JOURNAL OF INTERNATIONAL CRISIS AND RISK COMMUNICATION RESEARCH
ISSN: 2576-0017
2021, VOL 4, NO S4

Machine Learning Algorithms for Real-Time Fault Detection and Performance Enhancement in Solar Energy Systems

Venkata Narasareddy Annapareddy

Sr Enterprise Application Developer, venkata.narsreddy@gmail.com, ORCID ID:0009-0002-5385-5259

Abstract

This thesis presents machine learning algorithms and models for enabling real-time fault detection and performance enhancement of Solar Energy Systems. Modern Electroluminescence imaging technology and high-performance parallelizable ML models are used to scrutinize the condition of Electronic Focus Solar Energy Systems during power generation operations without shutdown. Utilizing the ML models trained on historical Solar Energy System generations and Electroluminescence information, we determine temporal performance degradation characteristics at otherwise latently generating Solar Energy Systems. Through temporal analysis of the actual degradation characteristics, (1) detection of a fault during generation operations, (2) prediction of cell failure time and degradation characteristic, and (3) estimation of the actual degradation characteristic, are enabled.

It is demonstrated that widening the training data window for ML model training enhances the temporal performance of the ML model. Furthermore, a higher risk of faults is identified for Solar Energy Systems located in dustier desert conditions, and these systems should be specially monitored or have preventive maintenance carried out. Experimental results indicate that the sizes and distributions of degradation zones may differ among some Solar Energy Systems, likely caused by effects like spatial temperature non-uniformity and undesired metal bridge pollution. For Solar Energy Systems attached with back contact solar cells, ML models indicate that there is a higher risk of cells with the back contacts being damaged and losing focus. To enhance the promulgation potential of the ML models proposed, we propose a new method of enfolding the temporal characteristics of degradation of Electronic Focus Solar Energy Systems in the period of risk detection based on temporal ML model enhancement. If any risk of fault or fault of an Electronic Focus Solar Energy System is detected, proper maintenance and preventive actions should be conducted to keep the systems from going through latently generating periods.

Keywords : Machine learning, algorithms, real-time, fault detection, performance enhancement, solar energy, photovoltaic systems, data analytics, predictive maintenance, anomaly detection, sensor data, energy efficiency, classification, regression, supervised learning, unsupervised learning, feature extraction, data preprocessing, renewable energy, system monitoring, neural networks, support vector machines, decision trees, random forests, deep learning, smart grid, IoT, data-driven models, model accuracy, optimization, diagnostics, power output, solar irradiance, system reliability, early warning, remote sensing, automation, real-time analytics, fault tolerance, prognostics.

1.Introduction

Increasing consumer energy demand due to rapid population growth has created the necessity of various energy generation techniques in addition to conventional methods. To meet economic growth and provide clean energy security, many governments have adopted policies that encourage investment in solar energy. However, solar energy systems encounter various problems during operation. These problems include energy conversion loss due to dust accumulation on photovoltaic panels, the impact of shading on energy output at certain times, decay of energy conversion efficiency due to frequent temperature changes in thin-film panels, and wear-out faults in energy storage devices caused by excessive temperature differences among cells. Faults in solar energy systems during operation lead to increased energy loss and operational cost, so real-time fault detection and performance enhancement in solar energy systems is a significant area of research.

Artificial intelligence-based techniques, including machine learning algorithms, have recently shown a rapid increase in development and application in various domains. Moreover, their vast benefits such as ability to deal with large amounts of data and well-developed techniques for making decisions at real-time speed, machine learning algorithms can detect faults in solar energy systems from operational data recorded in monitoring devices. Machine learning-based fault detection techniques have already been proposed for different solar energy system problems with acceptable performance. However, current research on solar energy systems lacks the use of different machine learning algorithms equipped with hyperparameter tuning strategies. Therefore, potential researchers may have difficulties in selecting the best machine learning-based fault detection technique for their solar energy system problems. There is also very little research on label and unsupervised machine learning-based performance enhancement techniques for solar energy systems. This chapter not only summarizes existing published machine learning-based research for fault detection in solar energy systems but also addresses issues in existing techniques and provides a comprehensive discussion on potential label-based and unsupervised machine learning-based performance enhancement techniques.

2. Overview of Solar Energy Systems

Solar power generation is achieving prominence and is being considered very useful on power generation avenue, as it is easy to harness, has no running costs, is pollution-free, is emissions-free, doesn't require chemical/nonchemical reaction, is capable of generating electricity for both localized load as well as grid-connected systems, requires compact solid-state gadgets, operates on P-N junction effect of semiconductors having variety in spectral responses, is not only environmentally-assistive and quiet but also generates reusable bio-products, is long-term sustainable with more than 60 years service life in case of crystalline class, silicon space modules used in solar cars and satellites to withstand harsh environment and not dependent on fossil fuels or any limited resource being abundantly available. However, capital cost is high and indium and gallium in case of thin-film technologies are limited resource. Being abundant source of solar availability, India with its geographical location has been entitled by enormous solar radiation for a longer period.

Solar energy has been a very popular choice of attention due to its free availability and also because of being one of the renewables. Innovation and research have been undergoing for past decades on solar energy technologies with involvement of many organizations to uplift the technology for future use and also reduce the capital cost for every household in order to cover widespread development and go for a budgeted solar PV system usage with storage facility for minimizing the gap on consumption and availability. Solar energy system consists of solar cell modules, inverter, storage system, load optimizer and control system. These components will be useful in order to do the integration of system onto a compact module.

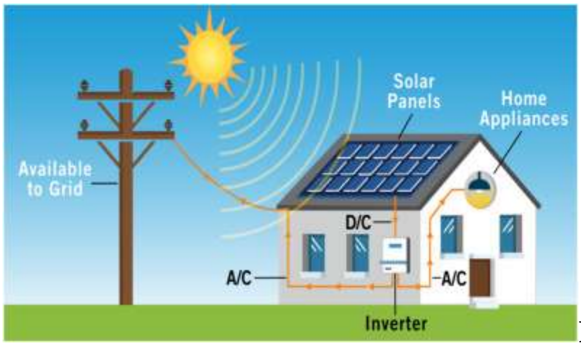

Fig 1 : Solar Photovoltaic System

2.1. Types of Solar Energy Systems

Solar energy is clean and natural energy generated from sunlight. Solar radiation is transformed into heat to be used directly or converted into other energy forms, such as heat, electricity, or fuels, through different technologies known as solar energy systems. There are two basic categories of solar energy systems: solar thermal systems and solar photovoltaic systems. Solar thermal systems collect thermal energy; use naturally occurring elements, such as water, to store heat; and are primarily employed for heat generation where sunlight is available, such as solar water heating, solar heating, solar air conditioning, and solar cooling. Solar photovoltaic systems convert solar radiation into electricity for use in households, businesses, and industries.

More precisely, solar energy systems are classified into two broad categories: solar thermal systems and solar photovoltaic systems. Solar photovoltaic systems convert solar radiation into electricity for own use in households, businesses, and industries, and for export into the utility grid. Every hour, the Earth receives more energy from the sun than what is used by humans in a whole year. For this reason, and given the frequently growing need for energy driven by pollution and fossil fuel depletion, many efforts, scientific and technological, have been put in the research for the design and implementation of efficient solar cells. All photovoltaic systems, from traditional monocrystalline technology to thin-film modules and concentrators, aim to maximize the efficiency, considering also the reduction of the resources needed for their production, stressing also the importance of recycling photovoltaic panels at the end of their life. Solar thermal collector systems collect thermal energy, which is then used to heat naturally occurring elements, such as water, and are, thus, mainly utilized for heat generation, as in solar water heating and solar air conditioning. These types of applications find a better location in countries where sunlight is available most of the time.

