Volume 19, Number 1, 2025

Technical University of Cluj-Napoca
North University Centre of Baia Mare
Faculty of Engineering
Electrical, Electronic and Computer Engineering Department

Volume 19, Number 1, 2025

Carpathian Journal of Electrical Engineering

Carpathian Journal of Electrical Engineering

ISSN 1843 - 7583

UTPRESS PUBLISHER

TECHNICAL UNIVERSITY
OF CLUJ-NAPOCA

NORTH UNIVERSITY CENTRE OF BAIA MARE
FACULTY of ENGINEERING

ELECTRICAL, ELECTRONIC AND
COMPUTER ENGINEERING DEPARTMENT

Carpathian Journal of Electrical Engineering

Volume 19, Number 1, 2025

ISSN 1843 – 7583
http://cee.utcluj.ro/cjee/

CONTENTS

Received: January 2025, Accepted: April 2025, Published: May 2025
Digital Object Identifier: https://doi.org/10.34302/CJEE/DACO1267

ORIENTATION OF PHOTOVOLTAIC PANELS USING A PYRAMIDAL SENSOR

Paul **NISTOR**, Ioan **ORHA**
Technical University of Cluj-Napoca, Romania
paul.nistor@ieec.utcluj.ro, ioan.orha@ieec.utcluj.ro

Keywords: sensors, solar panels, renewable energy sources, solar power generation

Abstract: *The work presents the realization of a solar cell orientation system using a pyramidal sensor. At the base of this system is the pyramidal sensor that captures the luminous flux and transmits it in the form of information to the block responsible for signal processing. The main component within this block is the operational amplifier AO 741, which amplifies the signal before reaching the motor drive block. Within this block we will meet an H-bridge that has the role of changing the direction of the current in the motors, thus resulting in the change of the direction of rotation. In this way, going through all the blocks described, the signal resulting from the pyramid sensor leads to the finality of the movement, namely, the top of the pyramid is oriented towards the sun. When the light flux will fall from a different angle on the sensor, the process will start again, reorienting the pyramidal sensor.*

1. INTRODUCTION

Concerns about climate change, excessive dependence on fossil fuels and their negative impact on the environment have led humanity to intensify its research on renewable and alternative energies in recent decades [1].

Thus, renewable energy has increasingly become a priority as awareness of environmental issues and the need to diversify our energy sources has grown worldwide. This fact has led to a continuous development in the field of solar, wind, hydroelectric and other forms of renewable energy [2].

This work only concerns the field of solar energy, and the idea of producing such a system comes from an analysis of solar panels and their orientation systems existing on the market.

A general analysis made on the energy efficiency indicates that the solar panels with an orientation system following the sun are clearly superior to those with a fixed system, registering a significant increase in terms of the level of energy production [3]. Although such a system increases productivity considerably, it present higher costs, mainly due to the need for additional components and more complex software-based technology. Therefore, we decided to implement a simple, reliable and much cheaper system based exclusively on hardware automation, but which would ensure the same performance as a software-based system.

In order to understand how the device works, we will briefly explain the steps taken by the system in the process of orienting the solar panels. It all starts with the pyramidal sensor that captures the light flux and converts it into a signal, which is transmitted to a processing block where the signal will be interpreted. As a result of this interpretation, a command signal is generated which is transmitted to the drive unit of the motors. From this point the motors will be actuated and will lead to the repositioning of the solar panels and the pyramid sensor.

Observing how the system works we can see that the value of this system is found in its simplicity.

2. SYSTEM DESCRIPTION

2.1. Description of Functional Blocks

The system shown in *fig. 1* follows a perfectly linear course of operation without interference or external adjustments, therefore the device once switched on acquires full autonomy. Even the orientation movement of solar panels based specifically on the concept of action-reaction amplifies this autonomy. The system illustrated in the figure below works exclusively through automatic hardware and in the following paragraphs we will present the electrical components found at the base of each operational block.

The pyramid sensor consists of four photovoltaic cells placed on each face of the pyramid. Photovoltaic cells are grouped in two pairs and connected in parallel, when one pair is illuminated it will provide energy, therefore transmitting a signal while the other pair will remain inactive. We mention the fact that the electrical voltage resulting from the sensor in a direction of travel, depending on how the light flux falls on the pyramid can be positive or negative. If the light flux will fall on both pairs, they will produce energy simultaneously, and due to the fact that they are connected in parallel, they will cancel each other out.

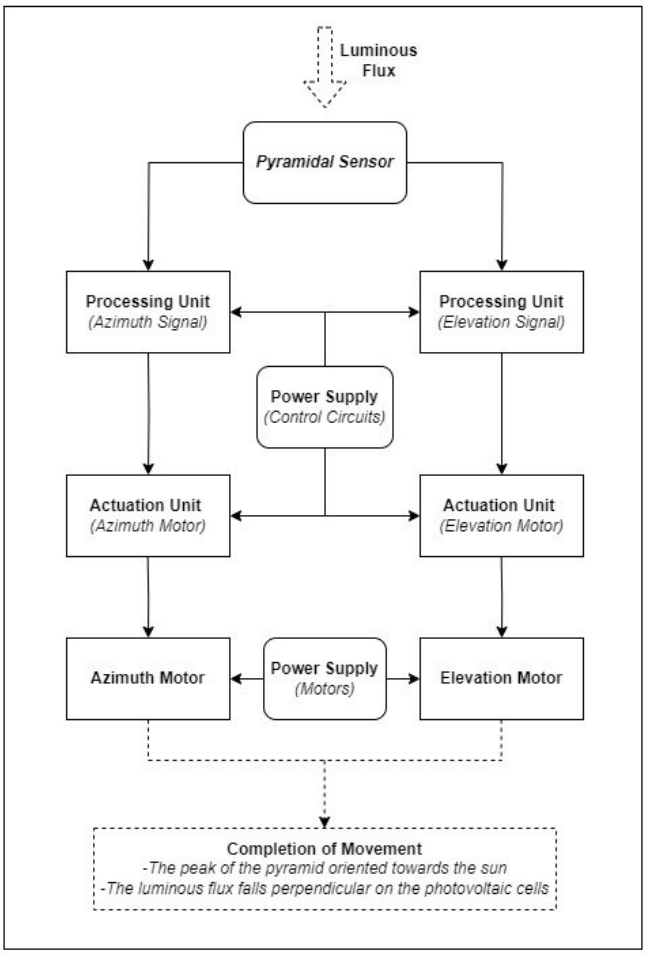

Fig. 1. System structure

It should be noted that this block responsible for capturing the light flux could be made in multiple ways, for example: the geometric shape of the sensor does not have to be a pyramid, it can even be a sphere. The light flux capture element itself does not need to be made with photovoltaic cells, if we wanted to minimize the project we could even use photoresistors.

The reason we used photovoltaic cells is due to the fact that they are built by design to follow the light spectrum that the photovoltaic panels were built for. An added advantage to using solar cells is that they provide easily processed voltages between 0-3V that can be easily filtered by optical and electrical interference.

Regarding the geometric shape used to implement the sensor, we chose the pyramid because it is ideal for capturing the light flux from all four cardinal points thus giving us the necessary signals for elevation and azimuth movement.

The next two blocks present in the system, as the name suggests, deal with the processing of the signal received from the sensor. These circuits are similar, with the only difference that one handles the horizontal movement and the other the vertical one, but the basis of both circuits is the AO741 in the amplifier assembly.

The AO741 circuit has adjustable amplification and can be modified from a semi-adjustable resistor. This feature will control the sensitivity of the orientation circuit and is adjusts at startup under normal lighting conditions.

The output of the operational amplifier drives a final stage in class B with load distributed in the collectors of the complementary transistors. Transistors being pure class B, they can never both conduct at the same time, having a voltage range at the output of the operational amplifier in which none of the transistors conducts. This creates an interval where the H-bridge is at rest, an event that occurs when opposite faces of the pyramid are equally illuminated. We mention that an RC high-pass filter is placed before the operational amplifier, it has the role of filtering the signal from optical interference.

The next blocks present in the system are those for driving the motors to obtain the two directions, namely azimuth and elevation. These blocks are also identical. Thus, each block contains an H-bridge, it is made with relays and is controlled by the previous block. The role of the H-bridge is to change the direction of rotation of the direct current motors, depending on the need to move the photovoltaic panel on which the pyramidal sensor is mounted.

The last two blocks present in the scheme represent the motors that make the movement of the solar panels and the pyramid sensor. These are direct current motors that change their direction of rotation depending on the change of polarity of the supply, they are motors with low consumption and with a simple actuation.

Once all the blocks have been presented, we reach the end of the scheme and also the end of the orientation process of the solar panel. At this point, the top of the pyramidal sensor will be oriented towards the sun and the luminous flux will fall perpendicular to the photovoltaic cells. When the position of the sun changes, the entire system orientation process will be restarted from the first step, going through all the blocks again and reorienting the solar panels.

2.2. Electrical Diagram of the System

Through the previously presented blocks, we can understand at least conceptually how the system works. But through the electrical scheme shown in *fig. 2*, we can observe and understand practically how it works.

Although at first look the wiring diagram may seem difficult to understand, it works and is structured exactly like the blocks already explained in the previous paragraphs. In the first part of the electrical diagram, we can observe how the signal reaches the operational amplifier whose amplification is adjusted with the help of potentiometer R3. The output of the amplifier drives the collectors of the two transistors which can never both conduct at once.

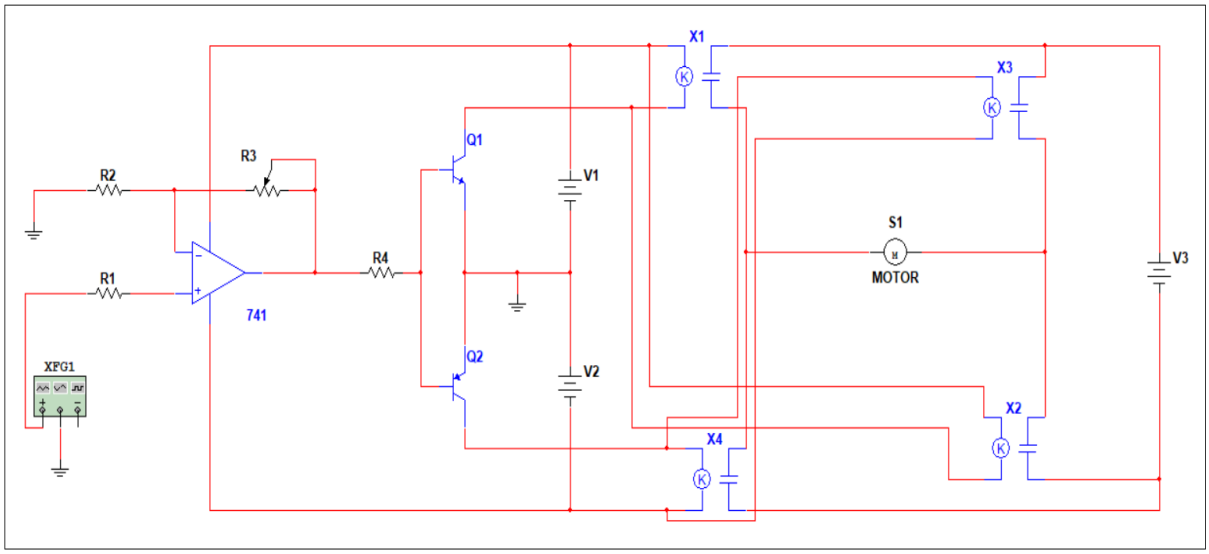

Fig. 2. Electrical diagram of the system

As can be seen in the diagram above, there are four relays that make up two groups. Each transistor is connected to a group of relays, and depending on the need, the required relay group will be activated by one of the transistors. In this way, the polarity of the motor supply is changed, leading to a change in its direction of rotation. Within the system there are two such electrical circuits, one for azimuth movement and the other for elevation movement. We mention the fact that from this electrical diagram itself we can notice the simplicity of the system.

3. SIMULATIONS AND RESULTS

In the following, we will present a comparison between a fixed solar panel and a solar panel with an orientation system. Based on these comparisons, it will be possible to see that a solar panel with an orientation system considerably increases the energy produced.

The reason this improvement in energy efficiency occurs is because an orientation system maximizes the solar panel's exposure to solar radiation over the course of a day. This fact is best highlighted on cloudy days or in diffuse light conditions, as the system will adjust the position of the solar panel to optimize the capture of the light flux. In contrast, this adaptability for such conditions is missing in the case of fixed solar panels.

To simulate the energy production of the photovoltaic system, we used the PVGIS (Photovoltaic Geographical Information System) platform [4]. The choice of the PVGIS platform was motivated by the fact that it provides real, long-term climate data, fully validated and used in photovoltaic research at European level.

3.1. Characteristics of the photovoltaic module used in the simulations

To perform the energy simulations, we used a monocrystalline photovoltaic module with a nominal power of 150 W, whose technical characteristics and electrical parameters under standard test conditions (STC) are presented in *fig. 3*.

Electrical Characteristics	
	STC
Maximum Power (Pmax)	150Wp
Maximum Power Voltage (Vmpp)	18.0V
Maximum Power Current (Impp)	8.33A
Open Circuit Voltage (Voc)	22.5V
Short Circuit Current (Isc)	9.00A
Power Tolerance(Positive)	5%
Module Efficiency STC	18.66%
Operating Temperature Range	-40℃ to +85℃
Maximum System Voltage	1000V
Series Fuse Rating	15A
Temperature Coefficient of Pmax	-0.40 %/℃
Temperature Coefficient of Voc	-0.30 %/℃
Temperature Coefficient of Isc	0.05 % / ℃
Nominal Operating Cell Temperature(NOCT)	45±2℃
Mechanical Characteristics	
Cell Type	Monocrystalline 156mm
Cell Number	36 (4x9)
Dimensions (mm)	670x1200x30
Weight(Kgs)	10.00
Front Glass	3.2 mm, High Transmission, Low Iron,Tempered Glass
Frame Type	Anodized Aluminium Alloy
Junction Box Protection Class	IP 67 Rated
Connector Type	MC4
Output Cables	2.5 mm^2, Length:900 mm

Fig. 3. Technical characteristics of the 150 W monocrystalline photovoltaic module [5]

The module has an efficiency at STC of 18.66%, a characteristic value for modern monocrystalline panels, which reflects the ratio between the incident light energy and the useful electrical energy obtained under ideal conditions. These parameters are essential for describing the electrical behavior of the module and serve as input data for modeling the

panel's performance both in a fixed configuration and in a mobile configuration with a tracking system.

Although the efficiency and STC parameters do not directly influence the energy simulation performed through PVGIS, their inclusion is necessary for the complete characterization of the module.

3.2. Simulation of the Fixed Photovoltaic System

The performance of the fixed photovoltaic system was evaluated using the PVGIS platform (version 5.3). The simulations were performed for the location of Baia Mare, Romania, characterized by the geographical coordinate latitude 47.656° N, longitude 23.595° E and an approximate altitude of 240 m. In the analysis, the influence of the local relief was taken into account by activating the automatic horizon calculation option (Use terrain shadows: calculated horizon).

The analyzed system consists of a single photovoltaic panel with a nominal power of 150 W (0.15 kWp), based on crystalline silicon technology. For the performance evaluation, the PVGIS-SARAH3 solar radiation database was used, and the total system losses were estimated at 14%, including losses due to temperature, cables, inverter and other operational factors.

The photovoltaic panel was considered to be mounted on an independent structure (free-standing), with a fixed mounting system. The tilt angle and the orientation towards the south were determined automatically by activating the slope and azimuth optimization option, thus ensuring operating conditions close to the optimal ones for the analyzed location. Based on these settings, the results regarding the energy production and the annual performance of the photovoltaic system were obtained, presented and analyzed in the following section.

Simulation outputs:	
Slope angle [°]:	37 (opt)
Azimuth angle [°]:	-2 (opt)
Yearly PV energy production [kWh]:	173.38
Yearly in-plane irradiation [kWh/m^2]:	1467.38
Year-to-year variability [kWh]:	7.38

Fig. 4. Main results for a fixed PV system

The results obtained in *fig. 4*. show that the analyzed system achieves an annual electricity production of 173.38 kWh, which corresponds to a specific efficiency of approximately 1155 kWh/kWp/year, a value representative for the areas of northern Romania. The annual solar irradiation on the panel plane was estimated at 1467.38 kWh/m², which confirms the existence of an adequate solar potential for the studied location. Following the

automatic optimization process, the optimal tilt angle of the panel was determined at "37°", with an azimuth very close to the southern direction "–2°".

Fig. 5. shows the monthly distribution of the electricity produced by the analyzed fixed photovoltaic system. A pronounced seasonal variation in energy production is observed, determined by the change in the level of solar irradiation throughout the year. The maximum production is recorded in the summer months, especially in the period June–August, when the production exceeds 20 kWh, while the minimum values occur in the cold season, reaching approximately 5.25 kWh in December. This behavior is characteristic of fixed-angle photovoltaic systems installed in temperate climate zones and confirms the direct dependence between the energy generated and the availability of the solar resource.

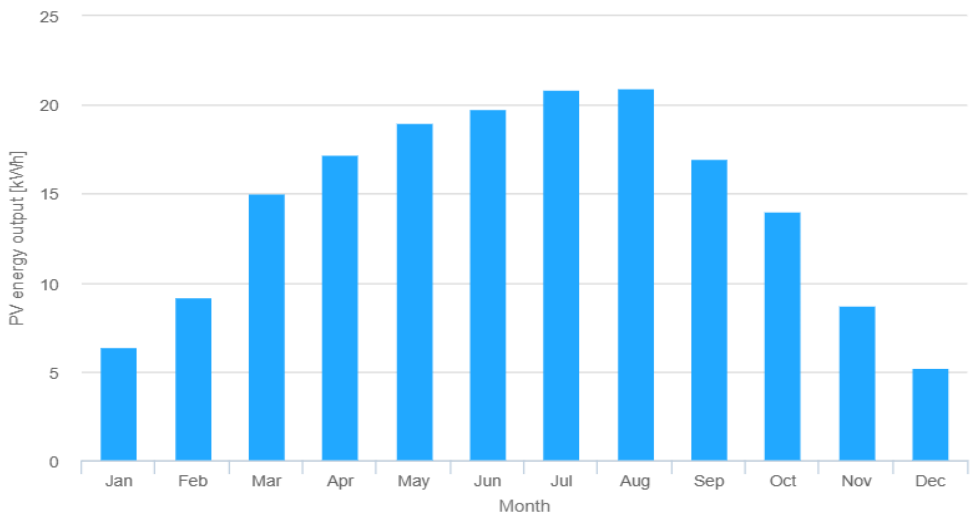

Fig. 5. Monthly energy output from fixed-angle PV system

3.3. Simulation of the Tracking Photovoltaic System

The performance of the tracking PV system (dual-axis tracking) was evaluated using the same simulation platform, PVGIS (version 5.3), to ensure direct comparability with the results obtained for the fixed PV system. The simulations were performed for the same location, keeping the fixed module options active.

