent performance of the LAD model with all subparts of the dataset.

The performance of the LAD and the most widely used machine learning and deep learning models on the dataset has been compared in Table 3 based on the F1 score. It also compares LAD results with previous studies against the same dataset as given in Hink et al. (2014). It is observed that LAD outperforms the state-of-the-art classification techniques as it has obtained an 86% score which is higher than the rest of the techniques. Apart from overall performance, the in-depth analysis and impact of introducing adaptions are visualized in Table 4 by comparing the performance between LAD and the adapted LAD model on a random subpart of the dataset. The adapted LAD model contains impure and hybrid patterns discussed in the Pattern Recognition section of the paper.

It can be deduced from the confusion matrix given in Table 4 that the standard LAD model provides very high precision, whereas the adaptations allow more accuracy and recall and overall create a positive impact for classification.

Fig. 3 LAD train and test results data



	Train				Test			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
File_1	0.881	0.896	0.958	0.926	0.805	0.848	0.911	0.878
File_2	0.777	0.793	0.941	0.861	0.711	0.745	0.921	0.824
File_3	0.860	0.898	0.895	0.896	0.814	0.862	0.862	0.862
File_4	0.880	0.911	0.934	0.922	0.784	0.844	0.880	0.861
File_5	0.927	0.929	0.979	0.953	0.810	0.836	0.931	0.881
File_6	0.814	0.902	0.828	0.863	0.800	0.895	0.815	0.853
File_7	0.875	0.884	0.956	0.918	0.761	0.809	0.883	0.844
File_8	0.884	0.893	0.960	0.925	0.777	0.815	0.906	0.858
File_9	0.844	0.913	0.862	0.887	0.818	0.880	0.859	0.870
File_10	0.887	0.878	0.980	0.926	0.802	0.823	0.923	0.870
File_11	0.849	0.899	0.890	0.895	0.788	0.838	0.875	0.856
File_12	0.844	0.901	0.867	0.884	0.812	0.876	0.843	0.859
File_13	0.855	0.861	0.953	0.905	0.769	0.799	0.909	0.851
File_14	0.909	0.910	0.982	0.945	0.800	0.839	0.927	0.881
File_15	0.822	0.897	0.833	0.864	0.795	0.868	0.822	0.845
Average :	0.861	0.891	0.921	0.905	0.790	0.839	0.884	0.860

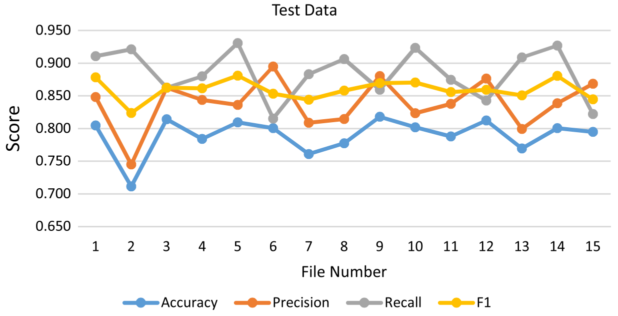


Fig. 4 LAD test results

7 Gap analysis and contribution

The LAD model is compared with other machine learning models based on the F1 score. We can observe from Table 3 that our model has achieved a score of 86%, which is the highest among all the classifiers mentioned. Thus, we can say that the LAD model outperforms most state-of-the-art classification techniques. We have also introduced adaptations in our LAD model, and these adaptations improvise the results even further. It can be observed from Table 4 that the false negatives have reduced for both train and test data. Using the adapted LAD model, we achieve a recall of 97.53% which is very high compared to the standard LAD model. However, more importantly, the LAD model introduces explainability within the classifier while generating results. It gives the knowledge of features involved in the attack, and thus we can focus more on those features which are vulnerable to attack. These all can potentially take place in real-time and with minimal computation. The dataset required to produce the desired results contains only 31514 observations. Thus, LAD does not require a huge dataset for classification. These observations distinguish LAD from the most commonly used machine learning algorithms.