2.2. Components and Functionality

Solar energy systems capture radiated solar energy to convert it directly into electricity by means of photovoltaic (PV) modules, or into storable thermal energy by means of thermal collectors or solar fuels. Each technology can be divided into different categories according to their configuration or how they utilize solar energy. In a PV module, a solar cell chip containing a semiconductor material converts sunlight into direct current (DC) electricity. While the module generates DC electricity, the inverter converts it into alternating current (AC) electricity of adequate voltage and frequency level to be injected into an electric grid or be utilized for running electrical loads at home. In thermal solar energy systems, a collector can be flat plate, evacuated tube, or parabolic trough type that utilizes sunlight to heat air or some working fluids such as water or oil. The heated fluid can be used to drive a turbine generator to produce electricity or be directly provided for heating services in different applications. Another technology is solar fuels in which solar energy is used to crack raw materials such as biomass or gasified coal to produce hydrogen or some other solar fuels. Here, we will first discuss the components of PV systems, both at the home level and utility scale, and their functions. A solar PV system comprised of PV modules connected to the electric grid consists of a few components. The array consists of several PV modules connected in series or parallel. The inverter converts the generated DC electricity into AC electricity. Then, the electric distribution board provides the AC power to the electrical loads and injects the excess power to the electric grid. Other accessories include markup and connection devices, battery, battery charger, combiner box, surge arrestors, lightning conductor, and meters. The utility scale solar PV plant is larger in power level than a grid-connected solar PV rooftop system installed at the home. Major components, configuration, and functions of both solar PV plants are quite similar with a hybrid configuration of PV-fuel cell battery system modified for better availability at home level.

3. Importance of Fault Detection

As is the case with a majority of the anomalous operation of the PV plant's components, FDs and diagnoses can facilitate timely and necessary actions to restore the normal operation of the power plant and recover its asset value by enabling timely interventions. The significance of FD and diagnosis is quantitatively seen in the potential for the augmentation of the availability and reduction of the outage of the systems. Both of these serve to substantiate the return on investment and overall profitability of the solar PV systems serving as a source of green energy. The early detection of faults allows timely maintenance or repair and can avert severe failures that would endanger the whole system, cause extended downtimes, and result in costly replacements of the components.

The research works on model-driven and model-less approaches mostly highlight the very significant impact of the most common fault on the performance and efficiency of the solar PV systems. Faults can occur in any of the components, they are inverters, modules, series fuses, etc. and if undetected they have a very severe effect on the performance of the solar PV systems. The effects of the most common faults on the PV system performance and safety characteristics are simply rooting at the reduction of the output energy which is a very commonly accepted view in all research works on fault detection.

Safety, performance, and investment protection are the key concern for any owner/investor of the solar PV systems. A timely detection and remedy of faults ensure the system plays its part in the energy transition and is contributing the prescribed energy quantities by the strategic plans at the local, regional, and national levels.

$$\text{Residual } r(t) = y(t) - \hat{y}(t)$$

Eqn 1 : Model-Based Diagnosis

Where:

- $y(t)$: Actual system output
- $\hat{y}(t)$: Estimated output from a model
- If $r(t)$ is significantly non-zero, a fault is likely present.

3.1. Impact of Faults on Performance

While solar photovoltaic (PV) energy generation systems are resilient and hassle-free, they are still subject to faults caused during design, material sourcing, and installation as well as those occurring due to environmental factors. The different systems composing solar PV generation systems comprise modules that include photovoltaic cells, inverters, battery energy storage systems, maximum power point tracker circuits, electrical wiring systems, and monitoring equipment. Any faults occurring in any of the system components would lead to a drop in performance. It is estimated that the failure rates over the first 5 years of operation range between 10% and 15%, leading to a permanent drop in nominal output of around 6%. More than half of current utility-scale PV systems are in regions dominated by energy-limited interconnection agreements. Moreover, while PV panels in operation are subject to performance degradation, the effects of high temperatures, moisture, and UV exposure also lead to gradual materials degradation. Therefore, while service times of grid-scale systems are currently in the 20- to 30-year range, the so-called "standard drop" in performance may drop in excess of 20% within this timeframe.

A decrease of efficiency directly affects the overall performance and also has long-term implications on the service life of the entire solar energy system. Hence, several solutions have been proposed to enhance the performance of solar energy systems. Some of the causes of faults in PV systems include temperature variations, humidity, irradiance level, and dirt accumulated on PV panel surfaces. Physical impacts or degradation in panel surfaces may cause reduced power generation from installed solar systems and may even lead to panels catching fire. Some of the potential faults in inverters could be damage to the circuit board due to overheating or faults-induced overcurrents or overvoltages.

3.2. Economic Implications

Solar energy is the most abundant source of all renewable energy sources available on the earth; being harnessed using the PV effect discovered over 100 years ago. With the need for tackling climate change and the limitation of fossil fuels, there is a growing interest in harnessing solar energy. Costs of solar panels have drastically reduced in recent years making solar power generation an economically viable option globally. The intermittency of solar energy has to be taken into consideration when utilizing it for electricity generation. The fluctuation in solar irradiation level mainly affects the lifetime of the solar photovoltaic (PV) modules and also the efficiency of the system. PV panels, when exposed to changing weather conditions, suffer from environmental pollution, cooling, thermal cycling stress, and faulty internal electrical circuits. These faults drastically reduce the energy output from the PV plants, leading to economic loss. Different types of faults like Partial Shading Conditions (PSC), bypass diode failure, cell fracture, or high-series resistance, suffering from unknown electrical parameters are addressed by different research works. With regards to this, Module-Current Sensible Conductance (MCSC) method is a new method proposed to detect cell fracture faults through current-voltage characteristics of PV modules under normal operational conditions.

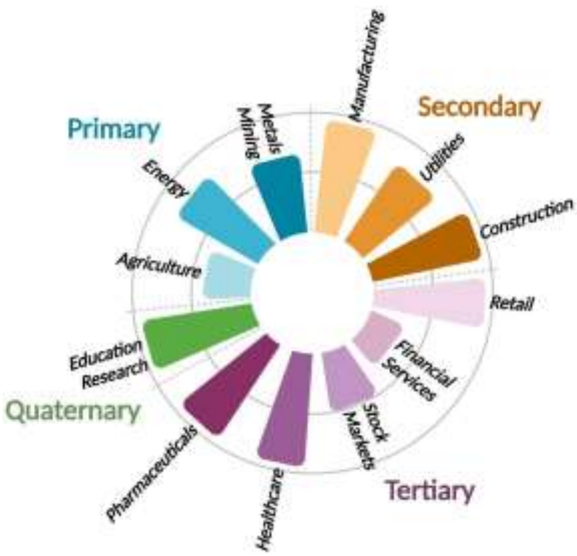

Fig 2 : Socio-Economic Impacts and Challenges of the Coronavirus Pandemic

Solar energy systems are increasing in size for distributed generation plants and on rooftops for power generation and reduction of electricity bills. Growing installed capacity for fossil-fuel-based generation is concerning because of the impact it has in terms of generation by polluting the environment and the adverse effect it may have on climate. With the increase in the installed capacity for solar energy systems globally, an equally important task for researchers and industry experts is to develop techniques and best operating practices that could enhance productivity and reliability, ensuring the generation of energy throughout its design life of PV plants. This is mainly due to the fact that a large portion of the total cost of solar energy systems is incurred during the materialization phase. By some estimates, for reasons of loss of energy production, operational expenses roughly account for 10% of the life-cycle balance of grid-connected solar energy systems.

4. Machine Learning Fundamentals

Machine learning is an artificial intelligence subfield that focuses on creating algorithms that enable computers to learn from data. Different data types and collections typically require different types of machine learning algorithms and would be described below. More formally, machine learning is the field of algorithm creation that allows learning as a result of experience or data and that typically means allowing performing a task without being explicitly programmed for that task. In more colloquial terms, machine learning is about building models from data instead of traditional programming, which is building models from code. Data are the inputs and ground truth is the expected outputs. In this regard, a machine learning model comprises a mathematical representation of the relationship between its inputs and outputs. Most commonly, the relationship is expressed by parameterized mathematical functions, referred to as model parameters that are typically based on the relationship expressed during the learning process as a result of solving an optimization problem.

The learning process consists of two main stages: a training phase and an evaluation or testing phase. During the training phase, the model parameters are initialized randomly and are then adjusted, usually based on some optimization technique, towards minimizing the training error or towards maximizing the likelihood of the data had come from the model defined as a function of the model parameters. After running the training phase, the model is said to be learned or trained. The estimation of the training error provides a measure of how well the model fits the training data, but it can also provide a false sense of security, as research in machine learning has shown that it is often better to have a low testing error than a low training error. The testing error estimates how likely the model was able to generalize to a new, completely different dataset, and hence how well it would be fairing on a real-world task.