The analyzed system consists of a single photovoltaic panel with a nominal power of 150 W (0.15 kWp), based on crystalline silicon technology, identical to that used in the fixed system. The same PVGIS-SARAH3 solar radiation dataset was utilized, while the total system losses were maintained at 14%. Thus, the only difference between the two analyzed configurations is represented by the orientation mechanism of the photovoltaic panel.

In this configuration, the photovoltaic panel was considered mounted on a dual-axis solar tracking system, which allows the continuous orientation of the photovoltaic panel surface in both azimuth and elevation directions, depending on the position of the sun

throughout the day. This mounting strategy ensures a favorable angle of incidence of solar radiation on the panel for a longer period of time compared to a fixed mounting system, leading to an increase in the captured energy.

Simulation outputs	Two axis
Slope angle [°]:	-
Yearly PV energy production [kWh]:	228.48
Yearly in-plane irradiation [kWh/m^2]:	1919.53
Year-to-year variability [kWh]:	11.2

Fig. 6. Main results for the tracking PV system

The results obtained in *fig. 6*. indicate an annual electricity production of 228.48 kWh, higher than that obtained in the case of the fixed installation system. This value corresponds to a specific efficiency of approximately 1523 kWh/kWp/year, highlighting the energy advantage of using the solar tracking mechanism. The annual solar irradiation on the panel plane was estimated at 1919.53 kWh/m², a significantly higher value than in the fixed system configuration, as a result of the continuous alignment of the panel with the position of the sun.

The monthly distribution of the electricity produced by the dual-axis tracking photovoltaic system is shown in *fig. 7.* The same seasonal variation characteristic of photovoltaic systems is observed, with maximum production values in the summer period, especially in the months of June–August, when the monthly production reaches approximately 29 kWh, and minimum values in the cold season, with a minimum of approximately 6.5 kWh in December. Compared to the fixed-mount system, the solar tracking configuration leads to an increase in monthly production throughout the year, the difference being more pronounced in periods of high solar irradiation.

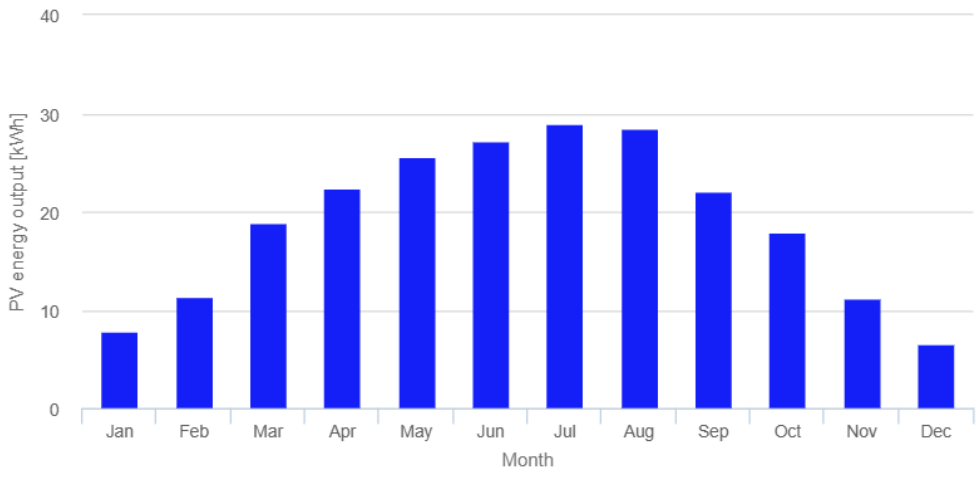

Fig. 7. Monthly energy output from tracking PV system

3.4. Comparison of Fixed and Tracking Photovoltaic Systems

Table 1. Monthly energy production comparison between fixed and tracking photovoltaic systems.

Month	Monthly energy production – fixed PV system (kWh)	Monthly energy production – tracking PV system (kWh)	Relative energy gain of tracking system (%)
January	6.38	7.82	22.57
February	9.22	11.44	24.07
March	15.00	18.91	26.07
April	17.22	22.34	29.75
May	18.98	25.58	34.77
June	19.76	27.22	37.76
July	20.84	29.04	39.35
August	20.98	28.49	35.79
September	16.98	22.03	29.74
October	14.01	17.91	27.84
November	8.76	11.14	27.17
December	5.25	6.55	24.76

In Table 1 presents the energy produced by the fixed photovoltaic system and the photovoltaic system equipped with a solar tracking mechanism. The comparison is performed for each month of the year under identical operating conditions. The fourth column presents the relative energy gain of the tracking system, expressed as the percentage increase in energy production compared to the fixed photovoltaic system.

The use of a dual-axis solar tracking system results in a significant increase in energy production throughout the entire year, with relative gains ranging from approximately 22% during the winter months to nearly 40% in summer, the highest improvements being observed between June and July. These results demonstrate that continuous adjustment of the panel orientation allows the system to maintain an optimal incidence angle, leading to enhanced performance, particularly under high solar irradiance conditions.

However, the table above illustrates an ideal case, in which only the gross energy production is considered, without taking into account the energy consumed by the tracking mechanism. In order to provide a more realistic assessment of the efficiency of the mobile system compared to the fixed configuration, the analysis was repeated by deducting the auxiliary energy consumption associated with the tracking system.

3.5. Comparison of Fixed and Tracking Photovoltaic Systems Considering the Energy Consumption of the Tracking Mechanism

It is important to specify that within the guidance system, the decision-making part, having a simple structure and being strictly implemented with hardware, generates low energy consumption. In contrast, the execution part, represented by the motors that adjust the position of the solar panel, requires a significant amount of energy. The tracking mechanism is considered at a conceptual level, without a specific motor implementation, and its energy consumption is estimated based on values reported in the literature for small-scale dual-axis tracking systems.

To ensure a conservative and realistic assessment, it was assumed that the auxiliary energy consumption of the tracking mechanism is equal to 5% of the total annual energy production of the tracking PV system. This annual auxiliary energy consumption represents the cumulative energy required for continuous orientation of the panels throughout the year. The estimated annual auxiliary energy consumption was then subtracted from the gross annual energy production obtained from the PVGIS simulations to determine the net energy yield of the tracking system.

Although the auxiliary energy consumption of the tracking mechanism was estimated annually (5% of the total annual energy production), a monthly comparison of net energy was required. Therefore, the annual auxiliary consumption was distributed proportionally over the months, resulting in a uniform 5% reduction in the monthly tracking energy values.

Table 2. Monthly net energy production comparison between fixed and tracking photovoltaic systems, including auxiliary energy consumption

Month	Monthly energy production – fixed PV system (kWh)	Monthly net energy production – tracking PV system (kWh)	Relative net energy gain of tracking system (%)
January	6.38	7.43	16.46
February	9.22	10.87	17.89
March	15.00	17.96	19.73
April	17.22	21.22	23.23
May	18.98	24.30	28.03
June	19.76	25.86	30.86
July	20.84	27.59	32.39
August	20.98	27.07	29.03
September	16.98	20.93	23.26
October	14.01	17.01	21.41
November	8.76	10.58	20.78
December	5.25	6.22	18.47

As shown in Table 2, after accounting for the auxiliary energy consumption of the tracking mechanism, the net annual energy production is reduced to approximately 217.04 kWh. Despite this reduction, the system achieves a specific net yield of about 1447 kWh/kWp/year, which clearly demonstrates the energy advantage of the solar tracking mechanism.

Although the relative energy gain of the tracking system is reduced after accounting for the auxiliary energy consumption, the results demonstrate that a significant net advantage is preserved throughout the entire year. For the small-scale system considered in this study (150 W), the absolute energy gain remains modest, which is expected given the low installed power.

However, it should be noted that the energy consumption of the tracking mechanism does not scale linearly with the installed photovoltaic power. Consequently, the relative benefit of solar tracking systems becomes increasingly relevant for medium and large-scale photovoltaic installations. In the present study, the proposed hardware-based tracking system is demonstrated for a small-scale application, serving as a proof of concept.

In *fig. 8*, we can see our solar panel orientation system. Although this is only a prototype, it is fully functional and successfully demonstrates the possibility of making the current guidance systems more economical and energy efficient using only hardware automation.

Therefore, the comparison between fixed and tracking installations demonstrates that tracking systems achieve a higher energy output, confirming their effectiveness in improving photovoltaic energy production.

Fig. 8. Actual photograph of the experimental prototype

4. CONCLUSION

This study presents the design and experimental validation of a photovoltaic panel orientation system implemented exclusively using hardware components. Unlike most commercial solar tracker solutions, which rely on microcontrollers, embedded software and digital signal processing, the proposed system achieves full autonomy through simple analog electronic circuits, resulting in a robust and low-complexity architecture that minimizes cost and potential failure points and removes the need for software programming, continuous calibration and maintenance.

From an economic perspective, the reduction in system complexity translates directly into lower implementation and maintenance costs. The use of widely available analog components, such as operational amplifiers, discrete transistors, and relays, makes the system very attractive for applications requiring low cost. In addition, the absence of digital control units reduces both the initial investment and long-term operating expenses, providing a competitive alternative to commercially available tracking solutions, especially in applications where simplicity and reliability are prioritized.

Another important advantage of the proposed system is the low auxiliary energy consumption at the control level. Unlike many tracking systems available on the market, which require continuous processing and permanently active control logic, the presented solution has low energy consumption at the control circuit level.

The system also shows a high level of flexibility and adaptability. Thanks to its modular design, individual components can be easily replaced or adjusted, including the sensor geometry and the light detection elements. In addition, the adjustable sensitivity of the guidance circuit allows the system to operate effectively under different ambient lighting conditions without requiring a complete redesign. This level of flexibility is rarely found in commercial tracking systems, which are difficult to customize.

In addition to the economic and structural advantages, the proposed solution demonstrates that efficient solar tracking can be achieved without sacrificing performance. Although the absolute energy gain remains modest for small-scale installations, the system serves as a functional proof of concept, validating the feasibility of hardware-based tracking solutions.

In conclusion, the proposed solar tracking system represents a simple, cost-effective and energy-efficient alternative to commercial software-based solutions. By focusing on simplicity, autonomy and low energy consumption, the solution demonstrates that reliable orientation of photovoltaic panels can be achieved with minimal resources, while providing a solid starting point for future improvements and practical applications.

REFERENCES

[1] Z. A. Smith, K. D. Taylor, Renewable and Alternative Energy Resources, ABC-CLIO, August 2008.

[2] M. Koussa, A. Cheknane, S. Hadji, M. Haddadi, S. Noureddine, *Measured and modelled improvement in solar energy yield from flat plate photovoltaic systems utilizing different tracking systems and under a range of environmental conditions*, Applied Energy, Vol. 88, No. 5, pp. 1756-1771, May 2011.

[3] N. A. Kablar, *Renewable Energy: Wind Turbines, Solar Cells, Small Hydroelectric Plants, Biomass, and Geothermal Sources of Energy*, Journal of Energy and Power Engineering, Vol. 13, pp. 162-172, 2019.

[4] Photovoltaic Geographical Information System (PVGIS). Available: https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis_en

[5] Monocrystalline Solar Panel 150W Datasheet. Available: https://solarfeeds-media.s3.ca-central-1.amazonaws.com/wp-content/uploads/2022/05/19001109/190520221652937068.pdf

[6] J. Rizk, Y. Chaiko, *Solar Tracking System: More Efficient Use of Solar Panels*, World Academy of Science, Engineering and Technology, Vol. 17, 2008.

[7] F. A. Khalil, M. Asif, S. Anwar, S. Haq, F. Illahi, *Solar Tracking Techniques and Implementation in Photovoltaic Power Plants: a Review*, Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences, Vol. 54, No. 3, pp. 231–241, 2017.

[8] P. E. Logan, B. W. Raichle, *Performance Comparison of Fixed, Single, and Dual Axis Tracking Systems for Small Photovoltaic Systems with Measured Direct Beam Fraction*, Department of Technology and Environmental Design, Appalachian State University, NC 28608

[9] A. Awasthi, A. K. Shukla, M. Manohar, C. Dondariya, K. N. Shukla, D. Porwal, G. Richhariya, *Review on sun tracking technology in solar PV system*, Energy Reports, Vol. 6, pp. 392-405, November 2020.

[10] A. Saleem, F. Rashid, K. Mehmood, *The Efficiency of Solar PV System*, Proceedings of the 2nd International Multi-Disciplinary Conference, Decembrie 2016.

[11] H. H. Pourasl, R. V. Barenji, V. M. Khojastehnezhad, *Solar energy status in the world: A comprehensive review*, Energy Reports, Vol. 10, pp. 3474-3493, November 2023.

[12] P. Sati, A. Kumari, S. Kumar, *Solar energy advancements and their environmental impacts*, Futuristic Trends in Renewable & Sustainable Energy, Vol. 2, March 2023.

Received: May 2025, Accepted: November 2025, Published: December 2025
Digital Object Identifier: https://doi.org/10.34302/CJEE/ECCK7473

SELECTING SWITCHES FOR SWITCHING INDUCTIVE LOADS: TECHNICAL ASPECTS AND PRACTICAL CONSIDERATIONS

Carlos Manuel **RUANO GONZÁLEZ**
Department of Electrical Engineering, Electromechanical Faculty, University of Camaguey, Cuba
cmruanog@gmail.com

Keywords: Switches of half tension, inductive loads, surges, electric arches

Abstract*: In power systems, reactive compensation is a common and necessary practice to maintain the quality of electrical service. Disconnection of reactors or reactor banks is frequent, sometimes up to two or three times a day. Therefore, circuit breakers for this application must operate satisfactorily under the energization and de-energization processes of "inductive currents" that can cause the chopping interruption phenomenon, as well as in cases of magnetizing currents from unloaded transformers and load currents from induction motors. The proper selection of circuit breakers to operate inductive loads in medium-voltage systems is a critical aspect in the design and operation of electrical networks. Inductive loads, such as motors, transformers, and reactors, present unique challenges due to the voltage and current transients associated with their connection and disconnection. This article addresses the essential technical criteria for breaker selection, including interrupting capacity, surge resistance, and arc flash management. In addition, relevant international regulations are discussed, and case studies illustrating industry's best practices are presented. The aim is to provide comprehensive guidance for engineers and designers seeking to optimize the reliability and safety of their electrical systems.*

1. INTRODUCTION

Reactive compensation in electrical power systems is vital for maintaining service quality and operational efficiency. It is commonly implemented using reactors or reactor banks that are connected and disconnected via medium-voltage circuit breakers. These breakers must withstand the significant transient events that occur during such switching operations.

Therefore, the correct selection of these devices is paramount, with special attention to the arc extinguishing medium, to prevent premature failures that compromise system reliability.

Inductive loads, such as motors, transformers, and reactors, store energy in their magnetic fields. Upon disconnection, this stored energy is released, generating voltage and current transients that can cause severe overvoltages [1, 2]. These phenomena, including inrush currents and electric arc generation, can compromise equipment integrity and reduce its operational lifespan [3]. Understanding these characteristics is essential for selecting a circuit breaker capable of handling these demanding conditions without failure.

This article presents a case study from a 220 kV substation where a reactor connected to a 34.5 kV bus, fed by the tertiary winding of a 230/125/34.5 kV autotransformer, was protected by several circuit breakers that failed prematurely. The reactor's nameplate data is shown in Table 1.

Table 1 Reactor data

Guy	PTD-20000/35T
Strain	34.5 kV
Current	209 A
Frequency	60 Hz
Ability	11800 kVA

Source: Chapa data

Initially, an air circuit breaker (installed in 1974) was used for protection and switching. It suffered severe damage, destroying one pole before its expected end of life. It was replaced by an oil circuit breaker, which experienced similar contact damage and was also replaced prematurely.

On June 12, 2013, a 36 GI-E25 type SF6 switch was installed with the following nominal data:

- ✓ Vn = 36 kV
- ✓ In = 630 A
- ✓ Interrupting Capacity = 25 kA

By June 2016, only three years into 20 years expected lifespan and after only 1,802 of a rated 10,000 operations the contacts required replacement due to excessive resistance. This recurrent failure pattern prompted a detailed investigation into the root causes, involving a review of testing protocols, operational records, and relevant literature.

This led us to begin a review and testing process to determine the possible cause of the failures in the aforementioned switches, consulting bibliographies and specialists in the field.

1.1. Brief theory on interruption of inductive currents

In a high-voltage system, reactors are used to compensate for reactive currents in the system. These reactors are connected to the tertiary delta winding of the autotransformers, using medium-voltage switches, generally located in high-voltage substations. These reactors or reactor banks are sometimes operated up to two or three times a day, so the switches must operate satisfactorily. Proper switch selection is of great importance to minimize the probability of failure.

The switching of reactors involves the interruption of small inductive currents, which can lead to the current chopping phenomenon and multiple reignitions during the interruption process [4]. These events generate high-magnitude, high-frequency transient overvoltages, formally known as Transient Recovery Voltage (TRV).

Also evident in the connection and disconnection of magnetizing currents of transformers without load, load currents of induction motors and load currents of transformers that feed shunt reactors [4].

When an inductive load is disconnected, the current lags the voltage by approximately 90°. As the contacts separate, an arc is established. At the first current zero crossing, if the dielectric strength between the contacts cannot withstand the rising recovery voltage, the arc reignites. This process can repeat, with the recovery voltage increasing at each successive zero crossing until the arc is finally extinguished. At that moment, the energy stored in the circuit's inductance (L) resonates with its inherent capacitance (C), generating a high-frequency oscillatory transient. The natural frequency (f_n) of this oscillation is given by:

$$fn = 1/2\Pi\sqrt{LC} \tag{1}$$

For these applications, SF6 circuit breakers are more limited in their ability to withstand sharp increases in the transient reset voltage (TRV), which is the effective value of the peak voltage of the first half-wave of the alternating current component that appears between the circuit breaker contacts after the current has been extinguished. It has a very significant influence on the circuit breaker's opening capacity and has a frequency of the order of thousands of Hertz, depending on the electrical parameters of the system in the operating zone. This voltage has two components: one at the system's nominal frequency and the other, superimposed, oscillating at the system's natural frequency than vacuum circuit breakers with similar nominal characteristics. Therefore, for applications where the use of vacuum circuit breakers is possible, selecting them is the best option.

Calculating the Transient Recovery Voltage (TRV) is a very cumbersome and illogical process. However, it is important to take into account the peak value of the Transient Recovery Voltage (TTR), which ranges approximately three times the phase-to-ground voltage, with a very rapid rate of rise (Rate of Rise) of kV/μs. This parameter is more capable of withstanding vacuum interrupters than SF6 interrupters, as they interrupt the current at its first zero crossing.

There are some solutions to reduce the growth rates of the Transient Reset Voltage TTR such as the installation of capacitors, these prevent the switch from being overstressed specifically during the thermal recovery period that occurs during the first 2 μs after the current interruption, this solution has the disadvantage of cost and added complexity, in case of having to use large capacities the ferroresonance phenomenon may occur. Another solution is the opening and closing synchronized with the oscillations of currents and voltages or the use of a synchronizing relay that enables the switch to operate in a synchronized manner with the phase voltage of the system, *figure 1* shows the interruption of inductive currents in the plane.