Table 3 Performance comparison of LAD model with other machine learning classifiers based on F1 Score

Model	F1
Support vector machine	0.2
Multi layer perceptron neural network	0.66
Random forest	0.74
Logistic regression	0.75
Decision trees	0.78
LAD model discussed in paper	0.86
Other models referenced in Hink et al. (2014)	
OneR	0.44
Nearest neighbor	0.169
Naive bayes	0.44
Jrip	0.649
Adaboost + JRiP	0.82

Table 4 Changes in confusion matrix due to adaptations

	Without Adaptation		With Adaptation	
	Train	Test	Train	Test
True positive	877	205	1212	317
False positive	0	47	118	92
True negative	404	54	286	9
False negative	421	120	86	8

8 Conclusions

As a classification technique, LAD appears to be competitive with the well-established methods in this area. It is easily interpretable and has wide applications, given that it is not bound to any specific specialties related to datasets. Its high classification accuracy, comparable to and frequently exceeding other methods, and ability to handle some missing data provide robustness to the model's applicability. It is also worth noting that imperfect patterns can also improvise the model given a threshold. Also, combining a few different kinds of patterns has helped reduce computation and time complexities. The results of the Power Attack Systems Case Study show many opportunities for LAD in developing new Smart Grids. The paper concludes by opening the following discussions:

- Exploring more applications of LAD in different sectors and expanding its concepts to ternary or even multi-class systems.
- While even degree three computations are compatible with most datasets, with the right resources, LAD could even be used for Big Data problems with the help of higher degrees along with combinations of higher degrees additionally.

Acknowledgements This work was funded by the Interdisciplinary Cyber Physical Systems (ICPS) division of the Department of Science and Technology, Government of India, under Research Grant DST/ICPS/Cluster/CS Research/2018(General).

Data Availability The Power System Attack Dataset that supports the findings of this study is available at https://www.kaggle.com/datasets/bachirbarika/power-system?resource=download and this dataset is explained in the paper Borges et al (2014).

Declarations

Conflict of interest The authors declare that there is no conflict of interest in this work.

References

Ahmed M, Pathan ASK (2020) False data injection attack (fdia): an overview and new metrics for fair evaluation of its countermeasure. Complex Adapt Syst Model 8(1):1–14

Alexe G, Alexe S, Bonates T et al (2007) Logical analysis of data - the vision of peter l. hammer. Ann Math Artif Intell 49:265–312. https://doi.org/10.1007/s10472-007-9065-2

Almuallim H, Dietterich TG (1994) Learning boolean concepts in the presence of many irrelevant features. Artif Intell 69(1–2):279–305

Bi S, Zhang YJ (2014) Graphical methods for defense against false-data injection attacks on power system state estimation. IEEE Trans Smart Grid 5(3):1216–1227

Bonates TO (2007) Optimization in logical analysis of data. Rutgers The State University of New Jersey-New Brunswick

Borges R, Beaver J, Buckner M et al (2014). Machine learning for power system disturbance and cyber-attack discrimination. https://doi.org/10.1109/ISRCS.2014.6900095

Boros E, Hammer PL, Ibaraki T et al (1997) Logical analysis of numerical data. Math Program 79(1):163–190

Boros E, Hammer PL, Ibaraki T et al (2000) An implementation of logical analysis of data. IEEE Trans Knowl Data Eng 12(2):292–306

Celik ZB, Fernandes E, Pauley E et al (2019) Program analysis of commodity iot applications for security and privacy: Challenges and opportunities. ACM Comput Surv 52:1–30. https://doi.org/10.1145/3333501

Chalapathy R, Chawla S (2019) Deep learning for anomaly detection: a survey. arXiv:1901.03407

Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv 10(1145/1541880):1541882

Cherdantseva Y, Burnap P, Blyth A et al. (2016) A review of cyber security risk assessment methods for scada systems. Comput Secur 56:1–27 https://doi.org/10.1016/j.cose.2015.09.009www.sciencedirect.com/science/article/pii/S0167404815001388