4.1. Supervised Learning

The automation of fault detection and performance enhancement in solar energy systems can be reliably addressed using a class of machine learning algorithms termed supervised learning. Lack of accurate models due to uncertainties associated with solar energy systems calls for the practical application of data-based fault detection methods and performance enhancement tuning strategies. At its core, supervised learning algorithms learn a mapping from an input feature space to an output target space, using a training data set containing paired input feature vector and output target vector. In the context of solar energy systems, the feature vector and target vector can belong to a wide spectrum of science fields, both theoretic and empirical. Such feature-target pairings can be based on physics-based performance models of the solar system, performance data of historical anomalous operating conditions, or domain specific expert knowledge. In a broad sense, supervised learning algorithms working with heterogeneous input feature-target pairings seek to approximate the relationships between system operating conditions and the underlying physics that govern the performance of solar energy systems.

Supervised learning provides more reliable data-driven solutions for fault detection and performance enhancement tuning strategy than unsupervised and reinforcement learning algorithms, as they use domain driven prior knowledge to efficiently model the system input-output relationships. The two main challenges in using supervised learning algorithms is selecting the right input features and obtaining sufficient labeled data for training. A compact feature vector must adequately represent the representative aspects of the solar energy system to expose its underlying physics when presented to a supervised learning model. Input feature selection can become more complicated, depending on whether the desire is to have a model that is interpretable or produces the highest performance. Performance data of historical anomalous operating conditions can be particularly useful in supervised learning applications. However, the challenge remains in acquiring high quantities of labeled fault data through domain specific expert knowledge. The main advantage of supervised learning algorithms is the model robustness that emerges from their learning ability to generalize.

4.2. Unsupervised Learning

The unsupervised learning dynamically recreates the input data measure mapping function. In actuality, it analyzes and examines the given sets of unlabeled input data to extract meaningful patterns or infer the data distribution that defines the input space. Unsupervised learning projects the most significant components but newly creates the original input data space. The components that describe the input data space more succinctly directly represent the input data and preserve more significant information. The applications of dimensionality reduction or data compression, such as data de-noising or data visualization, particularly utilize unsupervised learning function. Even if the input data representation loss internally occurs, it is too small or negligible. On the other hand, classification or generative modelling attempts to minimize the loss of the unlabeled data reconstruction. The generative models in unsupervised learning are generative attempts of the input data. They try to create viable, realistic data examples resembling the probabilities of given examples from the input data distribution. In this sense, they are also called density estimation models or density estimation attempts. They exactly fit the original input data to recreate their original space to the maximum extent possible through their probability distribution to perform their function correctly. The classification may appear as contraction attempts of the input data but based on specific, distinct categories or classes rather than very close examples. Many popular artificial intelligence applications have emerged from unsupervised processes.

Eqn 2 : Loss Function (typically Mean Squared Error

$$L(x, \hat{x}) = \|x - \hat{x}\|^2$$

Where:

- x: Input
- $\hat{x}$: Reconstructed input
- W, W': Weights
- f, g: Activation functions

4.3. Reinforcement Learning

Reinforcement learning (RL), the final branch of ML, is the domain wherein the model trials out various actions in a certain environment and with the help of rewards, learns how to achieve the goal. During the training phase, RL learns by an enumeration strategy to explore the action space and learn through comments or rewiring to exploit the best actions. On trial and error, the agent develops a policy, which is the mapping between situation (often denoted as state) and the actions. The main concept in RL development involves learning what to do in some conditions to make the most rewards over all time. The reward signal issued at feedback period reflects the degree of success from performing the actions. In RL, feedback continues to clarify whether or not the actions are helping to achieve the goal. In applications like video games, RL has solved problems where the goal was clear, such as achieving a high score.

RL is different from supervised, or unsupervised, learning; no teacher specifies the correct action for any input or state. RL is inspired by the operant conditioning of animals. It is one way in which animals learn to choose what to do in their environment. The animal operates on the environment and receives rewards or penalties. This feedback indicates how near or far the animal is from some goal or goals, such as exploring its environment or getting food. Based on its control over its environment, a reinforcement agent learns to optimize its deserved reward signal over time with limited knowledge of how its environment operates and conditions its performance. In principle, the environment for which its expert model ultimately is trained includes everything external to the agent that communicates capabilities and evolves in accordance with those capabilities over time.

5. Data Collection and Preprocessing

Solar energy systems serve as the leading source of renewable energy generation, nowadays. Such systems are generally comprising various apparatuses, including solar panels, power converters, batteries, and the grid connection. Since the deployment of solar energy systems has witnessed continuous growth over the years, the need for efficient System Performance Monitoring has become increasingly pressing. If a solar energy system is disconnected from the grid due to any underlying fault, the capability of providing energy back to the grid is compromised until the fault is recognized and repaired. Therefore, the challenge of fault detection in solar energy systems must be addressed in a timely manner in order to enhance sustainable and reliable operation of such renewable energy sources and avoid economic loss. Generally, such systems are monitored in real-time by utilizing performance metrics often wrangled or fused from sensors data. This chapter provides a comprehensive overview of data collection and processing methodologies. The main contributions of this chapter can be summarized in the items below: 1. Exploring various cleaning techniques to refine the raw data and improve data quality to allow for valid real-time monitoring and efficient model training. 2. Providing insights on feature selection and engineering techniques to transform raw data into performance metrics to suit different fault detection or diagnosis purposes. 3. Enabling new researchers and engineers to understand, utilize, and enhance existing performance monitoring methodologies to solve domain-specific issues. This section presents the needed description and procedures concerning the above objectives. Further, it also highlights the importance of each decision in the data preparation part towards ensuring efficient fault detection models are implemented. The proposed data processing methodology utilizes various cleaning and transformation techniques to convert Sensor-Combined AC Voltage and Current Measurements to key performance metrics: Solar Energy System Yield and Performance Ratio. Through this applied data processing, the entire data are first segmented into one-hour equally-sized bins before using pre-defined cleaning techniques to enhance data quality. Next, each bin is then converted into metric values that reflect the system performance during that time period.

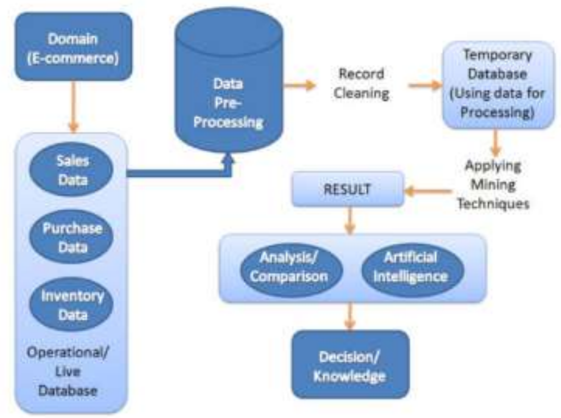

Fig 3 : Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process

5.1. Types of Data Collected

The commercialization of photovoltaic technology has made it possible to grow extensive solar farms with massive numbers of PV panels and power conditioning units to deliver an increasing contribution of power to emerging markets. Therefore, there is great commercial and environmental value in algorithms that can detect and monitor faults in real-time and promote the maintenance of these systems. A detailed list of the architecture of the solar field and the sensors required for this system to work is given in the following sections. The major solar irradiance and temperature sensors are imported and off-board. The site location, time of day, daily weather progression, and yearly schedule are used to calculate the expected solar energy output for comparison. Monitoring of the health of these solar systems using machine learning for fault prediction requires the collection of time-series data from the solar field as it is functioning during normal operating conditions, as well as the periods when faults are detected by the time-correlation analysis of the expected solar output and the time-stamped field data. Five different types of data have been collected during this project. The principal data are extracted from the site location, calculated efficiency maps, and time stamps for periods of operation at certain azimuth and elevation angles. This is used with the solar output for the average current flowing through the electrical wires over the period of record length. The solar irradiance output is compared with local data and the excellent correlation shown in this chapter is used to infer the estimated solar output levels.

5.2. Data Cleaning Techniques

Data acquired from real-world energy systems may not be perfect. Missing values, outliers, and noise are potential issues that corrupt real-world data and degrade the performance of data-driven models. For a machine learning model, the accuracy of predictions and decision processing relies on the quality of the input data. Forecasting the energy output of an energy system helps to detect faulty components based on the predicted values. Most of the energy systems use monitoring systems to monitor system behavior and any fault. These systems collect, monitor, and provide a stream of system-level time-series data that help to identify the current state of the system. Due to the malfunction of any of the sensors, data may get corrupted.