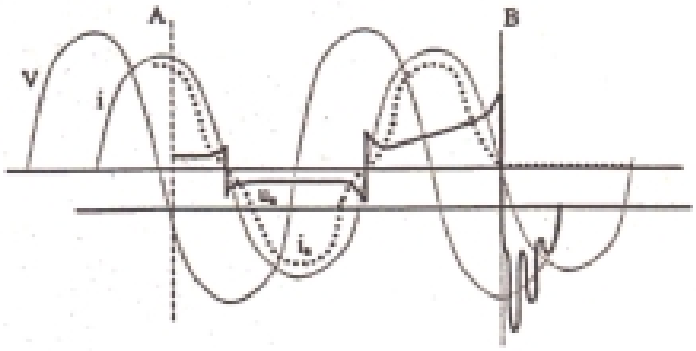

Fig. 1. Interruption of inductive currents.

2. METHODOLOGY

In this work the systemic method was used. To establish the relationship between the factors involved in the wear and tear or aging of the switches assigned to the connection and disconnection of inductive loads, through their testing protocols and the life records established by the company and by the statistical method for obtaining the behavior patterns of the state variables obtained from the documented records of their operation.

2.1. Review and testing process for the installed SF6 switch

During tests on the switch, a considerable deviation in contact resistance from the standard value of ≤ 100 μΩ for switches in service was observed. Measurements were taken during maintenance in 2016, and their values are shown in table 2.

Table 2: Results obtained in the tests carried out on the switch

Test I	Phase A	Phase B	Phase C
50 A	580 μΩ	82 μΩ	82 μΩ
100 A	468 μΩ	79 μΩ	82 μΩ
600 A	313 μΩ	81 μΩ	84 μΩ
100 A	232 μΩ	76 μΩ	91 μΩ

Source: [1]

According to the inquiries and bibliographical studies, the main causes of contact deterioration may be due to SF6 decomposition, such as oxidation and corrosion, contact wear, or pressure loss between contacts.

Several opening and closing operations were performed, seeking cleaning of the contacts and elimination of contamination if any, testing the contact resistance at 600 A where it can be used up to the nominal current, although deteriorated contact resistance values persist as shown in table 3.

Table 3: Contact resistance values after several opening and closing operations

Test I	Phase A	Phase B	Phase C
50 A	258 μΩ	70 μΩ	80 μΩ
100 A	269 μΩ	72 μΩ	74 μΩ
600 A	257 μΩ	72 μΩ	74 μΩ
100 A	266 μΩ	71 μΩ	74 μΩ

Source: [1]

The manufacturer's catalogue [2] establishes the contact resistance value in new switches as ≤50 μΩ, taking into consideration the latest values with 100A, we have that in phase A the contact resistance increases by 432% taking 50 μΩ as a base, in phase B it increases by 42% and in phase C by 48%, which shows serious problems in the contacts of the cutting chambers.

The manufacturer [2] proposes monitoring the wear of arc contacts and provides reference values:

- ✓ New arc contacts φ = 3.2°
- ✓ Arcing contacts with maximum permissible wear φ = 6.4°

To corroborate the above, the wear of the arc contacts is monitored, obtaining the following results table 4.

Table 4: Results of arc contact wear control

Switch phases	Poles of the failed switch (φ)
A	29°
B	27.5°
C	25°

Source: Chapa data

In the analysis of the contact resistance measurement results and in the arcing contacts wear control, the results show considerable wear on the main contacts and on the arcing contacts, the phase A the most damaged of the phases, which coincides with the phase that first opens, thus demonstrating the theory of the first pole to open.

Oscillographic tests were performed and all results were within standard, with no deterioration detected. The mechanical adjustment, i.e., the transmission regulation [5], was also reviewed, and it was verified that the value between points 1 and 2 coincides with the manufacturer's value of $D = 2$ mm, *fig. 2*.

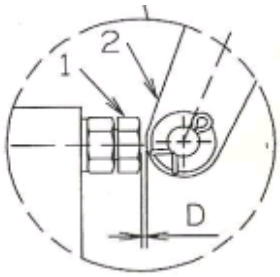

Fig. 2. Transmission regulation.

3. RESULTS

3.1. Analysis of measurement results

The diagnosis and monitoring of the condition of equipment and systems in real time comes from the 90s, particularly in the United States and Great Britain where the deregulation of the electricity market led to the search for new ways to reduce the costs of failures and increase reliability levels, in our case this monitoring and diagnosis is impossible and what is done is preventive maintenance that by procedure [6] is generally done once a year, bringing with it the detection of values that are sampled outside the parameter is late or in the worst case the failure surprises us.

The analysis of the contact resistance measurement results and the arcing contact wear monitoring in the SF6 circuit breaker show considerable wear on the main contacts and the arcing contacts, with phase A being the most damaged phase. This coincides with the phase that opens first, thus supporting the first-pole-to-open theory.

From the experience in the exploitation and the bibliographic analysis carried out [7] , the topic of selection of switches is of great importance, to avoid failures in the switches especially in the case at hand, the connection and disconnection of reactors or reactor banks, a process associated with the interruption of inductive currents that generate transient overvoltages of great magnitude with large frequencies [9] . Like the switches that are going to be installed near a generator, where the effects of the transient component of alternating current AC must be taken into account, analysis that is not the objective of the study [10].

Both practical and theoretical analysis have shown that the use of SF6 circuit breakers up to voltages of 34.5 kV is limited in the case of protecting reactors or reactor banks. Premature failures occur, shortening the circuit breaker's useful life. Although the literature proposes

various solutions for these cases to reduce the rate of rise of the Transient Recovery Voltage (TRV), all of which entail additional costs in terms of equipment and system configuration.

Therefore, for the connection and disconnection of reactors or reactor banks, up to a voltage level of 38 kV, vacuum circuit breakers [11] must be used, since they have a greater capacity to withstand strong increases in the growth of the Transient Restoration Voltage TTR. For voltages greater than 38 kV, SF6 circuit breakers are recommended, due to the limitation that vacuum circuit breakers have with the chambers for those voltage levels [12]. But since SF6 circuit breakers are assigned, and the procedures applied are not adequate for this type of maneuvers, this procedure was proposed in the case study, to avoid catastrophic failures.

3.2. Preventive Maintenance Tasks for Switches that Operate Inductive Loads in Substations

Therefore, a maintenance plan different from that designed for conventional switches was implemented [13], since the degradation of the contacts has an exponential behavior $f(x) = e^{x}$.

Therefore, the following set of actions is carried out to guarantee the type of maintenance necessary for the optimal service of these switches that operate inductive loads [14], table 3.

Table 3. Preventive Maintenance Tasks for Switches that operate inductive loads in substations

Part to inspect.	That verify.	Periodicity of the intervention.
Poles (Gas).	- Pressure of the gas, nominal value. - Quality of the gas.	One year after commissioning, and then twice a year, until year four and then every three months thereafter.
Insulators/Poles.	-Absence of dirt on the surface.	Dependent on environmental conditions.
Connection of the primary terminals.	- Absence of corrosion. - Bolt torque/tightness. - Presence of fat on the unions.	One year after commissioning and once per year thereafter.
Grounding.	- Absence of corrosion on grounding borders. - Tightness of connections. - Presence of grease on unions.	One year after commissioning and once per year thereafter.
External transmission mechanics.	- Absence of dirt/corrosion on friction points. - Proper lubrication (grease) on moving parts	One year after commissioning and once per year thereafter.
External structure.	- Absence of corrosion. - Absence of loose screws.	One year after commissioning and every five years thereafter

Static and dynamics contact resistance.	- Static and dynamic contact resistance values. - Compare against factory/normative standards.	One year after commissioning and then twice a year until four years and then every three months thereafter.
Arcing Contacts.	- Measurement of contact wear/erosion. - Verification against manufacturer limits.	One year after commissioning and then twice a year until four years and then every three months thereafter.
Oscillograph	- Find out simultaneousness of contacts	One year after commissioning and then twice a year until four years and then every three months thereafter.

4. CONCLUSIONS

1. The analysis of a failed SF_6 circuit breaker used for switching a 34.5 kV reactor confirmed that the primary failure mechanism was accelerated contact degradation caused by high Transient Recovery Voltage (TRV) and its high rate of rise (RRRV) associated with inductive current interruption.
2. Standard preventive maintenance cycles, typically annual, are inadequate for breakers performing frequent inductive load switching. The deterioration follows an exponential trend, necessitating a condition-based maintenance approach with significantly increased monitoring frequency, as proposed in this article.
3. For medium-voltage applications up to 38 kV involving the switching of reactors, capacitor banks, or unloaded transformers, vacuum circuit breakers are technically superior to SF_6 breakers due to their higher TRV withstand capability and interruption characteristics.
4. Implementing a tailored maintenance plan focused on frequent contact resistance measurement and visual inspection of contact wear is essential to prevent unexpected failures, ensure reliability, and optimize the total cost of ownership for circuit breakers in this demanding service.

REFERENCES

[1] W. Osorio and C.A. Culma, *Manual for the operation of electrical substations with voltage levels 115 kV, 33 kV and 13.2 kV*, Unpublished report, Technological University of Pereira, 2017. https://repositorio.utp.edu.co/server/api/core/bitstreams/45d3d748-ba03-4ca9-8c45-a405093484f7/content .

[2] I. Gondres and S. Lajes, *Partial maintenance management model for power circuit breakers using artificial intelligence*, Revista Chilena, Chile, vol. 26, no. 3, pp. 391-397, 2018.

[3] E. F. Álvarez, *Maintenance Management, Lean Maintenance and TPM, Master's thesis*, University of Oviedo. 2018.

[4] F. A Pérez, *General concepts in industrial maintenance management*, Unpublished report, Santo Tomas University, pp. 25-27, 2021. https://repository.usta.edu.co/bitstream/handle/11634/33276/9789588477923.pdf?sequence=4&isAllowed=y .

[5] S. García, *Corrective maintenance. Organization and management of breakdown repair*, Madrid, Spain, RENOVETEC. 2009.

[6] R. A Marrero and J. A. Vilalta, *Maintenance diagnosis-planning and control model*, Industrial Engineering, Havana, Cuba, vol. 40, no. 2, pp. 148-160,2019.

[7] L. R. Jaramillo and L.B. Miño, *Verification and maintenance manual for the Santo Domingo CC electrical substation*, Unpublished report, University of Cuenca, pp 15-20, 2018. http://192.188.48.14/handle/123456789/40?offset=80 .

[8] H. S Castelo, *Total production maintenance management model and its impact on operational performance in the area of extrusion of animal feed*, Unpublished report, Technical University of Ambato, 2017. https://repositorio.uta.edu.ec/bitstream/123456789/26702/1/Tesis_%20t1331mgo.pdf .

[9] L. P. Alcántara, *Notes on Electrical Substations*, Mexico, Underground Editions, pp. 155-156, 2015.

[10] J. H. Cuascuas, *Simulation of gas-powered circuit breakers in ATP Draw*, Unpublished report, National Polytechnic Institute, 2010. https://tesis.ipn.mx/jspui/bitstream/123456789/7182/1/4.pdf

[11] J. H. Cuascuas, Architecture and assembly of substations insulated with sulfur hexafluoride, Mexico City, Mexico: National Polytechnic Institute, pp. 85-105, 2008.

[12] J. P. Vasco, *Design of an automatic reconnection scheme for primary feeders in seven typical substations of the Quito Electric Company SA*, Unpublished report, Salesian Polytechnic University, 2015. https://dspace.ups.edu.ec/bitstream/123456789/8067/6/UPS-KT00957.pdf .

[13] F. H. Núñez, Generating Plants and Electrical Substations, Santo Domingo, Dominican Republic, Power Generation Plants and Substations. 2015.

[14] J. M. Prieto Montill, *Control Circuits for Cutting Devices in Electrical Substations*, University of Seville, 2017. https://biblus.us.es/bibing/proyectos/abreproy/91143/fichero/Primera+revisi%F3n+%28Recuperado+autom%E1ticamente%29.pdf .

[15] D.A. Peña, *Study of procedures for the maintenance of SF6 insulated switches in the EDC distribution equipment workshop, located in the Chacao service center*, Unpublished report, Central University of Venezuela, pp. 35-75, 2011. http://saber.ucv.ve/bitstream/10872/14125/1/TESIS.pdf .

[16] J. N . Morales, Electromagnetic Switches, pp 95-115, 2023. https://es.scribd.com/document/617708300/INTERRUPTORES-ELECTROMAGNETICOS .

[17] Q. Hu, L. Zhang, S. An, D. Zhang, *On Robust Fuzzy Rough Set Models,* IEEE Transactions on Fuzzy Systems, vol. 20, no. 4, pp. 636-651, 2012.

Received: September 2025, Accepted: November 2025, Published: December 2025
Digital Object Identifier: https://doi.org/10.34302/CJEE/BUEP2247

REGRESSION-BASED PREDICTION OF ANXIETY SEVERITY

Bogdan **CHIȘ**, Diana **DULF**, Mădălina **MORAR,** Eleonora **POP**
Technical University of Cluj-Napoca
chis.do.paul@student.utcluj.ro, dulf.da.diana@student.utcluj.ro,
jula.an.madalina@student.utcluj.ro, eleonora.pop@ieec.utcluj.ro

Keywords: Mental health, Anxiety disorders, Machine learning, Regression analysis

Abstract*: The present study explores the use of machine learning to predict self-reported anxiety levels based on demographic, behavioral, and physiological data. To this end, we used a dataset comprising 11,000 survey responses and applied multiple regression models, including Linear Regression, Gradient Boosting Regression, Extreme Gradient Boosting Regression Light Gradient-Boosting Machine Regression after data preprocessing. The performance of the models was evaluated using the mean absolute error, the root mean square error, and the coefficient of determination. Among the models evaluated, Gradient Boosting Regression achieved the best results, with a mean cross-validated R^2 of 0.759 after five-fold cross-validation.*

1. INTRODUCTION

Anxiety is among the most prevalent mental health concerns on a global scale, often evading detection in nonclinical populations due to its subjective and intricate nature [1]. In this context, Machine learning (ML) offers new opportunities to detect patterns associated with anxiety. These patterns are detected based on accessible, structured data.

The objective of this study is to predict anxiety levels using characteristics related to demographics, lifestyle, physiology, and psychological well-being. A series of multiple regression models were trained and evaluated.

A second objective of this study is the identification of the most significant predictive factors. The results of this study may support the development of data-driven screening tools for the early identification of people at risk of anxiety disorders.

2. RELATED WORK

This section reviews previous studies on mental health prediction using ML, organized first by general mental health disorders followed by anxiety, depression and stress.

2.1. General mental health

Chung and Teo [2] conducted an evaluation of classifiers for automated mental health screening, using response data from the Open Sourcing Mental Illness tech survey, https://osmhhelp.org/research.html. They compared five base model approaches with Extreme Gradient Boosting and a Deep Neural Network and a voting ensemble built from the same base learners. Under repeated cross-validation, Gradient Boosting Machine led all methods with an 0.88 accuracy. Extreme Gradient Boosting and Deep Neural Network achieved 0.87 and 0.86, respectively, while the voting ensemble reached 0.85 and the remaining base classifiers scored between 0.82 and 0.84.

Jain et al. [3] focused on predicting depression as a first step toward preventing suicide, using lifestyle and demographic attributes collected from 1,429 individuals. The study leveraged 76 features, ranging from income, marital status, and substance use to medical and financial information, to build predictive models using eight main ML algorithms. The Support Vector Machine classifier achieved the highest accuracy of 0.83.

Tate et al. [4] investigated the feasibility of predicting general mental health problems in adolescence using ML techniques. Using data from 7,638 participants in the Child and Adolescent Twin Study in Sweden [5], the authors trained models on 474 predictors derived from parental reports and national register data. Several common models were compared. Random Forest achieved the highest area under the receiver operating characteristic curve (AUROC) of 0.73.

Garriga et al. [6] developed a ML model to predict the probability of mental health crises over a 28-day horizon using electronic health records of 17,122 participants. Aimed at enabling proactive intervention, the model was designed for continuous patient monitoring and achieved an AUROC of 0.79, with 0.58 sensitivity and 0.85 specificity. Although the area under the precision-recall curve was only 0.15, the system demonstrated clinical utility. In a six-month prospective study, predictions were found helpful in 64% of cases, either by informing caseload management or mitigating crisis risk. This study is notable as one of the first to implement continuous, real-time risk prediction across a broad spectrum of mental health crises in a clinical setting.

2.2. Depression, anxiety and stress

Bhatnagar et al. [7] investigated the prevalence and impact of anxiety among Indian university students, focusing specifically on a sample of 127 engineering students. Using a Likert questionnaire, the study quantified anxiety levels and examined associated causes and effects. The authors applied multiple classifiers to predict anxiety severity. Among the tested algorithms, Random Forest achieved the highest accuracy of 0.78, followed by Support Vector Machine with accuracy of 0.75, and Naïve Bayes and Decision Tree both at 0.71.

Richter et al. [8] tackled the challenge of distinguishing anxiety from depression, using objective behavioral measures rather than self-report alone. In their study, 125 subclinical participants were stratified into four groups: high symptoms of depression, anxiety, or both and the non-symptomatic control group. Participants completed cognitive–emotional bias tasks. Models were trained to classify individuals based on their performance across these tasks. The analysis achieved 0.71 sensitivity and 0.70 specificity in distinguishing symptomatic from non-symptomatic participants. In a two-group model, the classifiers reached 0.68 accuracy for high-depression and 0.74 for high-anxiety groups.

Nemesure et al. [9] developed models to predict generalized anxiety disorder (GAD) and major depressive disorder (MDD) in a non-clinical setting using electronic health records. Drawing on a sample of 4,184 undergraduate students, the study excluded all direct psychiatric inputs and instead trained an ensemble ML pipeline on 59 biomedical and demographic features. On a test set, the model achieved an AUROC of 0.73 for GAD (sensitivity: 0.66, specificity: 0.70) and 0.67 for MDD (sensitivity: 0.55, specificity: 0.70). Feature attribution analysis using SHAP values revealed that variables such as satisfaction with living conditions and public health insurance were predictive of MDD, while up-to-date vaccinations and marijuana use were most predictive of GAD.

The Depression Anxiety Stress Scales (DASS) is a well-established self-report instrument designed to measure the emotional states of depression, anxiety, and stress. DASS-42 consists of 42 items and DASS-21 being its shorter 21-item version. Kumar et al. [10] investigated the prediction of anxiety, depression, and stress severity using two datasets: the DASS-42, sourced online, and the DASS-21, collected by the authors. The study employed eight ML algorithms along with a hybrid classification approach. Each model aimed to classify symptoms into five severity levels. While the hybrid method generally outperformed individual classifiers, the highest accuracy was achieved by the Radial Basis Function Network, which reached 0.97 for anxiety, 0.96 for depression, 0.96 for stress on DASS-42, and 0.82 for anxiety, 0.96 for depression, and 0.89 for stress on DASS-21.