Das TK, Adepu S, Zhou J (2020) Anomaly detection in industrial control systems using logical analysis of data. Comput Secur 96(101):935 https://doi.org/10.1016/j.cose.2020.101935www.sciencedirect.com/science/article/pii/S0167404820302121

Das TK, Ghosh S, Koley E, et al. (2019) Design of a fdia resilient protection scheme for power networks by securing minimal sensor set. In: International Conference on Applied Cryptography and Network Security, Springer, pp 156–171

Deng R, Liang H (2018) False data injection attacks with limited susceptance information and new countermeasures in smart grid. IEEE Trans Indust Info 15(3):1619–1628

Deng R, Xiao G, Lu R (2015) Defending against false data injection attacks on power system state estimation. IEEE Trans Indus Info 13(1):198–207

Deng R, Xiao G, Lu R et al (2016) False data injection on state estimation in power systems-attacks, impacts, and defense: A survey. IEEE Trans Indust Info 13(2):411–423

Giraldo J, Urbina D, Cardenas A et al (2018) A survey of physics-based attack detection in cyber-physical systems. ACM Comput Surv 51:1–36. https://doi.org/10.1145/3203245

Hammer PL, Bonates TO (2006) Logical analysis of data-an overview: From combinatorial optimization to medical applications. Annal Oper Res 148(1):203–225

Hink RCB, Beaver JM, Buckner MA, et al. (2014) Machine learning for power system disturbance and cyber-attack discrimination. In: 2014 7th International symposium on resilient control systems (ISRCS), IEEE, pp 1–8

Hossin M, Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manag Process 5(2):1

Li S, Yilmaz Y, Wang X (2014) Quickest detection of false data injection attack in wide-area smart grids. IEEE Trans Smart Grid 6(6):2725–2735

Liang G, Zhao J, Luo F et al (2016) A review of false data injection attacks against modern power systems. IEEE Trans Smart Grid 8(4):1630–1638

Liu X, Li Z (2017) False data attack models, impact analyses and defense strategies in the electricity grid. Elect J 30(4):35–42

Liu Y, Ning P, Reiter MK (2011) False data injection attacks against state estimation in electric power grids. ACM Trans Inform Syst Secur (TISSEC) 14(1):1–33

Liu L, Esmalifalak M, Ding Q et al (2014) Detecting false data injection attacks on power grid by sparse optimization. IEEE Trans Smart Grid 5(2):612–621

Liu X, Li Z, Li Z (2016) Optimal protection strategy against false data injection attacks in power systems. IEEE Trans Smart Grid 8(4):1802–1810

Liu C, Wu J, Long C et al (2018) Reactance perturbation for detecting and identifying fdi attacks in power system state estimation. IEEE J Sel Top Signal Process 12(4):763–776

Mitchell R, Chen IR (2014) A survey of intrusion detection techniques for cyber-physical systems. ACM Comput Surv (CSUR). https://doi.org/10.1145/2542049

Nazir S, Patel S, Patel D (2017) Assessing and augmenting scada cyber security-a survey of techniques. Comput Secur. https://doi.org/10.1016/j.cose.2017.06.010

Phadke A, Volskis H, de Moraes RM et al (2008) The wide world of wide-area measurement. IEEE Power Energy Mag 6(5):52–65

Singh SK, Khanna K, Bose R et al (2017) Joint-transformation-based detection of false data injection attacks in smart grid. IEEE Trans Industr Inform 14(1):89–97

Sridhar S, Hahn A, Govindarasu M (2011) Cyber-physical system security for the electric power grid. Proc IEEE 100(1):210–224

Yang Q, Yang J, Yu W et al (2013) On false data-injection attacks against power system state estimation: Modeling and countermeasures. IEEE Trans Parallel Distrib Syst 25(3):717–729

Yang L, Li Y, Li Z (2017) Improved-elm method for detecting false data attack in smart grid. Int J Electr Power Energy Syst 91:183–191

Zacchia Lun Y, D'Innocenzo A, Smarra F et al (2018) State of the art of cyber-physical systems security: an automatic control perspective. J Syst Softw. https://doi.org/10.1016/j.jss.2018.12.006