To give valid conclusions, separate studies have proposed and experimented with different techniques to evaluate and rectify real-time events affecting system performance and health. Not all the research work modifies faulty data to use them in the data-driven model, while others remove the faulty data. Different studies give a valid discussion about how to manage the faulty data, which helps in improving not only the performance of the model but also improve the model accuracy. Moreover, a similar approach may not be used for different systems. Each system has its own nature to deal with the detected fault for robust prediction. In addition, the information loss from the replacement of missing values or removal of the record from the data set tries to justify active decision models, but it also adds noise to the model accuracy. Many solutions have been proposed to deal with/mitigate these noise and faulty data detected from time series dataset.

5.3. Feature Selection and Engineering

Feature selection or feature engineering is the process of selecting the salient data features relevant to the classification task from the available data attributes during the data preparation step before the classification is performed. Selection of the most informative features can reduce the dimension of the feature space and increase the generalizability of the classification tools. It allows probabilistic classifiers to avoid the curse of dimensionality. Feature selection is important as it can potentially improve the performance of a classifier, improve generalization, and reduce the computation cost of modeling. The process involves the crucial step of deciding what features to include when training a machine-learning model.

$$Var(X_j) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$$

Eqn 3 : Variance Thresholding

Where:

- X_j: Feature j
- x_{ij}: Value of feature j for sample i
- $\bar{x}_j$: Mean of feature j

There are mainly three approaches to feature selection: filter, wrapper, and embedded methods. In filter methods, a metric is computed for every feature using a statistical measure, which is then used for selecting the feature set. In wrapper methods, a smaller subset of features is selected using a predictive classifier to evaluate the performance of any combination of features. The process is repeated by adding or deleting features until optimum performance is achieved and finally the model accuracy is cross-validated using another different dataset. In embedded methods, the feature selection process is combined with analysis and training of the model. The three types of feature selection methods are first described in more detail as follows, where we also briefly discuss how these methods can be used to conduct feature selection for real-time hardware-and-feature-constrained applications.

6. Fault Detection Algorithms

Systematic evaluation of the numerous existing Fault Detection, Diagnosis and Exclusion methods and algorithms that are traditionally developed is considered to be a complex process. In this study, a number of Fault Detection Tools, many relative to the ever-developing machine learning techniques, are assessed. The aim is to provide a comparison of the several researched algorithms and their key characteristics to the attention of practitioners. There are several algorithm types for fault detection by means of general machine learning methods. In the chosen application area, solar energy research, a set of supervised, unsupervised, classification, regression and ensemble learning techniques are implemented and classified.

While classification algorithms seek to divide data points into discrete categories, regression models attempt to predict continuous outputs. Shared use of both approaches is frequent. For instance, decision fusion applied on separate results from model parameters estimation and classification detection techniques are available and usually lead to an improvement in the ultimately provided results. Anomaly Detection methods, when applied in unsupervised mode, take the raw signals after preprocessing as input to the algorithms. Provided the official labels or class characteristics of the anomaly types are not included in an input dataset, artifacts other than the resident/target faults in the signals must be preprocessed or their presence eliminated from the raw signals.

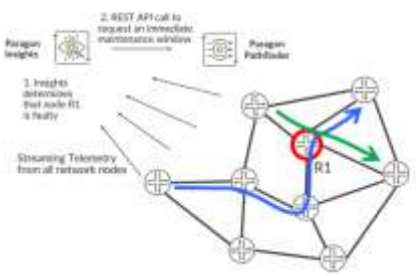

Fig 4 : Automated Fault Detection and Remediation Becomes a Reality with Paragon Automation

6.1. Anomaly Detection Techniques

Machine Learning is currently an evolving technology that is used in various sectors for a variety of applications. It aids in troubleshooting and fault diagnosis of power converters, wind energy systems, communication networks, solar photovoltaic systems, and spacecraft fault detection. People are applying different machine learning techniques over the years and helping in conversion improvements with a little assistance. Anomaly detection is an approach to the detection of faults in the characteristics of the data groups based on machine learning principles. The idea is to check the existence of the data in earlier developed singularity and monitoring.

The existing work on anomaly detection techniques suggests the application of one-class classifier methods in the implementation of the anomaly detection problem, since it has become a more generalized form of supervised learning. The major limitation of this method is giving the 'one-class' class label training sample, which causes difficulty for real-world problems. Two-class problems provide large amounts of instances for normal and abnormal situations. To uncover hidden instances, the concept of a local region around the query point embedded in a high-dimensional space is used. Then the local region is analyzed in search of sparsity. Among them, the Principal Component Analysis (PCA) is widely accepted for feature extraction to reduce the dimensional size of input data, and has been successfully employed to solve different applications. In solar photovoltaic systems, PCA is presented based on feedforward neural networks. The learning process can be realized in a supervised or in a structure-less way, if no label is available, which allows giving the labels afterwards. Other kinds of local region formulation can also be introduced in the PCA scheme or in different image processing techniques used for fault detection in solar panels.

6.2. Classification Algorithms

Classification algorithms can be efficiently deployed for classifying faults at the supervisory level, which has limited computational resources compared to the host computer or the cloud, which has nearly unlimited resources. Such a deployment is also potential for real-time applications with latency guarantees. Classification models can be trained as an external task. The input data for such a model is labeled interpretation of a fault, while features can be derived from different signals, for example, the string, inverter, or [...]

Several such classification models can also be trained in a hub-and-spoke architecture, where different error conditions are modeled as spokes, while a hub gathers the appropriate spokes in their respective subspaces to classify the type/severity of an issue. However, in these models, the supervision should be specified a priori. Machine learning-assisted methods trained on raw data can address the required flexibility. We also prefer model-agnostic models, especially when there are researchers in parameter estimation–based models because of the requirement of domain expertise and interpretability of the resulting model parameters in available models and the domain expertise required for designing a specific architecture in deep learning. These models have user-friendly interfaces allowing for drag-and-drop training of the models and consequently reducing the associated overhead.

The supervised ML methods, such as support vector regression or ensemble machine learning algorithms performance are extremely good with prediction accuracy for faults in module and inverter parameters as compared to the other models. The data-driven technique have been utilized to optimize PV installations, specifically in assessing and identifying the anomalies. The supervised machine learning algorithms have been trained with the sensors' data and reported accuracy to identify the anomalies.

6.3. Regression Models

Several regression models, including Regression Trees, Non-compensatory Weighted Scoring Function, M5 Model Tree, and Multilevel Multiple Regression, have been employed for FDI. The minimum output of the Capacitance/Voltage curve can be directly used to calculate the quality factor for the FDI task for the samples in this paper because the region between the two vertical dashed lines is characterized by a minimum value. The increase of defect density causes the draw-in of the defect-related levels, resulting in the increase of the contribution of the defect levels into the C/V, and consequently leading to a decrease of the quality factor. The defect-related levels are fabricated near the valence or conduction band, and the quality factor increases with either of the energy levels moving away from the respective band. As such, the defect levels, with low defect density, are far from the conduction or valence band; otherwise, the capacitance may not approach 0 at M/FCV.

Utilizing Poly-Si TFT device reliability-related characteristics to predict the defect density, the developed models served as FDIs for the generated devices in this work. A two-step procedure is employed for the fault detection and prediction for the devices in this paper. In the first step, for the test samples from a Poly-Si TFT device process, the regression models identify the defective samples, and the samples with low defect density but not yet passed the process qualification apply for the fault prediction in the second step. The post-retail review stage has been streamlined and formulated, thus the process customers will benefit from the more reliable Poly-Si devices.

7. Performance Enhancement Algorithms

The power generating capabilities of solar energy systems can be degraded by the presence or continuous operation under fault conditions. Unfortunately, the presence of some faults may not generate any disturbing indications, for instance, if the solar remittance is not suitable to perform a correct evaluation of the electrical performance. In both cases, the generation of power within the maximum capabilities of the system can be avoided. It is shown how predictive maintenance and optimization techniques can be gathered within machine learning models to minimize risk situations of failure in these systems, or to minimize the reduction of power generation capability if some faults are detected or if the system is operated under non-maximum conditions.