Priya and Garg [11] applied ML algorithms to classify the severity of anxiety, depression, and stress into five levels using responses from 348 participants of the standardized DASS-21 questionnaire. The dataset included individuals from varied professional and cultural backgrounds. Five common methods were tested. Although Naïve

Bayes achieved the highest raw accuracy, Random Forest was selected as the best performing model based on the F1 score, which was prioritized due to class imbalance in the data. The Random Forest model had 0.79 accuracy for depression, 0.71 for anxiety and 0.72 for stress. The study also analyzed feature importance, identifying “scared without any good reason”, “life was meaningless”, and “difficult to relax” as the most predictive items for anxiety, depression, and stress, respectively.

Chavanne et al. [12] examined the potential of early prediction of clinical anxiety in adolescents using a combination of magnetic resonance imaging derived brain features and psychometric data. In a longitudinal study of 580 participants, non-anxious individuals at age 14 were followed up to assess anxiety diagnoses between ages 18 and 23. A voting classifier combining Random Forest, Support Vector Machine, and Linear Regression was trained on baseline gray matter volumes and psychological measures. The model achieved moderate predictive performance for pooled anxiety disorders with AUROC of 0.68. Psychometric features such as neuroticism, hopelessness, and emotional symptoms were the primary contributors for prediction. While brain imaging data did not improve prediction for pooled anxiety outcomes, it did enhance the detection of GAD, particularly with contributions from the caudate and pallidum volumes.

3. METHODOLOGY

The workflow adopted begins with an exploratory data analysis phase comprising univariate, bivariate, and multivariate analyses to understand the distribution, relationships, and interactions within the dataset. This step provides insights that inform subsequent preprocessing decisions. Afterwards, ML experiments are conducted using various regression models. The performance of these models is then evaluated based on appropriate metrics to assess predictive accuracy and generalization capability.

3.1. Dataset description

The dataset used in this study was obtained from Kaggle, https://www.kaggle.com/datasets/natezhang123/social-anxiety-dataset, and is based on survey responses from a diverse population on factors believed to be associated with social anxiety.

The dataset consists of 11,000 rows and 19 columns in total. The target attribute is Anxiety Level, which is a self-reported numerical value that indicates the respondent's perceived level of anxiety. The remaining features are independent variables. An overview of the characteristics of the dataset, accompanied by their classifications and descriptions, is provided in *table 1*.

Table 1. Features description

Name	Description	Type
Age	Age of the respondent	Numerical
Gender	Biological sex of the respondent	Categorical
Occupation	Professional field or job role	Categorical
Sleep Hours	Average sleep hours per night	Numerical
Physical Activity (hrs/week)	Weekly physical activity duration	Numerical
Caffeine Intake (mg/day)	Average daily caffeine intake	Numerical
Alcohol Consumption (drinks/week)	Alcohol consumption per week	Numerical
Smoking	Whether the individual smokes	Boolean
Family History of Anxiety	Presence of anxiety in family history	Boolean
Stress Level (1-10)	Self-reported stress level	Numerical
Heart Rate (bpm)	Resting heart rate	Numerical
Breathing Rate (breaths/min)	Respiratory rate	Numerical
Sweating Level (1-5)	Degree of sweating	Numerical
Dizziness	Experience of dizziness	Boolean
Medication	Use of medication for anxiety	Boolean
Therapy Sessions (per month)	Monthly therapy sessions	Numerical
Recent Major Life Event	Recent stressful life event	Boolean
Diet Quality (1-10)	Nutritional quality of diet	Numerical
Anxiety Level (1-10)	Self-reported anxiety level	Numerical

3.2. Data analysis

In this section univariate, bivariate and multivariate analysis is carried out on the dataset to uncover underlying relationships between attributes. By conducting this analysis, integrity of the features is ensured before they are used as inputs for the predictive models.

3.2.1. Univariate analysis

An initial analysis revealed information on the distribution of data considering numerical attributes. The mean and median are nearly equal for most dataset columns as described in *table 2*. Such a distribution implies that the data is not heavily influenced by outliers and approximates a normal distribution, which provides a statistically sound basis for the subsequent application of regression algorithms. By examining the consistency of each attribute, we mitigate the risk of introducing bias. The symmetry in the data distribution is a key indicator of the reliability of the data collection process.

Table 2. Descriptive statistics of dataset columns

Feature	Mean	Median	Std	Min	25%	50%	75%	Max
Age	40.24	40.00	13.23	18.00	29.00	40.00	51.00	64.00
Sleep Hours	6.65	6.70	1.22	2.30	5.90	6.70	7.50	11.30
Physical Activity	2.94	2.80	1.82	0.00	1.50	2.80	4.20	10.10
Caffeine Intake	286.09	273.00	144.81	0.00	172.00	273.00	382.00	599.00
Alcohol Consumption	9.70	10.00	5.68	0.00	5.00	10.00	15.00	19.00
Stress Level	5.86	6.00	2.92	1.00	3.00	6.00	8.00	10.00
Heart Rate	90.92	92.00	17.32	60.00	76.00	92.00	106.00	119.00
Breathing Rate	20.96	21.00	5.16	12.00	17.00	21.00	25.00	29.00
Sweating Level	3.08	3.00	1.39	1.00	2.00	3.00	4.00	5.00
Therapy Sessions	2.43	2.00	2.18	0.00	1.00	2.00	4.00	12.00
Diet Quality	5.18	5.00	2.89	1.00	3.00	5.00	8.00	10.00
Anxiety Level	3.93	4.00	2.12	1.00	2.00	4.00	5.00	10.00

3.2.2. Bivariate analysis

A Pearson correlation matrix was computed to assess the relationships between numerical variables. The strongest positive correlation (r = 0.67) was observed between Anxiety Level and Stress Level, suggesting that stress is a significant contributing factor. A negative correlation (r = -0.49) was identified between Sleep Hours and Anxiety Level, suggesting that reduced sleep may exacerbate symptoms. A negative correlation (r = -0.41) between Diet Quality and Anxiety Level was also identified. The results of the therapeutic sessions indicated a slight mitigating effect (r = -0.22), while Caffeine Intake exhibited a weak positive correlation (r = 0.35) with Anxiety Level.

3.2.3. Multivariate analysis

Multivariate analysis indicates that people experiencing the highest levels of anxiety tend to have elevated stress levels, shorter sleep duration, and lower levels of physical activity as shown in *figure 1*.

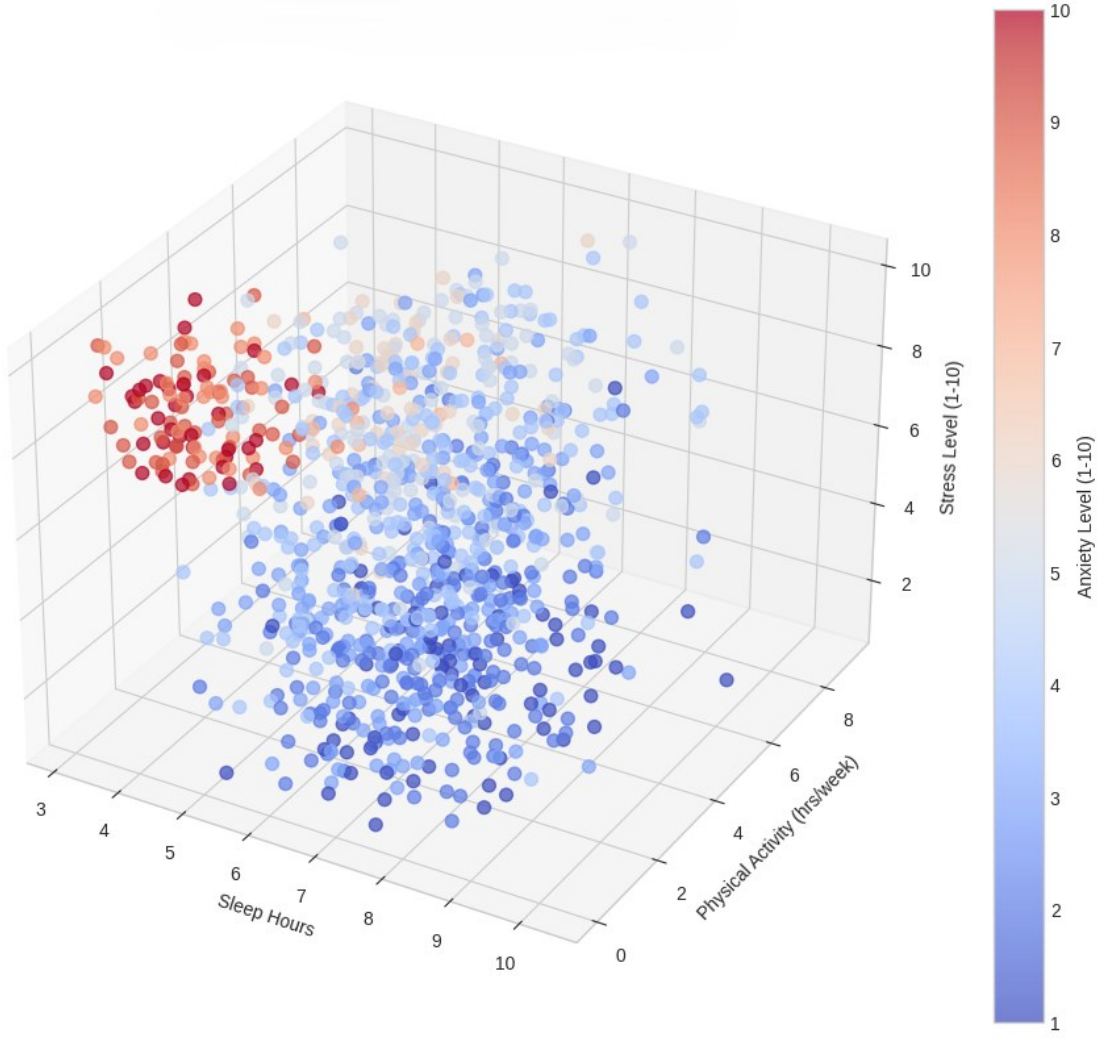

Fig. 1. Relationship between sleep, activity, stress and anxiety

3.3. Data preprocessing

To prepare the data for ML models, several preprocessing steps were implemented. Columns with a weak correlation with the target variable were eliminated, leaving five features, excluding the target variable. The remaining features are Sleep Hours, Caffeine Intake, Stress Level, Physical Activity and Therapy Sessions. The Physical Activity feature and the Therapy Sessions feature were transformed using log1p transformation to remove outlier bias. The values of all columns were scaled using min-max scaling. The data set was partitioned into a training set and a testing set, with the training set constituting 80% of the total data and the testing set comprising the remaining 20%.

4. EXPERIMENTS

Experimental procedures were conducted on the Google Colab platform using Python version 3.11. Data processing and visualization were supported by standard libraries, including Scikit-learn, Pandas, and Matplotlib.

The performance of all models is shown in *table 3*. Five-fold cross-validation was implemented for all models to obtain an average cross-validated R^2 score. Gradient Boosting Regression delivered the best performance, having a cross-validated R^2 of 0.759. Support Vector Regression and Light Gradient-Boosting Machine Regression attained comparable performance, with mean R^2 values of 0.758 and 0.756, respectively. Random Forest Regression demonstrated a slightly lower mean R^2 value of 0.749. Extreme Gradient Boosting Regression obtained a performance of 0.733. Linear Regression demonstrated significantly lower performance than previous models, with an R^2 value of 0.688. Decision Tree Regression exhibited the lowest performance, with a mean R^2 of 0.528.

Table 3. Performance comparison of the regression models

Model	MAE	RMSE	R^2	CV R^2 (mean)
Gradient Boosting Regression	0.822	1.032	0.770	0.759
Support Vector Regression	0.828	1.031	0.770	0.758
Light Gradient-Boosting Machine Regression	0.829	1.030	0.771	0.756
Random Forest Regression	0.851	1.058	0.759	0.749
Extreme Gradient Boosting Regression	0.869	1.079	0.749	0.733
Linear Regression	0.942	1.189	0.695	0.688
Decision Tree Regression	1.093	1.439	0.553	0.528

5. CONCLUSIONS

This study explored the predictive modeling of anxiety severity using a combination of demographic, behavioral, and physiological data derived from a large-scale survey dataset. After correlation analysis, the research identified stress level, number of sleep hours, physical activity, caffeine intake, and number of therapy sessions as factors significantly correlated with anxiety.

Among the suite of regression models evaluated, Gradient Boosting Regression emerged as the most effective, achieving the highest performance after cross-validation. Support Vector Regression and Light Gradient-Boosting Machine Regression also delivered

comparable results, demonstrating the utility of ensemble and kernel-based models in mental health prediction tasks.

Key insights from the analysis reaffirm the central role of stress, sleep, and physical activity in influencing anxiety levels. These findings are consistent with existing psychological research and underline the potential for ML to support data-driven mental health assessments and early intervention tools.

However, certain limitations must be acknowledged. The reliance on self-reported survey data introduces potential response bias, and the cross-sectional nature of the dataset prevents observation of temporal trends. These factors should be addressed in future research for more comprehensive mental health modeling.

REFERENCES

[1] World Health Organization, *World Mental Health Report: Transforming Mental Health for All*, World Health Organization, 2022.

[2] J. Chung, J. Teo, *Single classifier vs. ensemble machine learning approaches for mental health prediction*, Brain informatics, vol. 10, no. 1, p. 1, 2023.

[3] T. Jain, A. Jain, P.S. Hada, H. Kumar, V.K. Verma, A. Patni, *Machine learning techniques for prediction of mental health*, 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1606-1613, 2021.

[4] A.E. Tate, R.C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, R. Kuja-Halkola, *Predicting mental health problems in adolescence using machine learning techniques*, PloS one, vol. 15, no. 4, 2020.

[5] H. Anckarsäter, S. Lundström, L. Kollberg, N. Kerekes, C. Palm, E. Carlström, N. Långström, P.K. Magnusson, L. Halldner, S. Bölte, C. Gillberg, C. Gumpert, M. Råstam, P. Lichtenstein, *The child and adolescent twin in Sweeden (CATSS)*, Twin Research and Human Genetics, vol. 14, no. 6, pp. 495-508, 2011.

[6] R. Garriga, J. Mas, S. Abraha, J. Nolan, O. Harrison, G. Tadros, A. Matic, *Machine learning model to predict mental health crises from electronic health records*, Nature medicine, vol. 28, no. 6, pp. 1240-1248, 2022.

[7] S. Bhatnagar, J. Agarwal, O.R. Sharma, *Detection and classification of anxiety in university students through the application of machine learning*, Procedia Computer Science, vol. 218, pp. 1542-1550, 2023.

[8] T. Richter, B. Fishbain, A. Markus, G. Richter-Levin, H. Okon-Singer, *Using machine learning-based analysis for behavioral differentiation between anxiety and depression*, Scientific reports, vol. 10, no. 1, p. 16381, 2020.

[9] M.D. Nemesure, M.V. Heinz, R. Huang, N.C. Jacobson, *Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence*, Scientific reports, vol. 11, no. 1, p. 1980, 2021.

[10] P. Kumar, S. Garg, A. Garg, *Assessment of anxiety, depression and stress using machine*

learning models, Procedia Computer Science, vol. 171, pp. 1989-1998, 2020.

[11] A. Priya, S. Garg, N.P. Tigga, *Predicting anxiety, depression and stress in modern life using machine learning algorithms*, Procedia Computer Science, vol. 167, pp. 1258-1267, 2020.

[12] A.V. Chavanne, M.L. Paillere Martinot, J. Penttilä, Y. Grimmer, P. Conrod, A. Stringaris, B. Van Noort, C. Isensee, A. Becker, T. Banaschewski, et al., *Anxiety onset in adolescents: a machine-learning prediction*, Molecular psychiatry, vol. 28, no. 2, pp. 639-646, 2023.

Received: September 2025, Accepted: November 2025, Published: December 2025
Digital Object Identifier: https://doi.org/10.34302/CJEE/LDGV2056

SIMULATION OF AN IOT AIR QUALITY MONITORING SYSTEM

Bogdan **CHIȘ**, Cristinel **COSTEA,** Eleonora **POP**
Technical University of Cluj-Napoca
chis.do.paul@student.utcluj.ro, cristinel.costea@ieec.utcluj.ro, eleonora.pop@ieec.utcluj.ro

Keywords: Air quality, Air pollution, Anomaly detection, Internet of Things

Abstract: *Air pollution remains a major global health and environmental concern, driving the need for efficient and scalable monitoring systems. This paper presents a simulation-based Internet of Things architecture for air quality monitoring, focusing on real-time data transmission, anomaly detection, and visual analysis. The system simulates 25 monitoring stations, each with six virtual sensors representing key pollutants, totaling 150 sensor processes. These sensors communicate with dedicated gateway processes via MQTT and transmit data to a cloud-hosted web server for aggregation, air quality index calculation, visualization, and database storage. The simulation uses real-world data collected in Seoul to emulate realistic pollution conditions. An interactive web interface displays live air quality index values through maps and time series charts, allowing the detection of anomalies.*

1. INTRODUCTION

The expansion of industries and transportation has increased pollution, which poses a major challenge for developing countries. This issue has garnered increasing attention from both governments and citizens. According to a report, prolonged exposure to particulate matter, an airborne pollutant found both indoors and outdoors, has been linked to approximately five million deaths worldwide, ranking fifth among all global health risks [1]. If air quality continues to deteriorate, managing the economic and health costs of pollution will become a challenge for governments.

Air quality is a concern in major urban agglomerations around the world. Cities such as Paris [2], Athens [3], Sao Paulo [4], Beijing [5], and Tehran [6] face challenges in

managing air pollution levels due to factors such as high population density, industrial activity and transportation emissions.

To address this, air quality monitoring systems are essential for efficiently tracking pollution levels before they reach critical thresholds. Traditional air quality monitoring stations are often expensive to install, large, and expensive to maintain, limiting their widespread deployment in urban areas [7]. Although these systems can provide accurate measurements, they often involve lengthy offline processes. Furthermore, there is a growing demand for air quality information collected in wide areas and over a long period of time [8].

This research aims to enable pollution mitigation strategies through anomaly detection in a simulated IoT-based air quality monitoring system. The simulation focuses on developing a scalable architecture for handling data streams from multiple sensor nodes, with the goal of creating a real-world fog/edge air quality monitoring system. It also aims to incorporate real air quality data, reflecting real-world patterns and variability. While interactive visualization tools are created, their primary purpose is to enable analysis and exploration of air quality data to identify patterns and potential anomalies. Finally, the study evaluates the scalability and performance of the system.

Simulation is important because it provides a cost-effective way to use available air quality data and test the system architecture and anomaly detection methods before deploying expensive real-world networks. It also enables controlled experimentation and analysis of varied air quality conditions, which helps the development of effective anomaly detection algorithms.

2. RELATED WORK

An IoT system was developed by Dhingra et al. [9] using gas sensors, Arduino development boards, and Wi-Fi modules. This system can be installed in different locations to monitor air pollution. The gas sensors collect data from the air and send it to the Arduino, which then transmits the data to the cloud via the Wi-Fi module. An Android app called IoT-Mobair was also created, which allows users to access air quality information from the cloud. The app predicts the pollution levels along a user's travel route and provides a warning if the pollution exceeds safe levels. In addition, the collected air quality data can be used to forecast future levels of the air quality index (AQI).

Wearable devices used to monitor air quality were also developed. Park et al. [10] used a GeoAir2 portable air monitoring system, which integrates a $PM_{2.5}$ sensor and a GPS. In this study, a travel activity diary was created. The diary includes questions about the type of location visited (e.g. home, workplace, school, transit stop), the activity performed, the exact time the activity began, and the mode of transportation used. The Research Electronic

Data Capture [11], a web platform to create and manage online surveys and databases, was used to develop the surveys.

Reshi et al. [12] designed a platform based on a wireless sensor network, called VehNode, that provided automobiles with the ability to monitor the level of pollutants in exhaust fumes released by the vehicle.

Several studies have focused on the integration of IoT technology for monitoring both air quality and noise pollution, offering a comprehensive approach to environmental monitoring. These systems typically utilize a combination of sensors to measure various levels of pollutants and noise [13, 14]. This dual monitoring approach allows for a more detailed view of environmental quality.

Niculae [15] explores the use of big data and machine learning techniques to predict air quality, using data collected between 2018 and 2021 from measurement probes in Romania for PM_{10}, NO_2, O_3 and SO_2. The analysis reveals that time series models perform better than traditional models. In addition, artificial neural network models are effective in classifying pollutants' AQI levels but do not accurately predict their actual values.

3. AIR QUALITY INDEX CONCEPTS AND CALCULATION

The AQI is a numerical scale that is used to communicate the quality of air and its potential health effects. It simplifies air quality reporting by converting pollutant concentrations into a standardized scale, usually ranging from 0 to 500 [16]. Each pollutant has a specific subindex calculated based on its measured concentration using breakpoints provided by environmental or governmental agencies.

In *equation 1*, AQI_p represents the AQI for the specific pollutant p, C_p is the concentration of the pollutant p, C_{low} is the concentration breakpoint that is less than or equal to C_p, C_{high} is the concentration breakpoint that is greater than or equal to C_p, I_{low} is the AQI value corresponding to C_{low} and I_{high} is the AQI value corresponding to C_{high}.

$$AQI_p = \frac{(I_{high} - I_{low})}{(C_{high} - C_{low})} \times (C_p - C_{low}) + I_{low} \quad (1)$$

4. SYSTEM ARCHITECTURE

To effectively manage data collection from many distributed sensors, the system is based on an architecture with sensor groups and gateway processes, as shown in *figure 1*. The sensors are organized into groups $(S_1, S_2, \ldots, S_m)$, where each group represents the sensors deployed at a specific monitoring station.

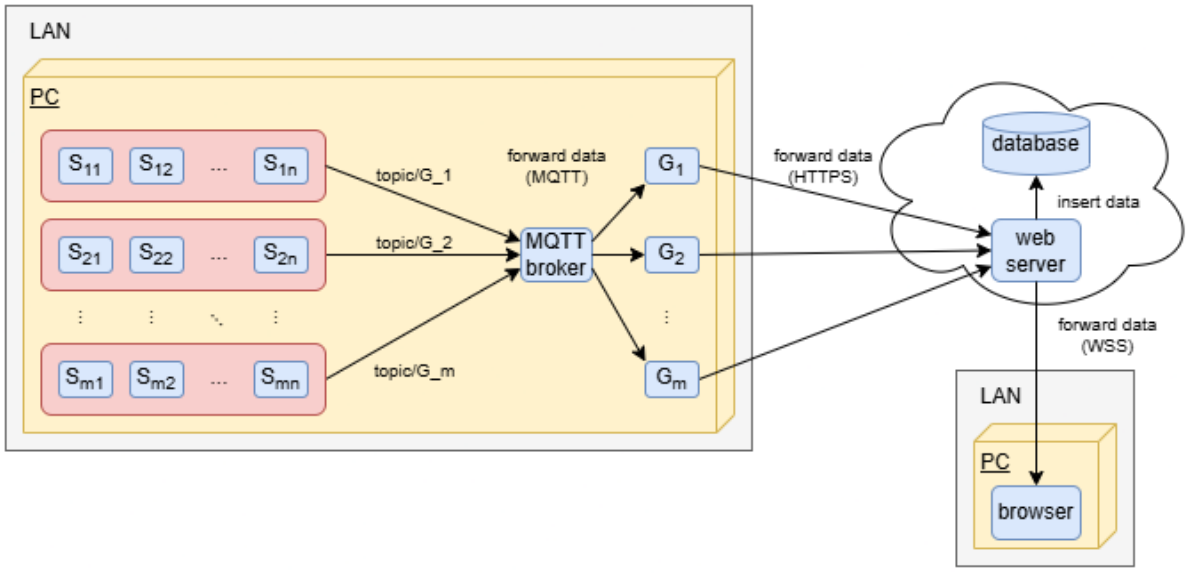

Fig. 1. The architecture of the system used in simulation

Within each sensor group, *n* individual sensors communicate data to a central MQTT broker. Each of the *m* gateway processes, representing distinct air monitoring stations, is associated with a specific MQTT topic. All sensors belonging to a specific sensor group publish their data to the corresponding topic.

The MQTT broker routes the data to the appropriate gateway process $(G_1, G_2, ..., G_m)$. Acting as subscribers, these gateway processes aggregate the incoming data into a standardized format before forwarding it to the web server via HTTPS.

The web server is deployed in the cloud to store the data, calculate the AQI, and send them to the client application. WebSockets are used to enable communication between the web server and the browser application, ensuring immediate data updates.

The client application, accessible through a web browser, displays air quality data using interactive visualizations. The key features are map visualization and time series charts.

5. EXPERIMENTAL SETUP

The data represents the real values recorded hourly in Seoul from January 2017 to December 2019. The data set contains records from 25 air quality monitoring stations. The relevant substances to be monitored are: SO_2, NO_2, CO, O_3, PM_{10} and $PM_{2.5}$. The dataset is available on Kaggle, https://www.kaggle.com/datasets/bappekim/air-pollution-in-seoul.

In the experimental setup, a total of 25 gateway processes were instantiated, each representing a separate air quality monitoring station. For each gateway, there are six sensor processes each corresponding to a monitored pollutant. These sensors were simulated as individual processes that have a total of 150 sensors across all stations. All sensors and

gateway processes ran on a single local machine used for development and testing, equipped with an Intel Core i7-1165G7 CPU @ 2.80GHz, 8 GB RAM, running Widows 10. The web server and database were hosted remotely.

The sensor processes and gateway processes were implemented as individual Python processes. The MQTT broker used was Mosquitto. The web server was built using the Flask framework. Data storage was handled by a MongoDB database. The web client was developed using HTML, CSS, and JavaScript.

6. DISCUSSIONS

6.1. Key findings

The chart in *figure 2* illustrates fluctuations in air quality indicators over time for a specific monitoring station, and the presence of sudden spikes in pollutants. The analysis indicates that $PM_{2.5}$ levels determine the worst AQI readings, as evidenced by the overlapping plots for the highest AQI and $PM_{2.5}$ concentrations. From an anomaly detection perspective, spikes in $PM_{2.5}$ levels emphasize the importance of closely monitoring concentrations to identify possible pollution events.

Recognizing such widespread anomalies can be useful for system notifications, prompting automated alerts or warnings to stakeholders and the public. Incorporating notifications into the monitoring framework ensures timely responses to pollution spikes, allowing immediate actions to mitigate potential health impacts.

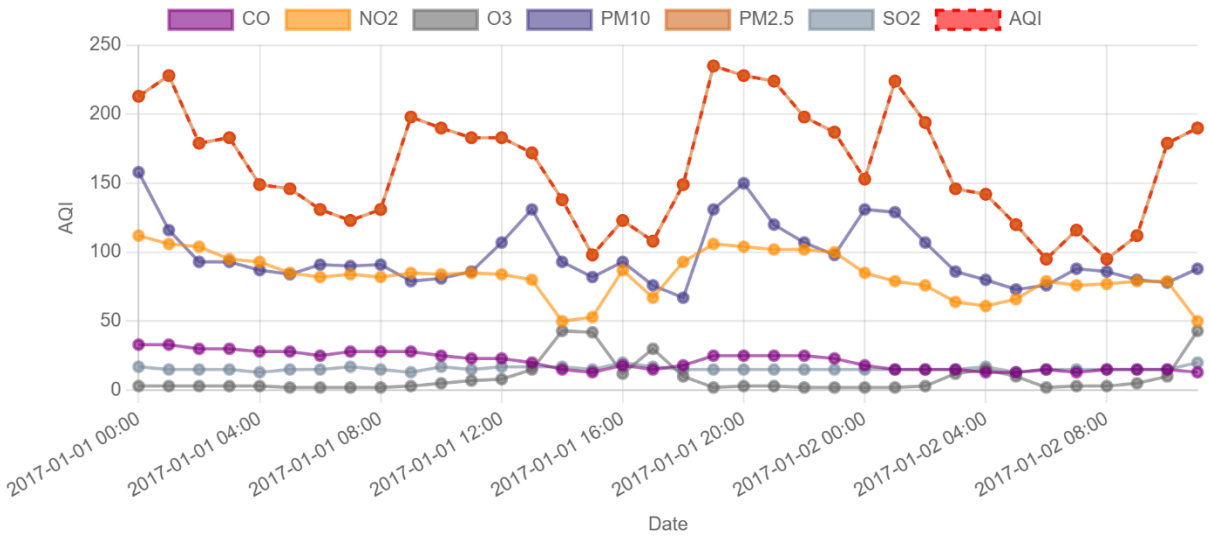

Fig. 2. Air quality index history chart for a station

6.2. Scalability analysis

The scalability analysis was performed in a simulated environment on a single machine. This limits the maximum number of simulated sensors and gateways. Using 150 sensor processes and 25 gateway processes, the system ran smoothly without any performance problems.

The MQTT broker, serving as the communication backbone for sensor data transmission limitations that affect its throughput and connection capacity. High volumes of concurrent sensor data streams can lead to broker saturation, message delays, or dropped messages.

The web server can become a bottleneck under increased load. Web servers have finite resources and processing capacity. When the number of simultaneous users or calls exceeds these limits, the response times increase or the server may not respond.

Database systems also pose scalability challenges. As data volume increases due to the continuous storage and retrieval of sensor data, demands increase.

The findings of this study are subject to limitations due to the scope of the dataset. Seoul's unique air quality patterns, influenced by local factors, and the limited time frame may not capture long-term trends. Furthermore, hourly measurements may miss short-term pollution spikes. The pollutant range of the dataset, specific to Seoul, may also limit its applicability elsewhere.

7. FUTURE WORK

Future developments of this system aim to enhance its analytical capabilities and practical applicability. A notable extension involves the use of Voronoi diagrams to partition areas based on the level of pollution [3].

Another enhancement could be the integration of algorithms to detect potential sources of air pollution. By combining spatial pollutant concentration gradients with meteorological data (e.g., wind speed and direction), it may be possible to infer the likely sources of pollution, providing actionable insights for regulatory authorities.

Machine learning can enhance anomaly detection by reducing false positives. For instance, Isolation Forest algorithm isolates abnormal data points in complex datasets, reducing false positives in IoT sensor networks compared to traditional methods [17, 18]. Meanwhile, autoencoders learn normal pollution patterns and detect anomalies through reconstruction errors, proving effective in virtual monitoring systems [19].

In addition, the system may incorporate time series forecasting models for AQI values. This would support early warning systems and proactive public health responses by

anticipating pollution peaks and trends. One model that could be implemented is Long Short-Term Memory since the model is able to handle long time series [20].

These extensions would transform the current simulation into a decision support tool, capable of both retrospective analysis and forward-looking prediction.

8. CONCLUSIONS

This study presents a simulation-based IoT architecture for air quality monitoring, emphasizing scalability, real-time visualization, and anomaly detection. Using MQTT communication, cloud-hosted web server and database, and interactive client interfaces, the system enables continuous monitoring and analysis of key pollutants in multiple locations.

The experimental setup involved 25 gateway processes and 150 individual sensor processes that send real-world pollutant data from Seoul. The results showed that $PM_{2.5}$ concentrations consistently led to the highest AQI readings, underscoring its critical role in pollution-related health risks. The system also proved stable under simulation, with no performance issues encountered when managing all processes on a single development machine.

In addition to monitoring pollutant levels, the platform supports anomaly detection by identifying sudden spikes in concentration levels. These events may trigger alerts and provide early indicators of pollution episodes.

REFERENCES

[1] World Health Organization, *7 Million Premature Deaths Annually Linked to Air Pollution*, 2014. [Online]. Available: https://www.who.int/news/item/25-03-2014-7-million-premature-deaths-annually-linked-to-air-pollution.

[2] EOLEAF, *Air Quality in Paris Presents Ongoing Concern for Olympic and Paralympic Games*, 2024. [Online]. Available: https://eoleaf.com/blogs/blog/air-quality-in-paris-presents-ongoing-concern-for-olympic-and-paralympic-games.

[3] D. Deligiorgi, K. Philippopoulos, *Spatial Interpolation Methodologies in Urban Air Pollution Modeling: Application for the Greater Area of Metropolitan Athens, Greece*, Advanced air pollution, vol. 17, pp. 341-362, 2011.

[4] M.C.Q.D. Oliveira, R.M. de Miranda, M. de Fátima Andrade, P. Kumar, *Impact of urban green areas on air quality: An integrated analysis in the metropolitan area of São Paulo*, Environmental Pollution, vol. 372, p. 126082, 2025.

[5] H. Wang, J. Zhou, X. Li, Q. Ling, H. Wei, L. Gao, Y. He, M. Zhu, X. Xiao, Y. Liu, et al., *Review on recent progress in on-line monitoring technology for atmospheric pollution source emissions in China*, Journal of Environmental Sciences, vol. 123, pp. 367-386, 2023.

[6] A. Rad, M.J. Nematollahi, A. Pak, M. Mahmoudi, *Predictive modeling of air quality in the Tehran megacity via deep learning techniques*, Scientific Reports, vol. 15, no. 1, p. 1367, 2025.

[7] K.B. Shaban, A. Kadri, E. Rezk, *Urban Pollution Monitoring System with Forecasting Models*, IEEE Sensors Journal, vol. 16, no. 8, p. 2606, 2015.

[8] W.Y. Yi, K.M. Lo, T. Mak, K.S. Leung, Y. Leung, M.L. Meng, *A survey of wireless sensor network based air pollution monitoring systems*, Sensors, vol. 15, no. 12, pp. 31392-31427, 2015.

[9] S. Dhingra, R.B. Madda, A.H. Gandomi, R. Patan, M. Daneshmand, *Internet of Things mobile--air pollution monitoring system (IoT-Mobair)*, IEEE Internet of Things Journal, vol. 6, no. 3, pp. 5577-5584, 2019.

[10] Y.M. Park, D. Chavez, S. Sousan, N. Figueroa-Bernal, J.R. Alvarez, J. Rocha-Peralta, *Personal exposure monitoring using GPS-enabled portable air pollution sensors: A strategy to promote citizen awareness and behavioral changes regarding indoor and outdoor air pollution*, Journal of exposure science & environmental epidemiology, vol. 33, no. 3, pp. 347-357, 2023.

[11] E.F. Patridge, T.P. Bardyn, *Research electronic data capture (REDCap)*, Journal of the Medical Library Association: JMLA, vol. 106, no. 1, p. 142, 2018.

[12] A.A. Reshi, S. Shafi, A. Kumaravel, *VehNode: Wireless sensor network platform for automobile pollution control*, 2013 IEEE Conference on Information & Communication Technologies, pp. 963-966, 2013.

[13] K. Cornelius, N.K. Kumar, S. Pradhan, P. Patel, N. Vinay, *An efficient tracking system for air and sound pollution using IoT*, 2020 6th International conference on advanced computing and communication systems (ICACCS), pp. 22-25, 2020.

[14] S. Deshmukh, S. Surendran, M.P. Sardey, *Air and Sound Pollution Monitoring System using IoT*, International Journal on Recent and Innovation Trends in Computing and Communication, vol. 5, no. 6, pp. 175-178, 2017.

[15] A.M. Niculae, *Analysis of Romanian Air Quality using Machine Learning Techniques*, Database Systems Journal, vol. 13, 2022.

[16] Y.P. Kim, G. Lee, *Trend of air quality in Seoul: Policy and science*, Aerosol and Air Quality Research, vol. 18, no. 9, pp. 2141-2156, 2018.

[17] M.R. Dileep, V. Balasubramaniam, R. Mahajan, *Artificial Intelligence in Air Quality Assessment: A Comprehensive Review of Anomaly Detection Methods for Environmental Sustainability*, SGS-Engineering & Sciences, vol. 1, no. 1, 2025.

[18] V. Evagelopoulos, N.D. Charisiou, P. Begou, *Fault detection of air quality measurements using artificial intelligence*, E3S Web of Conferences, vol. 436, p. 10005, 2023.

[19] R. Abrol, *AI-Powered Anomaly Detection in Air Pollution for Smart Environmental Monitoring*, Indian Journal of Artificial Intelligence and Neural Networking (IJAINN), vol. 5, no. 3, pp. 1-5, 2025.

[20] J.K. Makala, L. Chidzulo, *Forecasting of Air Quality with Machine Learning, Impact Assessment in the Age of Artificial Intelligence*, 2025.

Received: September 2025, Accepted: November 2025, Published: December 2025
Digital Object Identifier: https://doi.org/10.34302/CJEE/MEIZ3377

A DATA-DRIVEN ANALYSIS OF REMOTE WORK SALARIES AND SATISFACTION

Ioan Alexandru **VLAD**, Dacian **TEPFENHART**
Technical University of Cluj-Napoca, Romania
al3xro@gmail.com, dacianXpaul@gmail.com

Keywords: remote work, salary analysis, data science, job satisfaction

Abstract: *This paper provides a data-driven analysis of remote work salaries, leveraging a real-world dataset processed with Python. The study investigates how industry, experience, employment type, and remote flexibility impact salary and job satisfaction. Insights are visualized using statistical plots and support a broader understanding of global salary trends in remote work environments.*

1. INTRODUCTION

The rapid adoption of remote work has significantly reshaped global labor markets, influencing both salary structures and employee satisfaction. Advances in digital technologies and increased workplace flexibility have enabled organizations to recruit talent globally, while employees benefit from improved work-life balance. Understanding how factors such as industry, experience level, and employment type affect compensation in remote environments has become increasingly important. This paper presents a data-driven analysis of remote work salaries and job satisfaction using a real-world dataset, aiming to identify key trends and relationships that characterize contemporary remote work dynamics.

2. METHODOLOGY – DATA SOURCES AND TOOLS

The present analysis is based on a dataset containing remote work salary data, processed and visualized using Python. The references used provided both a socio-economic context and methodological foundation for the analysis.