Predictive maintenance tries to avoid a fault to happen and try to optimize the performance of the solar system. The use of machine learning models together with accumulated data permits to calibrate precursors for each specific PV system and be used to provide future indications of fault probability, and used in predictive maintenance tasks.

Optimization techniques will be used to help increase the power output of a solar energy system. In this case, machine learning algorithms will consider the installed capacity as the main variable defining the objective function in conjunction with weather data. The explanation of the location of these models is that they allow to minimize the error related to the PV production model. The parameters defining the DC model will be optimized by making use of a greater dimension of the data variables. The decision capture are important for the final location of the performance enhancement systems.

7.1. Predictive Maintenance

Machine learning is increasingly being used for predictive maintenance, which aims to address defects like module mismatch and module soiling in solar panels before they lead to failure. Predictive maintenance is preferred over reactive maintenance, which is the most commonly used strategy today, as the latter suffers from downsides like increased energy generation losses, potential health risks resulting from unsafe working conditions for technicians, and increased replacement costs from complete failure. Implementing predictive maintenance strategies based on machine learning algorithms have the benefits of minimizing maintenance and repair costs and maximizing energy generation and utilization efficiency, but more research is still needed towards their integration in the solar energy industry.

Several researchers have already demonstrated the effectiveness and validity of using machine learning algorithms for predictive maintenance. A method for module mismatch detection and prediction based on pattern recognition and an artificial neural network using the environment temperature, maximum power, and PID temperature values has been designed. Machine learning has been utilized for predicting the defects in PV systems for understanding the risks at a regional level. A predictive maintenance strategy for AC fuse degradation in PV inverters has been proposed. The k-nearest neighbor algorithm has been used for predicting the remaining working time of PV field workers. A convolutional neural network has been applied to predict the temperature of PV elements in an image captured via a drone carrying an infrared thermal camera. Machine learning techniques have been investigated for predicting and identifying PV faults. Additionally, a method based on precision hazard recognition for predicting the fault location in PV devices has been proposed. Machine learning algorithms have been used for predicting the failure of PV inverters. A predictive maintenance of a PV performance monitoring system based on a multi-layer feedforward neural network has been proposed.

7.2. Optimization Techniques

Optimization is used for renewable energy applications to enhance energy generation and full use of the resource. The optimization techniques are also referred to as tracking techniques that enable the continuous operation of the energy system at maximum efficiency. These techniques are used to track the peak and maintain the performance at this point during energy generation. The various optimization techniques are continuously changing. The major categories of optimization algorithms include mathematical, curve fitting, heuristic, and soft-computing algorithms. The mathematical algorithms are time-consuming. But they provide more accuracy than other algorithms. The curve-fitting algorithms compute the equation of voltage/current versus irradiance. The heuristic algorithms reduce the number of function calls to some degree. The soft-computing algorithms are the latest techniques aimed at using machine-learning techniques to reduce response time with high accuracy. The soft-computing techniques operate continuously without knowledge of the characteristics of the energy-generating element during the tracking operation. Soft computing techniques are developed using artificial neural networks, genetic algorithms, fuzzy logic controllers, etc.

Some examples of mathematical optimization styles are numerical search techniques, the exact solution of the conditions for optimality, or classical optimization techniques. The optimization with these techniques has limitations. Numerical search techniques can converge slowly, while classical optimization techniques primarily work on convex problems. Curve-fitting optimization techniques perform best for static or quasi-static problems, like 1-D or 2-D Parameter Estimation Problems. These techniques execute generally better than heuristic techniques, but worse than model evolution techniques.

8. Real-Time Implementation Challenges

In the work, we proposed novel machine learning algorithms for fault detection in photovoltaic plants, based on solar plant output data. In order to utilize them in real-time, we need to transform them into real-time classifiers, which work on resource constrained devices and highly volatile data. In this chapter, we focus on the issues related to real-time implementation on such devices and in real-time environments. As far as notification and notifications are concerned, we investigate the challenges for real-time detection of faults in solar plants. These challenges are called data latency, integration related, and computational resource related problems. Data latency is a major issue when using learning based models to detect faults in a solar plant, since the variables relevant to fault detection are fluctuating with time. Data latency makes the data classification very instable, which in turn makes the classifier less usable in real-time. Classifier confidence based approaches provide a solution to this problem. However, they can work only if the model is not computationally heavy. Resource constrained monitoring devices would use the model implemented by low complexity computational resource constrained classifier.

On the other hand, notification of faults with the help of process monitoring sensors would require the machine learning based classifier operating in synchronization with sensor data transfer. Such a real-time operating condition would require the timely availability of data at processing node. Furthermore, in this case, the model should be implemented by computationally light models in order to provide real-time notifications. So the choice of the right classifier is crucial for the real-time implementation. Moreover, sensor placement strategy also has a serious impact on classification accuracy, especially, when high velocity of data transfer is observed. Integration of the classifier based system with existing sensor based fault detection system is a challenge. If both systems are to work semi-independently, then designing a upper layer communication strategy framework suitable for the application is very important. Then the framework is supposed to handle conflict between sensor based and model based fault detections.

Fiq 5 : Top challenges in ERP implementation- Sage Software

8.1. Data Latency Issues

As previously described, a system's reaction time mainly depends on the selection of communication protocol it uses to transfer data across devices. Such reaction time can be categorized into two components: one is the data latency, and the other is the optimized estimation. The data latency accumulates when the data is being transmitted from one end to another, and deducted from the total reaction time. Latency interference can be classified in two ways. The first application interference can trigger the delay of readout values being sent back to the data receiver.

In our proposed system, the data receiver protocol is not designed for the real-time communication. The receiver script is being executed at a very low frequency in order to maintain its normal operation without affecting the SCADA. In consequence, the required data from the SCADA will be returned only at that low frequency. Hence, if the machine learning applications are not designed properly with that frequency correction, the data receiver will produce unforeseen results and may fail at some point. All those electrochemical sensors have readout operations on microseconds basis, and they have to wait for the data receiver which works mainly on minute basis. Therefore, the control is based on machine learning algorithm results actually from the data receiver triggering time way above the microseconds requirement.

To avoid this situation, it is inevitable to optimize the readout data frequency from the SCADA down to several seconds, which was successfully achieved with the SCADA design and implementation assistance from a specialist engineer. The second kind of data interference can occur actually within the machine learning algorithms, more specifically, during the training stage. It is closely related to how the considered input training data will impact the desired results from the models on its data arrays. With particular regards to our discussed machine learning algorithms, the required training data needs to satisfy the requirements for the ‘input-output’ structure, such that even an unrelated observer just looks at the output learning arrays, can predict the expected model values with high possibility. For instance, if an observer is watching rain accumulation data on the training input dataset, obviously, he/she will know that when the rain stops, flooding may still occur within a specific time for discharge, and vice versa – flooding duration will definitely not be equal to the rain duration.

8.2. Computational Resource Constraints

Most MLAST algorithms for RTFDD and PH in Solar Energy Systems are proposed as software only schemes, and their computational burden and performance efficiency are not considered. However, these algorithms are expected to be implemented in dedicated hardware with limited computational resources i.e. low memory space, number of computational cores, and/or low processing speed. RTFDD of a Solar PV Panel requires data from all the sensors for many trajectory samples. As a result, it generates heavy data traffic at the time of preparation of the fault detection model, during training. Other algorithms require continuous data exchange for ML and performance execution for PH. Therefore, along with the RTFDD algorithm, an effective and efficient scheme must be developed to handle the huge data in a limited duration and memory space. Particle swarm optimizer is used for initializing the weight matrix of CNN and GRNN methods for RTFDD in a smaller time duration and efficient manner. These models also must be optimized, reduced, and loaded in the limited memory space for efficient usage of HDL architecture for CNN and GRNN.