Context:

- Sources [1], [2], and [3] provide insight into remote work trends, global salary expectations, and worker preferences, which helped frame the broader relevance of the dataset.
- Economic relevance: The OECD report [4] highlights how the COVID-19 pandemic influenced remote work, justifying the need for salary structure analysis in this new normal.
- Technical tools: Python libraries such as Pandas [5], Matplotlib [6], and Scikit-learn [7] were used for data cleaning, transformation, statistical summary generation, and data visualization. These tools are recognized and validated in the field of data science.

3. CODE FOR ANALYSIS IN GOOGLE COLAB:

```
# Remote Salary Analysis - Google Colab Notebook
import pandas as pd import seaborn as sns import
matplotlib.pyplot as plt # Load dataset df =
pd.read_csv("/content/Work_From_Anywhere_Salary_Data.
csv")
# Preview print(df.head()) print(df.info())
# Descriptive statistics print(df.describe())
# Check for missing values print(df.isnull().sum())
# Salary distribution plt.figure(figsize=(10, 6)) sns.histplot(df['Salary (Annual)'], kde=True, bins=30,
color='skyblue') plt.title('Annual Salary Distribution') plt.xlabel('Salary (Annual)') plt.ylabel('Frequency')
plt.show()
# Boxplot - Salary by Experience Level plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='Experience Level', y='Salary (Annual)', palette='
    Set2') plt.title('Salary by Experience Level') plt.show()
# Average salary per industry industry_salary = df.groupby('Industry')['Salary (Annual)'].mean().
    sort_values()
plt.figure(figsize=(10, 8)) industry_salary.plot(kind='barh', color='teal')
plt.title('Average Salary by Industry') plt.xlabel('Average Salary (Annual)')
plt.ylabel('Industry') plt.show()
# Comparison: Remote vs Onsite plt.figure(figsize=(8, 6))
sns.boxplot(data=df, x='Remote Flexibility', y='Salary (Annual)', palette
    ='coolwarm') plt.title('Salary by Remote Flexibility') plt.show()
# Average job satisfaction by Tech Stack
tech_stack_satisfaction = df.groupby('Tech Stack')['Job Satisfaction
    Score (1-10)'].mean().sort_values() plt.figure(figsize=(10, 8))
tech_stack_satisfaction.plot(kind='barh', color='orange')
plt.title('Average Job Satisfaction by Tech Stack') plt.xlabel('Average
Satisfaction Score') plt.ylabel('Tech Stack') plt.show()
# Employment types distribution employment_type_counts = df['Employment Type'].value_counts()
plt.figure(figsize=(6, 6)) employment_type_counts.plot(kind='pie', autopct='%1.1f%%', startangle
```

```
    =140) plt.title('Distribution of Employment Types') plt.ylabel('')
plt.show()
# Correlation matrix for numeric variables plt.figure(figsize=(10, 6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm', fmt
    ='.2f') plt.title('Correlation Matrix') plt.show()
```

4. DATASET PREVIEW AND STRUCTURE

To understand the structure and context of the analysis, Table 1 presents the first five entries of the dataset. Each row represents a job position in a remote or flexible working environment, and each column captures a specific characteristic of the job or employee:

Table 1. Entries of the dataset

	Company	Job Title	Industry	Location	Employment Type
1	Microsoft	Data Analyst	Media	Austin	Part-time
2	Apple	Data Scientist	Retail	San Francisco	Part-time
3	Amazon	Software Engineer	Healthcare	San Francisco	Full-time
4	Tesla	Data Analyst	Retail	Austin	Contract
5	Adobe	DevOps Engineer	Healthcare	New York	Contract

This snapshot highlights the diversity in employer type, industry, geographical location, and employment contract. Notably:

Industry Variety: The dataset includes entries from Media, Retail, and Healthcare industries, allowing for cross-sectoral comparisons.

Employment Types: Full-time, part-time, and contract-based employment are all represented, offering insight into different compensation structures and job stability.

Geographic Distribution: The jobs are based in prominent U.S. tech hubs like San Francisco, Austin, and New York, which are known for high employment demand and competitive salaries.

Job Roles: Positions vary from Data Analysts to Software and DevOps Engineers, covering a wide range of technical expertise. This initial view, summarized in Table 2, supports the generalizability of the dataset and establishes a solid foundation for further exploratory and statistical analysis.

Table 2. Dataset structure and job characteristics

	Experience Level	Remote Flexibility	Salary (annual)	Currency
1.	Mid	Remote	155200.11	AUD
2.	Lead	Remote	106365.54	INR
3.	Lead	Remote	91026.49	INR
4.	Mid	Onsite	41824.38	EUR
5.	Senior	Remote	143929.78	USD

This subset illustrates how experience level and remote flexibility interact with salary levels and global currencies:

Experience Impact: Lead and Senior professionals command higher salaries on average than Mid-level employees, consistent with broader labor market trends.

Remote Flexibility: Positions marked as fully remote tend to be better compensated than onsite roles, even when accounting for differences in geographic location and currency.

... geographic and contractual flexibility. The salary distribution presented in *fig. 1* further highlights these variations across the dataset.

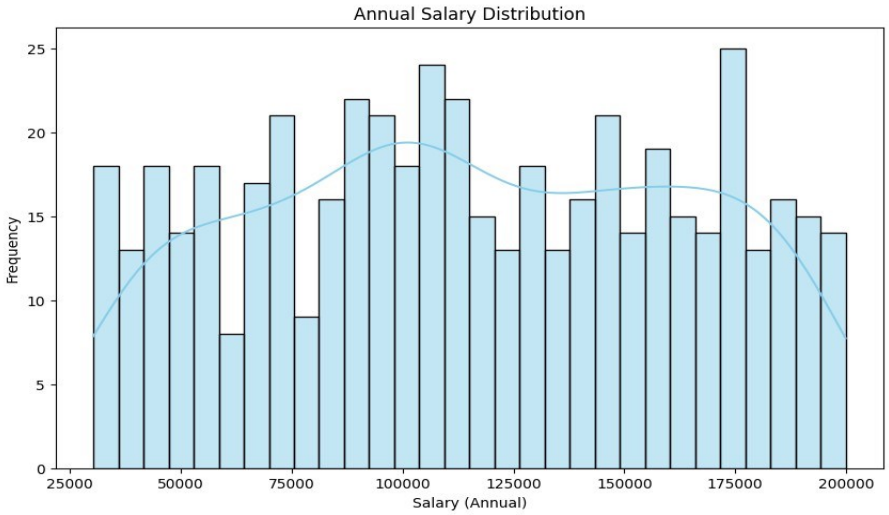

Fig. 1. Annual Salary Distribution (Histogram)

The salary distribution shown in *fig. 2* follows a slightly right-skewed curve, suggesting that while most employees earn within a moderate range, a minority earn substantially higher or lower salaries. These outliers may be attributed to senior roles, niche technical skills, or geographic salary differentials.

Fig. 2. Salary by Experience Level (Boxplot)

This boxplot compares annual salaries across four categories of experience level: Junior, Mid, Senior, and Lead. A positive correlation between experience and compensation is evident. While junior employees have more clustered earnings, higher-level professionals display a broader salary range, which may reflect diverse negotiation power, specializations, or company size.

The average annual salaries grouped by industry, illustrated in *fig. 3*, show that sectors such as Technology, Finance, and Healthcare offer higher average salaries, while Retail and Media tend to fall below the overall average. These findings highlight how industry type plays a crucial role in shaping remote salary structures.

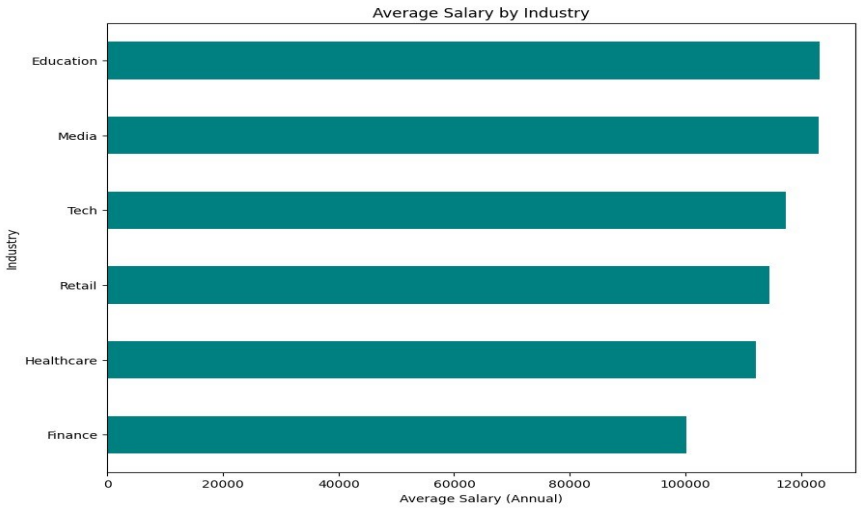

Fig. 3. Average Salary by Industry (Horizontal Bar Plot)

Salary variation by work arrangement type—Fully Remote, Hybrid, or Onsite—is illustrated in *fig. 4*. Fully remote roles demonstrate the highest median salary and wider distribution, which may result from global hiring practices and increased demand for flexible work. Onsite positions, by contrast, show more constrained salary ranges.

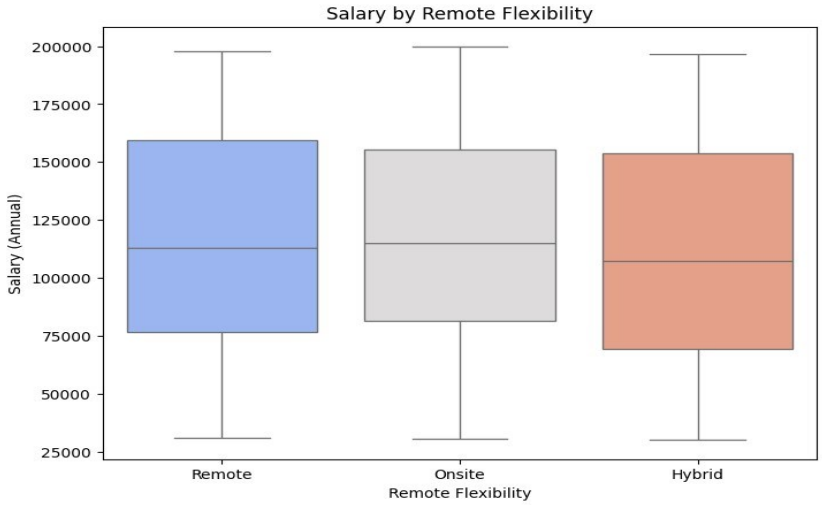

Fig. 4. Salary by Remote Flexibility (Boxplot)

Average job satisfaction scores by technological stack are illustrated in *fig. 5*. Employees using stacks such as Python, Go, and Kubernetes report significantly higher satisfaction levels, while roles based on older or more standardized stacks like JavaScript or .NET show lower satisfaction averages, potentially due to project maturity or limited innovation.

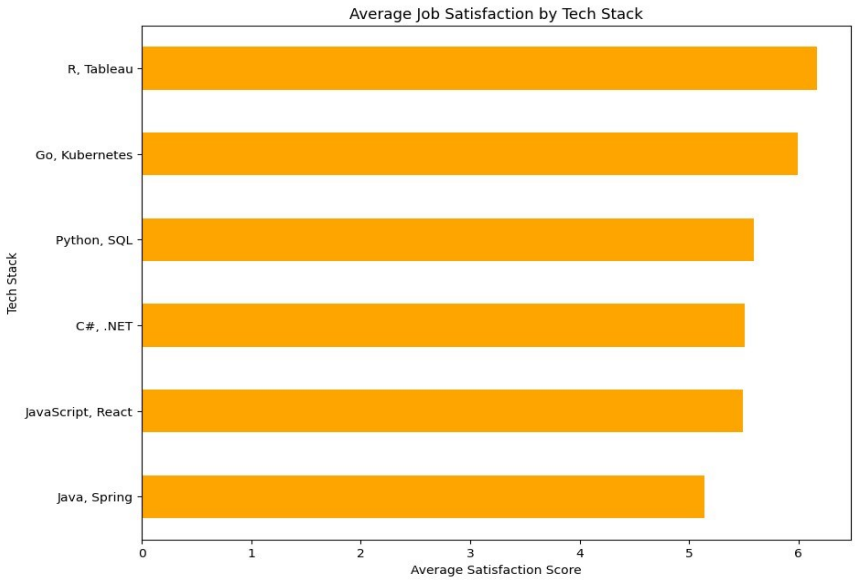

Fig. 5. Job Satisfaction by Tech Stack (Bar Plot)

The distribution of employment contracts in the dataset is illustrated in *fig. 6*. Full-time positions represent the majority, reflecting companies' tendency to maintain long-term employment relationships even in remote settings. Part-time and contract roles are also present but account for smaller segments of the workforce.

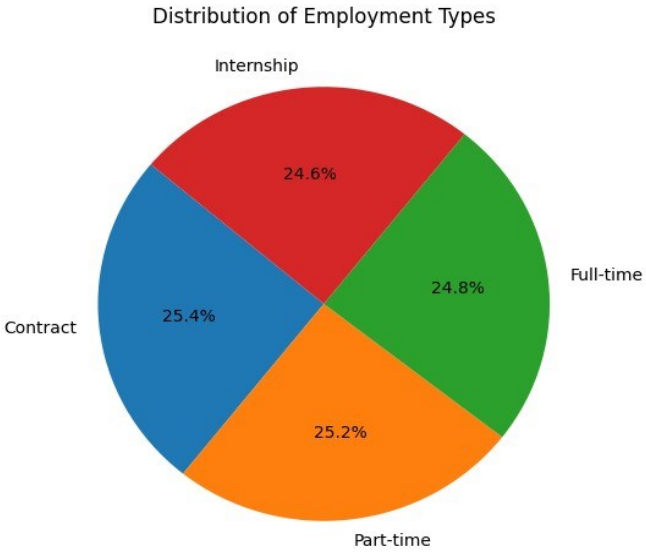

Fig. 6. Employment Type Distribution (Pie Chart)

The correlations among key numerical variables—including annual salary, years of experience, job satisfaction score, and time since last promotion—are illustrated in *fig. 7*. As expected, years of experience positively correlates with salary, while job satisfaction shows weak correlation with both salary and promotion recency, suggesting that factors such as work-life balance and team culture may have a stronger influence on satisfaction.

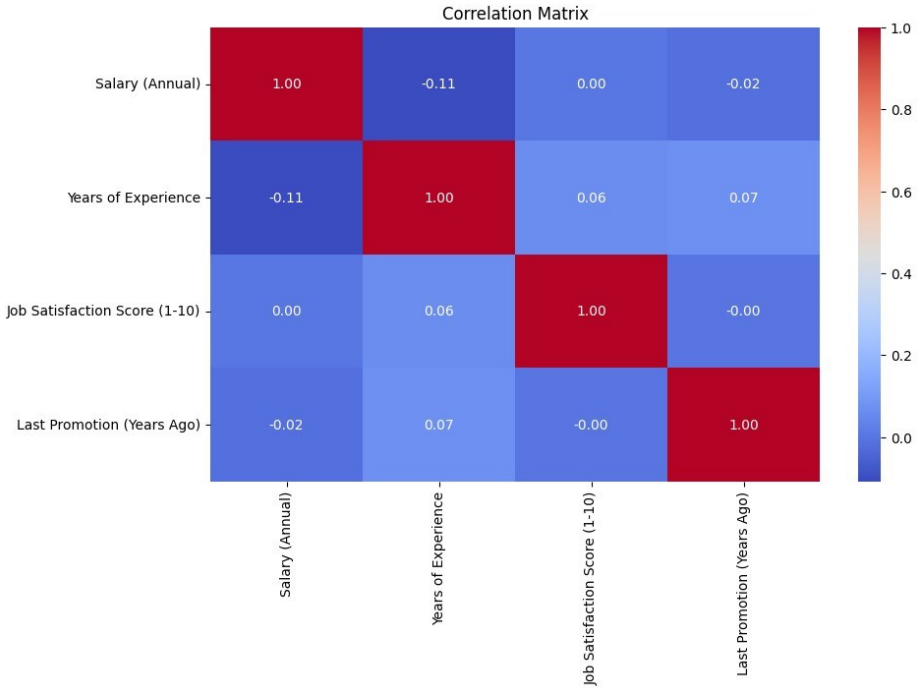

Fig. 7. Correlation Matrix (Heatmap)

5. CONCLUSIONS

In conclusion, the analysis highlights the industries that offer the highest salaries in remote work, the impact of experience and promotions on compensation, and the connection between work flexibility and employee satisfaction.

REFERENCES

[1] Globalization Partners, *Work from Anywhere: The Evolution of Global RemoteWork*, https://www.globalization-partners.com, 2022.

[2] Buffer, *State of Remote Work 2023*, https://buffer.com/stateof-remote-work, 2023..

[3] Glassdoor Economic Research, *Remote Work: What We've Learned So Far*, https://www.glassdoor.com/research, 2022.

[4] OECD, *Teleworking in the COVID-19 pandemic: Trends and prospects*, OECD Publishing, https://doi.org/10.1787/5fad3b8e-en, 2021.

[5] The Pandas Development Team, *pandas-dev/pandas: Powerful Python data analysis toolkit (v2.1)*, Zenodo, https://doi.org/10.5281/zenodo.3509134, 2023.

[6] J. D. Hunter, Matplotlib: *A 2D Graphics Environment*, Computing in Science& Engineering, vol. 9, no. 3, pp. 90–95, doi: 10.1109/MCSE.2007.55, 2007.

[7] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

Received: October 2024, Accepted: December 2025, Published: December 2025
Digital Object Identifier: https://doi.org/10.34302/CJEE/LRBT6157

ENHANCING ISLANDING DETECTION IN PV-BASED DISTRIBUTED SYSTEMS USING CEEMDAN AND PATTERN RECOGNITION NEURAL NETWORK

Sulayman **KUJABI,** Emmanuel Asuming **FRIMPONG,** Francis Boafo **EFFAH**
Department of Electrical and Electronic Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
kujabisaul@yahoo.com

Keywords: **Distributed generation, islanding detection, empirical mode decomposition, pattern recognition neural network, CEEMDAN**.

Abstract: *This paper presents an enhanced and practically validated islanding detection framework for grid-connected solar PV systems, integrating Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and a Pattern Recognition Neural Network (PANN). The method processes negative sequence voltage signals at the Point of Common Coupling (PCC) to extract intrinsic mode functions (IMFs), mitigating mode mixing and improving signal fidelity. Significant IMFs are selected based on their power percentile, and three statistical features—maximum value, standard deviation, and entropy—are extracted and normalized before classification by the PANN. Unlike prior studies, this work extends evaluation to zero-power mismatch scenarios, noisy environments, and load-switching conditions, providing practical validation of real-time detection performance. Simulation results demonstrate a classification accuracy of 98.6% with a detection time of 0.1806 seconds, complying with IEEE 1547 standards. The proposed approach exhibits robust and reliable islanding detection across diverse operating conditions, significantly reducing the non-detection zone (NDZ) and enhancing the safety and reliability of modern distribution systems.*

1. INTRODUCTION

The increasing penetration of solar photovoltaic (PV) systems into modern distribution networks has introduced significant operational and protection challenges. Among these, islanding detection remains one of the most critical issues. Islanding occurs when a portion of

the power system continues to be energized by distributed generators (DGs) after being disconnected from the main utility grid[1],[2]. Failure to detect islanding conditions promptly may compromise personnel safety, damage equipment, and degrade power quality. Therefore, fast and reliable islanding detection mechanisms are essential to ensure compliance with IEEE 1547 standards and maintain grid stability.

Conventional islanding detection methods (IDMs) are generally classified as passive, active, or hybrid approaches. Passive methods monitor electrical parameters such as voltage magnitude, frequency deviation, and harmonic distortion. Although relatively simple and cost-effective, these methods often suffer from large non-detection zones (NDZs), particularly under small power mismatch conditions[3,4]. Active methods introduce controlled perturbations into the system to provoke detectable changes during islanding events. While this approach reduces the NDZ, it may adversely affect power quality and increase implementation complexity[4],[5]. Hybrid methods attempt to combine the advantages of passive and active techniques; however, they still face challenges related to reliability, coordination, and practical deployment[6].

Recent advancements in signal processing and artificial intelligence (AI) have enabled more effective islanding detection. Techniques such as Fourier Transform, Wavelet Transform, Hilbert–Huang Transform (HHT), and Empirical Mode Decomposition (EMD) have improved feature extraction from voltage signals [7],[8],[9]. Similarly, machine learning approaches, including artificial neural networks (ANN), support vector machines (SVM), and pattern recognition neural networks (PANN), have demonstrated high accuracy in classifying islanding events[10], [11], [19], [20],[14]

Among advanced decomposition techniques, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) has emerged as an effective solution to the mode-mixing limitations of classical EMD. By adaptively adding noise and ensuring complete signal reconstruction, CEEMDAN provides more reliable intrinsic mode functions (IMFs) for feature extraction. When combined with intelligent classifiers, CEEMDAN-based frameworks offer strong potential for robust islanding detection.

In our previous study[15], a CEEMDAN–PANN-based islanding detection framework was developed using negative sequence voltage measured at the Point of Common Coupling (PCC). Three statistical features derived from selected IMFs—maximum value, standard deviation, and entropy—were used to train the classifier. That study demonstrated high detection accuracy under standard operating conditions and moderate power mismatch scenarios.

However, several practical aspects were not comprehensively investigated in the earlier work, including:

- Zero-power mismatch conditions, where islanding discrimination becomes particularly challenging;
- Extensive fault resistance variations at different grid locations;
- Large-scale load switching disturbances;

- Detailed cross-validation-based robustness assessment;
- Explicit evaluation of real-time detection performance under noisy measurement conditions.

To address these limitations, the present study extends and strengthens the previously proposed CEEMDAN–PANN framework through expanded simulation scenarios and enhanced validation procedures. Specifically, this paper:

1. Evaluates detection performance under zero-power mismatch, representing the most critical NDZ condition.
2. Incorporates noisy signal environments to assess robustness against measurement disturbances.
3. Analyzes fault cases with resistance values ranging from 1 Ω to 70 Ω at varying distances from the PCC.
4. Examines load switching magnitudes up to 20 MW to ensure reliable discrimination from islanding events.
5. Assesses detection time performance to confirm compliance with IEEE 1547 requirements.

By broadening the validation domain and strengthening performance evaluation, this study enhances the practical applicability and robustness of the CEEMDAN–PANN methodology for real-world PV-based distributed generation systems.

The remainder of this paper presents the system modeling, extended simulation scenarios, signal decomposition and feature extraction methodology, classifier training and validation results, and a comparative performance analysis with existing islanding detection approaches.

2. METHODOLOGY

2.1. CEEMDAN-Based Signal Decomposition

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is used to decompose the negative sequence voltage signal into Intrinsic Mode Functions (IMFs). CEEMDAN improves upon classical EMD and EEMD by adaptively adding Gaussian noise at each decomposition step, ensuring accurate extraction of IMFs and eliminating mode mixing.

The CEEMDAN decomposition process is as follows[16]:

1. Add white Gaussian noise to the original signal:

$$x_i(t) = x(t) + \varepsilon_0 n_i(t) \tag{1}$$

where, ε_0=0.6 is the noise standard deviation amplitude of the added white noise, $x(t)$: Original signal, $n_i(t)$ is the i-th noisy realization.

2. Extract the first IMF from the noisy signal using EMD and average over 100 realizations:

$$IMF_1(t) = \frac{1}{N}\sum_{i=1}^{N} IMF_{1,i}\,(t) \tag{2}$$

3. Residual Calculation:

$$r_1(t) = x(t) - IMF_1(t) \tag{3}$$

4. Repeat adaptive noise addition and IMF extraction iteratively until the residual becomes monotonic or the maximum number of IMFs is reached:

$$x(t) = \sum_{j=1}^{M} IMF_j(t) + r(t) \tag{4}$$

5. Significant IMFs selection: The power of each IMF is computed as:

$$P_i = \sum_{t=1}^{N}[C_i(t)]^2 \tag{5}$$

IMFs with power above the 90th percentile are selected for feature extraction.

2.2. Feature Extraction and Selection

From each significant IMF, three statistical features are extracted:

1. Maximum value:

$$Maximum = max(IMF)$$

2. Standard deviation:

$$standard\ deviation = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

3. Entropy:

$$entropy = -\sum_{i=1}^{N} p(x_i)\ log_2\big(p(x_i)\big)$$

where, $p(x_i)$ is the probability of each unique IMF value.

2.3. Feature Normalization

The extracted features are normalized using min-max scaling to ensure uniform magnitude and facilitate PANN training:

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (6)$$

X_{min} *and* X_{max} are the minimum and maximum values of the feature across the dataset.

2.4. PANN Classifier

The Pattern Recognition Neural Network (PANN) structure in *figure 1* is used for islanding detection. The network consists of:

- **Input layer**: 3 neurons (for max, standard deviation, entropy)
- **Hidden layer**: 3 neurons
- **Output layer**: 1 neuron (0 = non-islanding, 1 = islanding)

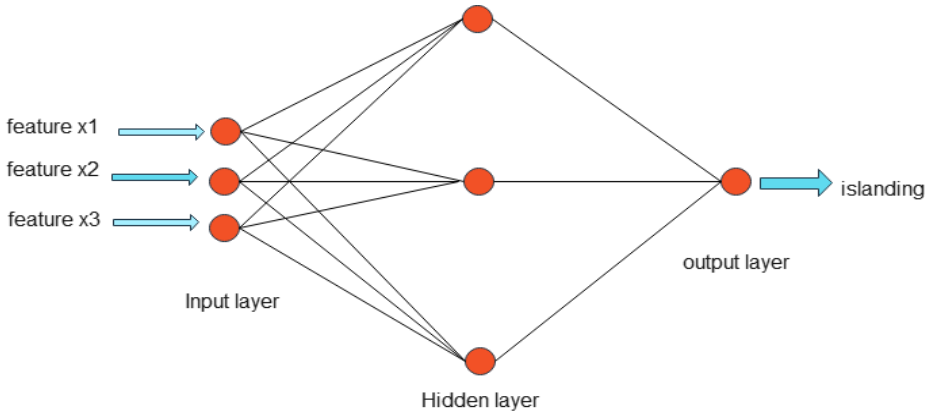

Fig. 1. PANN structure with 3 inputs

The PANN is trained using supervised learning and backpropagation, adjusting weights to minimize prediction errors. The classifier learns patterns and relationships in the features to distinguish islanding from non-islanding events effectively.

2.5. Proposed Detection Algorithm

The overall algorithm is summarized as follows:

1. Acquire negative sequence voltage data from islanding and non-islanding scenarios at the PCC.
2. Add white Gaussian noise (std = 0.6) to enhance decomposition robustness.
3. Apply CEEMDAN across 100 noisy realizations to extract IMFs.

4. Compute IMF power and select significant IMFs (≥90th percentile).
5. Extract max, standard deviation, and entropy from significant IMFs.
6. Normalize the features using min-max scaling.
7. Train the PANN classifier on the extracted features.
8. Classify new voltage signals to detect islanding events.

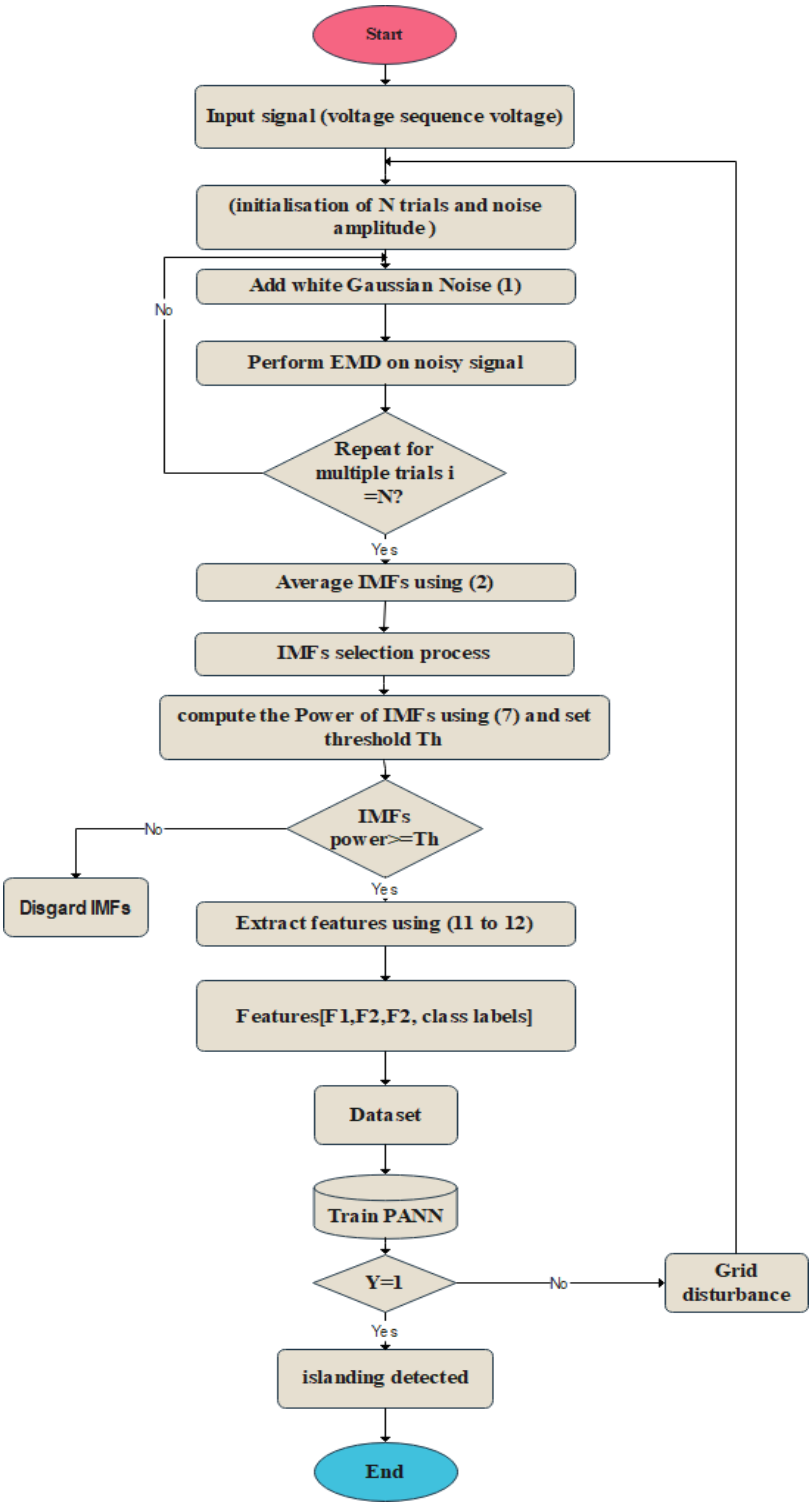

Fig. 2. Flow chart for the proposed islanding detection

Figure 2 illustrates the CEEMDAN–PANN-based islanding detection framework, showing steps from signal acquisition to classification.

3. RESULTS AND DISCUSSION

3.1. Test System for Islanding Detection

The proposed CEEMDAN–PANN islanding detection method was evaluated using a standard test system widely employed in the literature for validating islanding detection algorithms[17],[18]. The system was modeled in MATLAB/Simulink, incorporating realistic grid conditions, distributed solar PV generation, and various load types. *Figure 3* illustrates the studied distributed generation system.

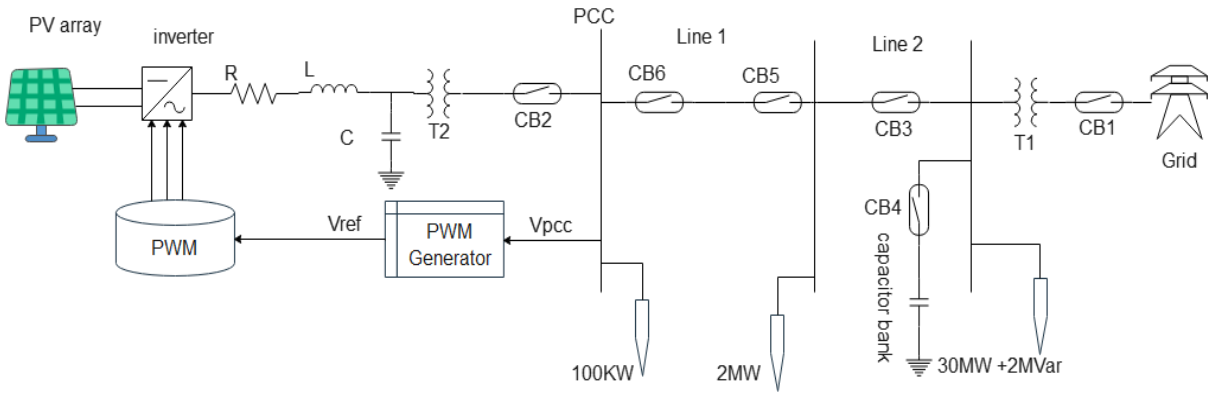

Fig. 3: The studied distributed power generation system

Key components of the system include:

- ✓ **PV-DG unit**: Represents solar photovoltaic generation connected to the grid.
- ✓ **Loads**: Both local and grid-connected loads of varying magnitude and characteristics.
- ✓ **Utility grid and circuit breakers**: Enable simulation of disconnection events and islanding conditions.
- ✓ **Point of Common Coupling (PCC):** Location where negative sequence voltage is measured for monitoring and analysis.

A total of 294 distinct disturbance scenarios were simulated (Table 1), encompassing islanding, fault, and load-switching conditions. The negative sequence voltage was recorded at a sampling frequency of 3.84 kHz over 2.5 s, capturing both transient and steady-state characteristics necessary for robust islanding detection.

Table 1 Simulated grid disturbances

labels	scenarios	Scenarios description	Number of tests
C1	islanding	Tripping circuit breaker during different power mismatch	144
C2	Non islanding	Fault events for 3-phase, 2-phase and single phase	132
C3	Non islanding	Switching of local and grid loads	18

Name of system parameters	Specifications
PV system	Module: sun-power Module model: SunPower SPR-305E-WHT-D Modules in series: 5 No of parallel string: 66 parallel strings PV power rating: 100.7kw Reference voltage: 500V DC Inverter nominal frequency: 60Hz Frequency of the PWM carrier: 33x60Hz Voltage integral and proportional gain: *Ki*:800, *Kp*:7 Current integral and proportional gain: *Ki*:20, *Kp*:0.3
Electric power grid	Rating: 120Kv, and 2500MVA
Transmission line	Resistance: R=0.413 Inductance: L=3.32x10^{-3}H Capacitance: C=5.01x10^{-9}F Rating: $L1$=100kW, $L2$=2MW, $L3$= 30MW+j2MVAr Line voltage: 25KV Length of the line: line-1 is 14km, and line-2 is 5km
Transformer	Voltage level:120kV/25kV Rating: 47MVA for T1, 100kVA, 25kV/0.67kV for T2

3.2. Islanding Scenario (C1)

Islanding events were simulated by tripping the grid-side circuit breaker. *Figure 4* illustrates islanding of the negative sequence voltage response during various power mismatch conditions.

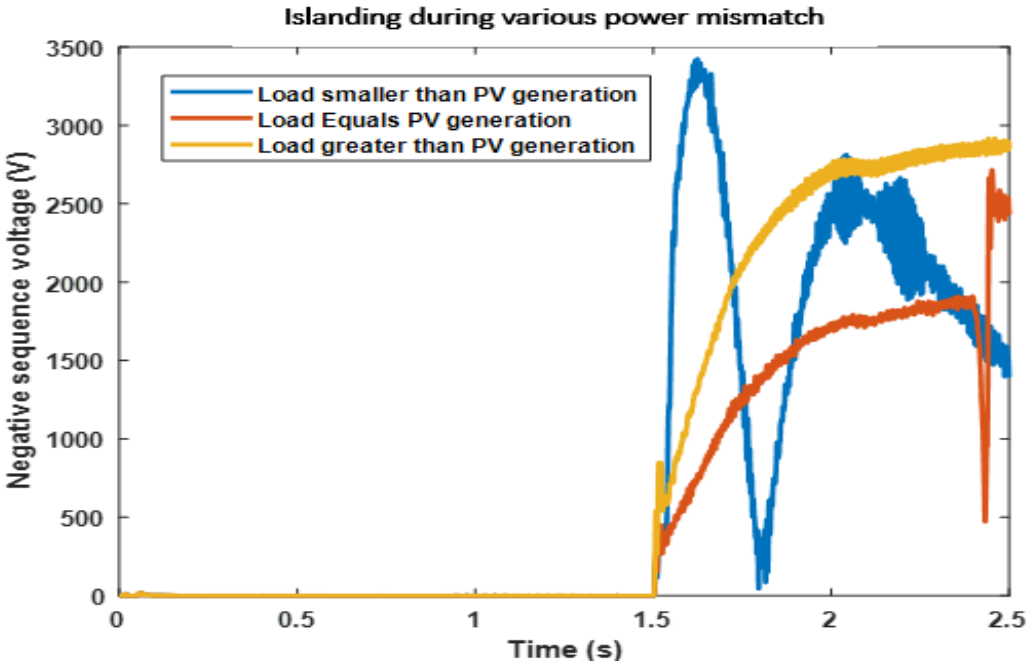

Fig. 4. Islanding during various power mismatch

Under zero-power mismatch, the signal and its IMFs (*Figure 5*) exhibit minimal variation, making differentiation from other disturbances challenging. Three IMFs were extracted:

- ✓ IMF1: Highest frequency content, capturing rapid transient changes.
- ✓ IMF2 & IMF3: Lower frequency content, reflecting slower system dynamics crucial for detection.