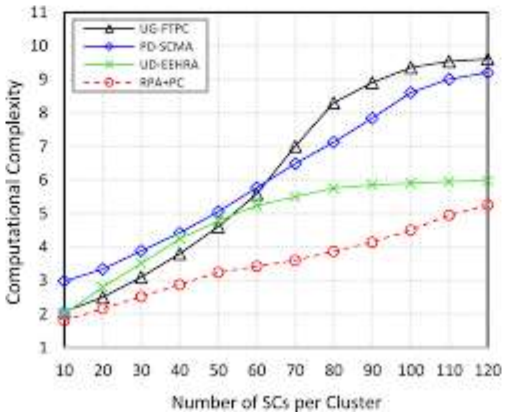

Fig : Heterogeneous Computational Resource Allocation

The model should also be designed for utilization of scalable and parallel manpower during the execution of the embedded system algorithm. From the perspective of design, the choice of control unit is dominant for the architecture. Most of the ML embedded engine systems use DSP and/or FPGAs as architectural control units. In most of the ML embedded engines, neural networks require the use of FPGAs, which is efficient in terms of chip area and power. The highly parallel nature of the processing element reduces the execution time significantly with an increase in power cost. When the order of execution of the activated parts is optimized, the number of uncontrolled looks makes optimization at a chip level difficult. These types of design tradeoffs are usually crucial during the course of development.

8.3. Integration with Existing Systems

Even when models are sufficiently fast and accurate, potential end users of these models often have real concerns about the ‘integrability’ of these algorithms into their existing management systems and processes. For forecasting or fault detection algorithms to be truly useful, they need to fit into the existing data pipeline of module performance monitoring and evaluation systems and business. An algorithm that generates monitoring alerts or performance forecasts in real time needs to communicate with the central data warehouse or dashboard for the user or manager to act on these alerts and take decisions. If there are multiple servers at different sites generating forwarded alerts or multiple forecasts, these localized alerts or forecasts need to get collated into a centralized system. Further, it is important to emphasize that the models are not purely predictive, but also add value to enhance the existing prediction capabilities of statistical models. Thus, in case of real-time forecasting, the models need to act either as wrappers for supplementary corrective modeling, seamlessly integrating with the existing overall forecasting framework and user interface or take the output from the existing business forecasting systems as input for postprocessing.

However, real-time fault prediction models can also coexist as entirely stand-alone systems, for both fault detection and performance forecasting components. End-user organizations may prefer to keep different systems completely isolated due to internal policies, security concerns, and risk mitigation strategies. Such a bifurcated system is not uncommon in real-time systems, and seamless data transfer from one system to the other, possible state-level alerts, and possible actions taken based on these alerts can be established. In this case, the integration and automation concerns reduce to purely passive modes for the fault detection and forecasting components.

9. Case Studies

In preceding sections, various machine algorithms are discussed for fault detection and performance enhancement in solar energy systems. The developed algorithms can assist in timely fault detection, making solar energy systems more reliable and efficient. In this section, the presented machine learning models are implemented to three case studies: residential, commercial, and large-scale solar systems. These presented case studies are useful in fault detection, performance evaluation, and performance enhancement of solar energy systems during the design and operational phases.

9.1. Case Study 1: Residential Solar Systems

The first case is conducted using a residential solar system located in Dallas, Texas. The developed fault detection and diagnosis systems can timely detect the fault and identify the fault type. This can optimize the inspection and repair process, and prevent the battery overcharge and the damage of inverter and PV panels, if any. The implemented models can evaluate the installed PV system performance ratio, and the prediction models can provide the expected system output. The prediction error and performance ratio are used to identify the potential fault. The evaluation and prediction results are discussed in detail in this section. The system contains twelve PV panels with a capacity of 50 Wp installed on the rooftop as a grid connected system. The system has two PV arrays, and each array has six PV modules connected in series. The inverter is used in the system. The available data set consists of weather data (ambient temperature, global horizontal solar irradiance, and wind speed) and system operating data (ambient temperature, module surface temperature, irradiance-dependent module voltage and current, inverter output voltage and current, and output power) collected at 5-minute intervals using a data acquisition system. The data set is very clean, and the missing data percentage is less than 4%.

9.1. Case Study 1: Residential Solar Systems

The residential power generation capacity is achieved through PV systems located on rooftops. Such systems are gaining popularity due to energy independence and lowering electricity bills. With many of these PV systems being grid-tied, the economics of sizing the PV systems are favorable, since excess power generated can be sent to the grid for a reward, and power not generated can be drawn from the grid. In this section, we present a residential 7 kW PV system which was installed in the summer of 2011 on the rooftop of a dwelling. The system, which consists of 28 panels with a monitor and optimizer, and an inverter cluster of micro-inverters, started normal operation after extensive testing and calibration phases in the summer of 2011. This PV system has been continuously monitored for experimental data, including inside and outside array temperatures, array and inverter output currents and voltages, ac power to the grid, and array plane of array incident solar radiation. This data has been collected in predefined and diverse operating modes metered during sunny days without any cloud cover, and further used in conjunction with existing clear-sky irradiance models to deduce correction factors to estimate hourly incident plane array solar radiation during cloudy and other non-clear sky conditions. A large set of these factors is then used in conjunction with local weather data to create a composite clear-sky solar radiation database of monthly median values.

The objective of the present case study is to analyze and model the performance of the 7 kW PV system under clear-sky and diverse weather conditions using data obtained from the PV system itself. Special attention is paid to anomalous performance dips observed on the short-term while recovering after heavy snow accumulation periods. Using the data correction factors, these PV system weather condition dependent performance models are developed in this case study, and then used to not only detect faults due to these major weather dependencies, but also detect alarm events when the performance behaves abnormally on the short-term due to other reasons. We therefore present a rather unique research thesis work in this section.

9.2. Case Study 2: Commercial Solar Installations

We demonstrate app installations in demanding environments to determine if they are practical in commercial scalability: a rooftop partially shaded installation with modules at different orientations, glass of different coloration and cleanness, and with multiple inverter connections to the grid: a system with NA. A PV-Field with partial shading resulting from trees spanning the roof held several modules at different orientations and different positions of cleanness because some debris have been blown across the site. Elements of various properties and colors are in close proximity. Planning tests required extensive effort to determine appropriate times when modules would exhibit simultaneous output voltages and currents and their MAPD and Kai metrics would be different from each other. The PV-Field had two digitally connected strings with six modules each, one string field potentiometer, and one string flow meter. The input signal was the AI 10 absorbed at the module in the center of each string.

For the tests we used a single PV-Field GUI collecting data, followed by visual inspection of the recorded output. Each GUI executed a prototype application with a separate destination for the AI 10. The GUI was structured as a top panel with command buttons for database queries, graph draws, and parameter file editing, and a lower panel with visuals for the current GUI operation. An input file was generated and image capture in one of the cameras was initiated. The photos stored in the module camera were automatically uploaded to the remote server. The installation produced scattered MAPD spikes associated with high photovoltaic module temperatures during the early afternoon of hot summer sunny days, as expected, as well as early mornings and moments of intermittent cloud cover. Full module shading on either string produced shallow spikes in the AI 10 associated with changes in the sensor temperatures, but no transitional spikes as with the module segments heater blocks, as expected. Data pattern inspection indicated sufficient Ai 10 accuracy for module status detection.

9.3. Case Study 3: Utility-Scale Solar Farms

Utility-scale solar photovoltaic (PV) energy generation systems harness solar energy on a large scale designed to produce electricity with minimal restrictions for transfer to the electric grid. Utility-scale PV facilities deploy thousands of PV modules and inverters, potentially costing millions to assess and manage over their lifecycle as proper O&M of these PV systems is critical thereby incurring high costs. As a result, PV farms could very well benefit from improved cleanness monitoring systems, especially given the large extent of the land areas typically involved often with varying payback periods depending on geographic location and climate as well as the vast diversity among PV system configurations and deployment modes. Automated property surveillance services could very well allow for remote monitoring of large expansive installations, thus enabling reduced costs and risk while potentially increasing utility-scale facility profits.

The primary objectives of this utility-scale solar portfolio case study are both fault detection and performance enhancement for solar PV systems located in Sunbelt states, with specific emphasis placed on remote monitoring of PV system field components. The work mainly analyzes an assortment of disparate utility-scale PV concepts -fixed angle, 1- and 2-axis tracking -usually in widely varying geographic locations; and deploys stationed imagery capture systems including terrestrial-based cameras and unmanned aerial vehicles at the PV sites to gather both weather data including air temperature and sky conditions along with visual data throughout the solar cycle to identify existing field anomalies. The data is then analyzed to reveal additional performance deficiencies beyond those detected from just regular array current monitoring including soiling, shadowing, module overheating and delaminating, PV interconnection problems, cable insulation degradation, and inverter faults.

10. Evaluation Metrics

Evaluation metrics serve as vital tools for gauging the efficacy of machine learning models, guiding their application in real-world scenarios. In the realm of fault detection, an erroneous prediction entails a latent defect in the system that may necessitate costly rectification. This prediction incurs operational expenses but may not yield any immediate returns. Consequently, we prioritize accuracy and precision scores. In the subsequent energy prediction chapter, we predict energy values with absolute average errors ranging from 0.067 kWh to 1.03 kWh. We consider these scores to be highly relevant as they are critical to proper operations.

The accuracy, precision, and sensitivity scores are well-established metrics used in the evaluation of classification tasks. The true positive (TP), true negative (TN), false positive (FP), and false negative (FN) scores, computed as follows, are essential for their estimation: TP is determined as the count of ground-truth class positive predictions that are made for positive samples; TN is determined as the count of ground-truth class zero predictions made for zero samples; FP is determined as the count of predictions for positive samples that are inaccurate, due to their correlation with the ground-truth zero class; FN is determined as the count of ground-truth class one predictions made for positive samples that are inaccurate. Utilizing these four scores, we compute the overall accuracy for each classification algorithm according to its definition as the ratio of all correct predictions to the total number of samples. We also compute the precision, defined as the ratio of correct TP predictions to the total TP predictions that belong to the faulty class, and the sensitivity (or recall), defined as the ratio of TP predictions to the total number of ground-truth samples that belong to the faulty class.

10.1. Accuracy and Precision

The experimental performance of any applied algorithm is dependent on the required inputs and objectives. The analysis presented here is independent of a specific problem as it could be equally applied to any classification algorithm. Presented here are common error metrics such as accuracy, precision, recall, confusion matrix, and Cost-Benefit Analysis to measure the successfully detected faults or rejected normal behavior for the five datasets. Accuracy considers all predictions equally whether they are positive or negative, as it is a set ratio of TP + TN to all overall predictions made: Accuracy = TP + TN / TP + TN + FP + FN. Here, TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative. However, in our research problem stated before, the negative predictions (i.e., no faults) consider a much larger majority class sample in comparison to the detected fault prediction (positive class). Such class imbalance can make accuracy misleading. Accuracy, in general, cannot determine well how well a model classify the minority positive class; because although many classes classified to be negative are correct: if some from the positive class are incorrectly classified to be negative, the accuracy can overlook that. It is common to use precision, recall as complementary metrics along with accuracy. Precision gives a sample ratio of correct positive predictions to the actual positive samples determined: Precision = TP / TP + FP.

10.2. Recall and F1 Score

Another metric to assess the performance of a model is Recall. The Recall is the measure of the proportion of actual positives that are correctly predicted as positive by a model. In other words, it tells how many of the actual positives the model captured. However, it is important to note that Recall considers only the positive class. It does not care or account for the model's performance with the negatives. In the case of the validation data used for this work, there are 672 anomalies out of 307885 data points. However, these 672 anomalies have small peaks in their physical features, meaning that anomalies could be easily missed by the model without affecting the overall performance of it. In situations like this one, we want to ensure maximum coverage of the anomalies, hence why Recall is an important metric to analyze.

F1 Score combines the precision and recall metrics into one single metric to evaluate the performance of a model. The problem with precision or recall is that they account for only one class. By balancing both metrics, F1 Score allows us to look at the positives and negatives together. In the case of the validation data used for this work, since the anomalies only represent less than 0.2% of the data, having high precision and low recall or low precision and high recall may lead to outputs that have no significant impact because they either contain just a few anomalies or too many.

10.3. Cost-Benefit Analysis

Classification methods focus on minimizing the error rates for determining the negative and positive class. For our research, the cost of negative class was determined as the cost incurred by the PV system when an anomalous condition is prevailing but is actually predicted as a normal when reaching to the negative class. If not diagnosed and found by the field engineers or the system operators, energy is wasted and ultimately it affects the degradation of the PV plant, influencing the results and strategic future approaches. The cost associated with this situation is therefore very high. Moreover, this situation can last for weeks (or years). For these reasons, a lower bound for the acceptable false negative rate, associated with the cost incurred per hour of energy production during an event when the state is actually an emergency alarm. This cost per hour can easily amount to several hundred dollars in the long term. In most applications, the false positive misclassification cost is low and therefore, the user can accept a lower classification accuracy on the positive class. This happens, for example, in the following cases: a misclassification errors, leads to further thermal cycling of the equipment that will eventually lead to increased costs for the user and, lastly, misclassification that then needs to be validated by technical staff.

In order to consider the possibility of introducing an external factor into the two-class classification model, we used the concept of Cost-Sensitive Modelling. This modelling transforms the classification problem, integrated into a cost-sensitive environment by allowing the knowledge of the total cost paid by the company for the false classifications to allow, in turn, the estimation of the flexible error thresholds. These thresholds will allow us to optimize the accuracy of a service for which the cost of false classification is the main element that can lead to an offering with a high utility.

11. Future Trends in Machine Learning for Solar Energy

Machine learning has made a significant impact on the solar energy industry, enabling a wide range of forward-looking applications, researching and developing new algorithms that solve practical business problems, and creating new products that utilize the technology in helpful ways. The biggest barrier to the adoption of machine learning in the solar industry is still primarily the availability of labeled or quality data sets. However, as data sets continue to grow and cloud providers are offering new, powerful and easy-to-use tools and platforms, algorithm development and refinement are probable to accelerate quickly. In the coming years, with the breadth of machine learning tools assumed to grow, there will likely be three main avenues of growth for practical applications of machine learning. First will be advancements in new models, algorithms, and approaches. These broadly useful improvements assume to make a broad variety of some tasks easier, faster, or more capable. The technology of solar energy has also advanced over the past few decades, showing a general trend toward reducing cost and enhancing performance. With that trend also, many manufacturing processes related to MPPT have evolved and improved. It is in this area of technology too that a new product relies on machine learning models that predict motion patterns and allows for a new, different, and optimal tracking strategy. Models that adaptively maintain the shadow of the sun on the focal point along arbitrary trajectories with minuscule computational needs are presented here, facilities addressed via these tools are enough to positively impact the long-term stability and balance of the plant, the development of this structure also can led to the profitable and profitable operation of the currently implemented systems.

11.1. Advancements in Algorithms

Solar energy remains remarkably under-exploited—while the generation cost has recently approached that of fossil fuels, the efficiency is far from its physical potential limits. To further deploy solar energy at a large scale it remains imperative to convince both regulators and customers, to overcome the three major hurdles of safety, reliability, and efficiency. In this chapter, we highlight the potential impact of advanced Machine Learning algorithms. We argue that recent advancements in Machine Learning create unprecedented and timely opportunities to deploy predictive and prescriptive analytics for Solar energy. These algorithms can predict several key operational characteristics (and their interdependencies) in real time—current, future performance, efficiency level, current and future safety, maintenance needs. By itself, predicting these quantities would only be the first step to utilize their potential. Indeed, their predictive power needs to be complemented with a whole set of powerful prescriptive analytic algorithms that can design customized action plans to actualize these predictions and provide effective data driven recommendations to users in their respective decision making processes.

We specifically illustrate the impact of three recent advancements in algorithms capabilities: adversarial learning to deal with the critical issue of imbalanced data problem, DDMs for predictive analytics, for temporal variation in time series (including inter-dependencies across modelled interrelated output time series), and prescriptive analytics thanks to the recently proposed Dynamic Weapon of Influence algorithm. We argue that the timely deployment of this set of algorithms can have a significant operational value. Specifically, we claim that they can enhance the capability to generate power, while improving demand matching and overall efficiency. Furthermore, they can greatly reduce safety hazards, while better anticipating maintenance needs appropriately, increasing the perceived quality and accelerating the development of the solar industry. Finally, recommendations provided by these algorithms can be easily operationalized.

11.2. Emerging Technologies

Researchers in different parts of the world are continuously searching for developing novel technologies for harnessing solar energy. In the years to come, the prediction for solar energy predicts that the colonization of planetary bodies will rely on space-based solar energy systems that will convert solar energy for generating power in orbit. The demand for electricity is expected to increase in the future due to increasing demand for digital information like communication, information processing, transportation, etc. Power generating devices will be located in the desert and the synthesized energy will get transported for fulfilling excess need. A technology has been developed for increasing the efficiency of solar cells. The photovoltaic cells are said to be about 14% efficient, which means they can convert about 14% of direct solar energy to electric energy. The technology being developed will help in increasing efficiency beyond 30% in concentrated photovoltaic. Hence it is expected that in the future, concentrated photovoltaic which increases the intensity of sunlight several times on a single cell, will become modular and commercially available.

The 3D Printing Technology has laid the foundation for the development of future solar technology comprising solar bio-harvesters. Combining solar energy utilization and bio-integration technology, solar bio-harvesters or biohybrid devices will be incorporated which are printed by 3D printing to develop dispositifs able to generate and export electric power, on-demand, while converting both organic and inorganic substrates from the surrounding environment therein. These ground-breaking technologies will also address future challenges in supporting new strategies for energy-efficient solutions, which can sustainably minimize the end-user costs, energy consumption, and environmental impact of energy systems. Further, the integration of wireless technology with solar energy technologies will be implemented in such a manner so as to devise solar photovoltaic systems and components to enable both photovoltaic energy generation and wireless communication in the millimeter wave frequency range in a fundamental and seamless manner.

12. Conclusion

"Solar energy systems are among the active areas of contemporary scientific and engineering research. Achieving efficiency and foolproof working of solar energy systems requires implementation of real-time fault detection and performance enhancement techniques based on novel methods. The current work presented several advanced machine learning techniques based solutions for fault detection and troubleshooting in photovoltaic panel and solar water heater systems. Specifically, the decision-oriented responsibility diagram structure helped reduce the existing knowledge on industrial photovoltaic panel systems fault knowledge. Then, the physics-oriented baseline comparison impeded the task of fault detection, classification, and severity scoring of the space-heating flat plate solar water heater system. The issue of aquatic solar still performance enhancement strategies focusing on structuring of latent heat of vaporization and efficiency enhancement factor relied on an artificial intelligence-based search for optimization. The sensorless control methodology was proposed as a novel alternative to miss the existing solar photovoltaic system maximum power tracking techniques, with special focus on short circuit methodology based tracking. Finally, fault detection techniques concerning industrial scale thermal energy storage tank system have been studied using the wavelet transform and trainable intelligent transformers for the first time."

References:

[1] Paleti, S., Singireddy, J., Dodda, A., Burugulla, J. K. R., & Challa, K. (2021). Innovative Financial Technologies: Strengthening Compliance, Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures. Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures (December 27, 2021).

[2] Gadi, A. L., Kannan, S., Nanan, B. P., Komaragiri, V. B., & Singireddy, S. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization. Universal Journal of Finance and Economics, 1(1), 87-100.

[3] Someshwar Mashetty. (2020). Affordable Housing Through Smart Mortgage Financing: Technology, Analytics, And Innovation. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 99–110. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11581.

[4] Sriram, H. K., ADUSUPALLI, B., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks.

[5] Chava, K., Chakilam, C., Suura, S. R., & Recharla, M. (2021). Advancing Healthcare Innovation in 2021: Integrating AI, Digital Health Technologies, and Precision Medicine for Improved Patient Outcomes. Global Journal of Medical Case Reports, 1(1), 29-41.

[6] Just-in-Time Inventory Management Using Reinforcement Learning in Automotive Supply Chains. (2021). International Journal of Engineering and Computer Science, 10(12), 25586-25605. https://doi.org/10.18535/ijecs.v10i12.4666

[7] Koppolu, H. K. R. (2021). Leveraging 5G Services for Next-Generation Telecom and Media Innovation. International Journal of Scientific Research and Modern Technology, 89–106. https://doi.org/10.38124/ijsrmt.v1i12.472

[8] Adusupalli, B., Singireddy, S., Sriram, H. K., Kaulwar, P. K., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks. Universal Journal of Finance and Economics, 1(1), 101-122.
[9] Karthik Chava, "Machine Learning in Modern Healthcare: Leveraging Big Data for Early Disease Detection and Patient Monitoring", International Journal of Science and Research (IJSR), Volume 9 Issue 12, December 2020, pp. 1899-1910, https://www.ijsr.net/getabstract.php?paperid=SR201212164722, DOI: https://www.doi.org/10.21275/SR201212164722
[10] AI-Based Financial Advisory Systems: Revolutionizing Personalized Investment Strategies. (2021). International Journal of Engineering and Computer Science, 10(12). https://doi.org/10.18535/ijecs.v10i12.4655
[11] Cloud Native Architecture for Scalable Fintech Applications with Real Time Payments. (2021). International Journal of Engineering and Computer Science, 10(12), 25501-25515. https://doi.org/10.18535/ijecs.v10i12.4654

[12] Innovations in Spinal Muscular Atrophy: From Gene Therapy to Disease-Modifying Treatments. (2021). International Journal of Engineering and Computer Science, 10(12), 25531-25551. https://doi.org/10.18535/ijecs.v10i12.4659
[13] Pallav Kumar Kaulwar. (2021). From Code to Counsel: Deep Learning and Data Engineering Synergy for Intelligent Tax Strategy Generation. Journal of International Crisis and Risk Communication Research , 1–20. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/2967
[14] Raviteja Meda. (2021). Machine Learning-Based Color Recommendation Engines for Enhanced Customer Personalization. Journal of International Crisis and Risk Communication Research , 124–140. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3018
[15] Nuka, S. T., Annapareddy, V. N., Koppolu, H. K. R., & Kannan, S. (2021). Advancements in Smart Medical and Industrial Devices: Enhancing Efficiency and Connectivity with High-Speed Telecom Networks. Open Journal of Medical Sciences, 1(1), 55-72.
[16] Chava, K., Chakilam, C., Suura, S. R., & Recharla, M. (2021). Advancing Healthcare Innovation in 2021: Integrating AI, Digital Health Technologies, and Precision Medicine for Improved Patient Outcomes. Global Journal of Medical Case Reports, 1(1), 29-41.
[17] Kannan, S., Gadi, A. L., Preethish Nanan, B., & Kommaragiri, V. B. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization.
[18] Implementing Infrastructure-as-Code for Telecom Networks: Challenges and Best Practices for Scalable Service Orchestration. (2021). International Journal of Engineering and Computer Science, 10(12), 25631-25650. https://doi.org/10.18535/ijecs.v10i12.4671
[19] Srinivasa Rao Challa. (2021). From Data to Decisions: Leveraging Machine Learning and Cloud Computing in Modern Wealth Management. Journal of International Crisis and Risk Communication Research , 102–123. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3017

[20] Paleti, S. (2021). Cognitive Core Banking: A Data-Engineered, AI-Infused Architecture for Proactive Risk Compliance Management. AI-Infused Architecture for Proactive Risk Compliance Management (December 21, 2021).
[21] Vamsee Pamisetty. (2020). Optimizing Tax Compliance and Fraud Prevention through Intelligent Systems: The Role of Technology in Public Finance Innovation. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 111–127. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11582
[22] Venkata Bhardwaj Komaragiri. (2021). Machine Learning Models for Predictive Maintenance and Performance Optimization in Telecom Infrastructure. Journal of International Crisis and Risk Communication Research , 141–167. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3019
[23] Transforming Renewable Energy and Educational Technologies Through AI, Machine Learning, Big Data Analytics, and Cloud-Based IT Integrations. (2021). International Journal of Engineering and Computer Science, 10(12), 25572-25585. https://doi.org/10.18535/ijecs.v10i12.4665
[24] Kommaragiri, V. B. (2021). Enhancing Telecom Security Through Big Data Analytics and Cloud-Based Threat Intelligence. Available at SSRN 5240140.
[25] Rao Suura, S. (2021). Personalized Health Care Decisions Powered By Big Data And Generative Artificial Intelligence In Genomic Diagnostics. Journal of Survey in Fisheries Sciences. https://doi.org/10.53555/sfs.v7i3.3558
[26] Data Engineering Architectures for Real-Time Quality Monitoring in Paint Production Lines. (2020). International Journal of Engineering and Computer Science, 9(12), 25289-25303. https://doi.org/10.18535/ijecs.v9i12.4587
[27] Mandala, V. (2018). From Reactive to Proactive: Employing AI and ML in Automotive Brakes and Parking Systems to Enhance Road Safety. International Journal of Science and Research (IJSR), 7(11), 1992-1996.