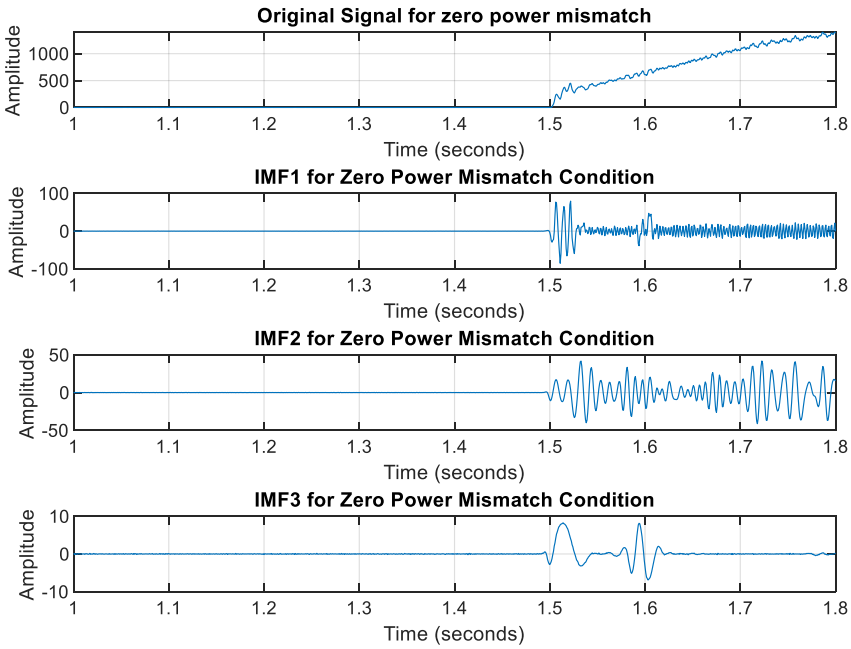

Fig. 5. Islanding case of Negative sequence voltage and IMFs during zero power mismatch

With a 50% power mismatch, larger voltage variations are observed (*Figure 6*), improving discrimination between islanding and non-islanding events. IMF1 captures fast transients, while IMF2 and IMF3 reflect slower dynamics, highlighting the impact of power mismatch on islanding detection.

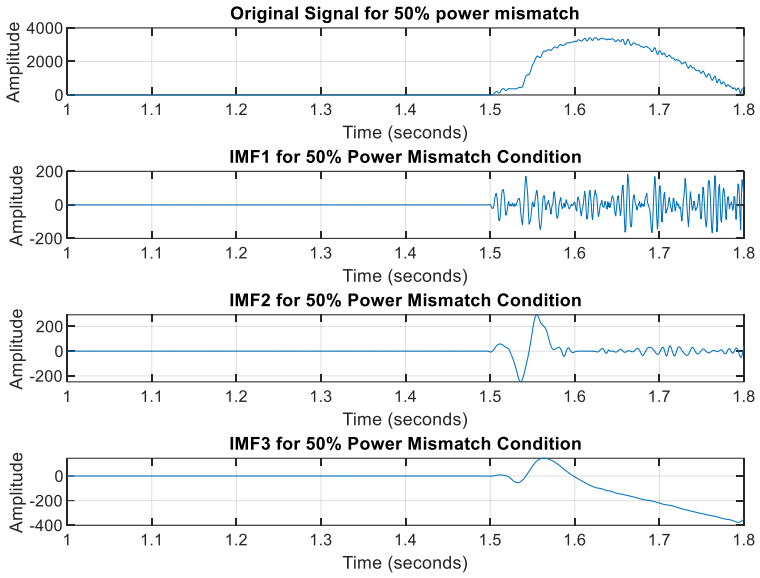

Fig. 6. Islanding case of negative sequence voltage and IMFs @50% power mismatch

3.3. Fault Scenario (C2)

Various faults, including single-phase, two-phase, and three-phase faults, were simulated at different distances from the PCC (5 km and 14 km). Resistance values ranged from 1 Ω to 70 Ω and were cleared after 300 ms.

Figures 7–9 illustrate negative sequence voltage responses for different fault types. Lower voltages at higher fault resistance complicate islanding detection.

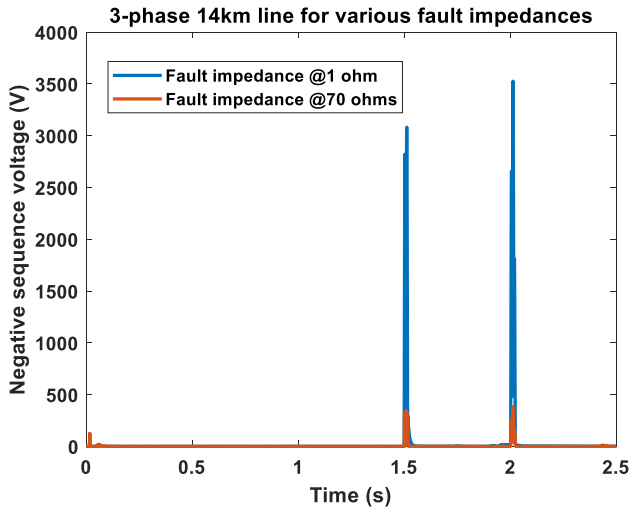

Fig. 7. 3-phase fault inception time 1.5 seconds negative sequence voltage waveform

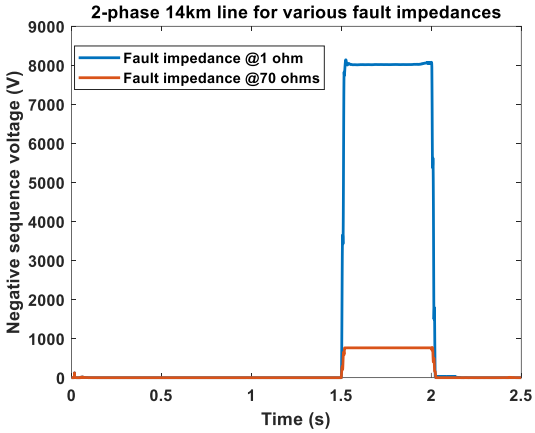

Fig. 8. 2-phase fault inception time 1.5 seconds negative sequence voltage waveform

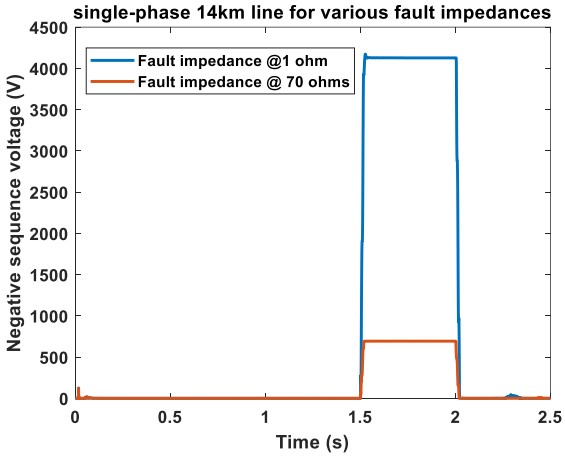

Fig. 9. 1-phase fault inception time 1.5 seconds negative sequence voltage waveform

CEEMDAN was applied to extract IMF features from the 1.5–1.8 s window around the fault event (*Figure 10*).

IMF1 captures high-frequency transient response while IMF2 & IMF3: Reveal lower-frequency components essential for distinguishing faults from islanding.

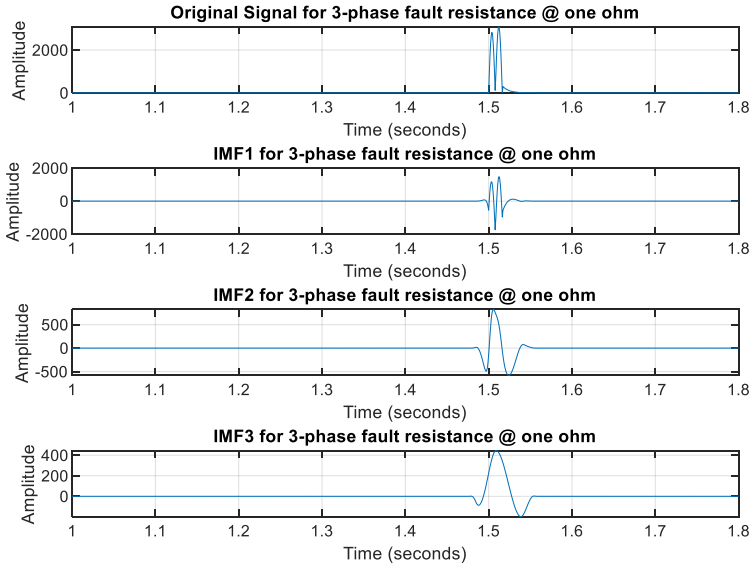

Fig. 10. 3-phase grid fault and IMFs for a fault resistance of one ohm

3.4. Load Switching Scenario (C3)

Load switching events were simulated at t = 1.5 s, with capacities ranging from 10 MW to 20 MW. *Figures 11–12* display voltage variations at the PCC for local and grid load switching.

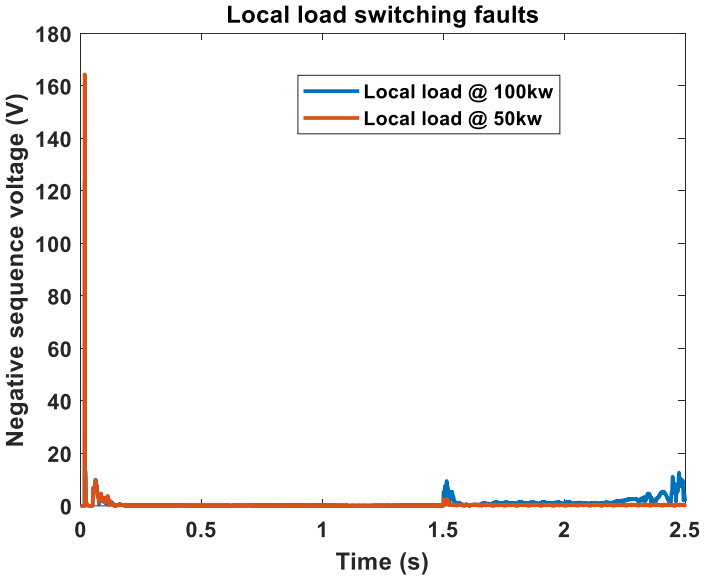

Fig. 11. Local load switching fault

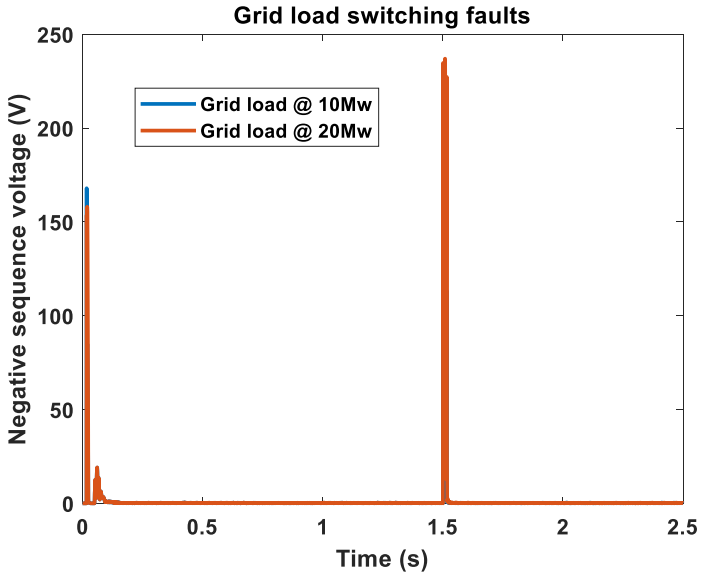

Fig. 12. Grid load switching fault

IMFs corresponding to different load magnitudes (*Figures 13–14*) demonstrate that larger load changes induce more significant negative sequence voltage variations, which are critical for detecting disturbances that could be misinterpreted as islanding.

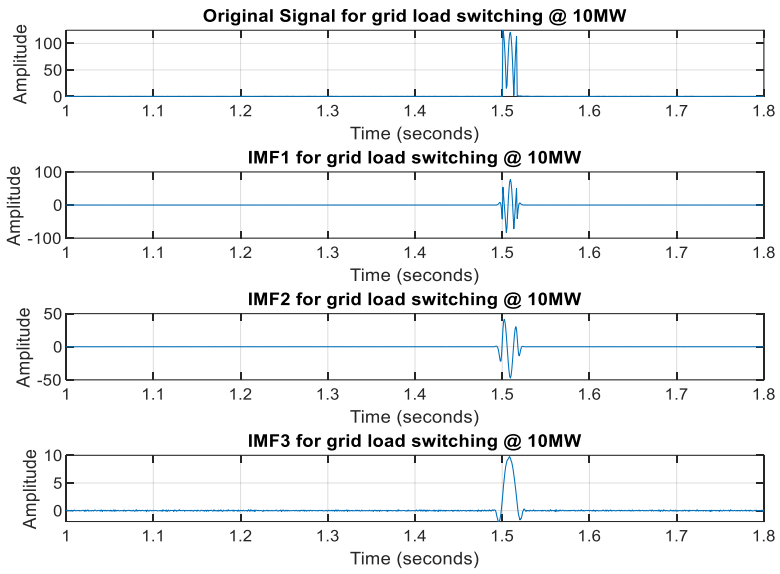

Fig. 13. Switching fault and IMFs for various loading effect(10MW)

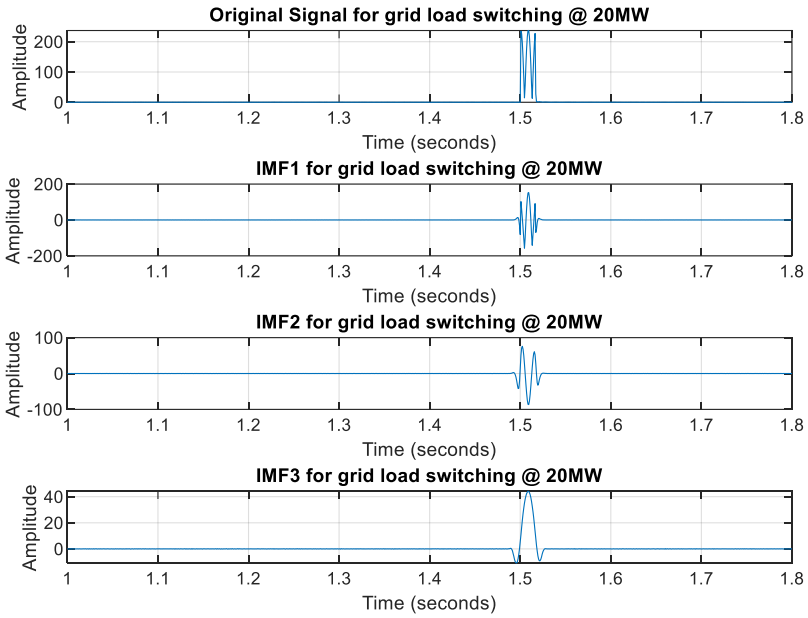

Fig. 14. Switching fault and IMFs for various loading effect(20MW)

3.5. Feature Analysis for Islanding Detection

Three statistical features were extracted from the IMFs for islanding detection: normalized maximum value, standard deviation, and entropy.

Normalized Max (*Figure 15*) has higher during islanding events, lower for non-islanding scenarios.

Normalized Standard Deviation (*Figure 16*) has larger deviations during islanding.

Normalized Entropy (*Figure 17*) has lower entropy in islanding conditions, higher in non-islanding events.

These features provided clear discrimination between islanding and non-islanding scenarios, forming the input for the PANN classifier.

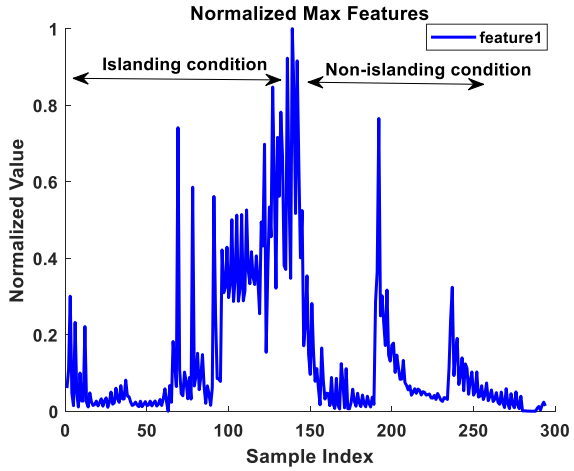

Fig. 15. Normalized max feature analysis

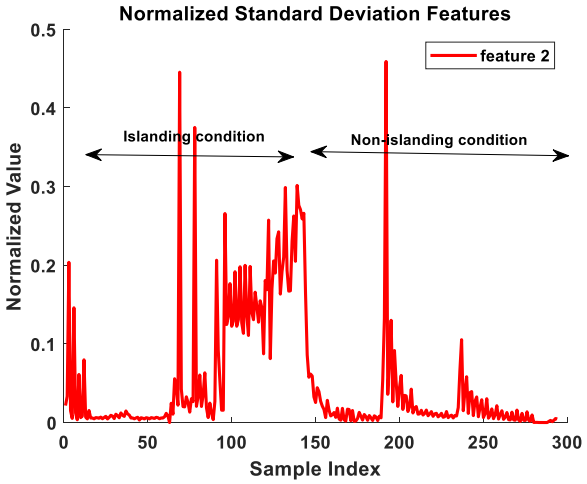

Fig. 16. Normalized std features analysis

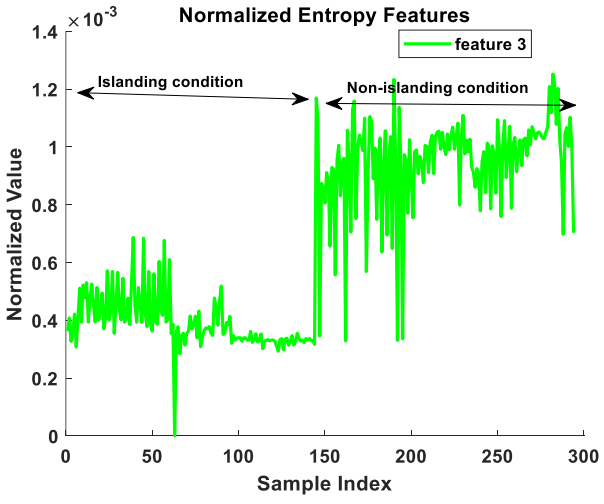

Fig. 17. Normalized entropy features analysis

3.6. PANN Training and Performance

The PANN with a single hidden layer of 3 neurons was trained over 24 iterations. Performance metrics included: Training accuracy: 96.4%, Validation accuracy: 97.1%, Testing accuracy: 100%, Overall accuracy: 97.0%.

Figure 18 shows regression performance, and *Figure 19* highlights cross-entropy loss reduction to 0.09399 at epoch 18, indicating effective feature learning and model convergence.

Training and testing times were 4109.85 ms and 18.97 ms, respectively, demonstrating real-time applicability.

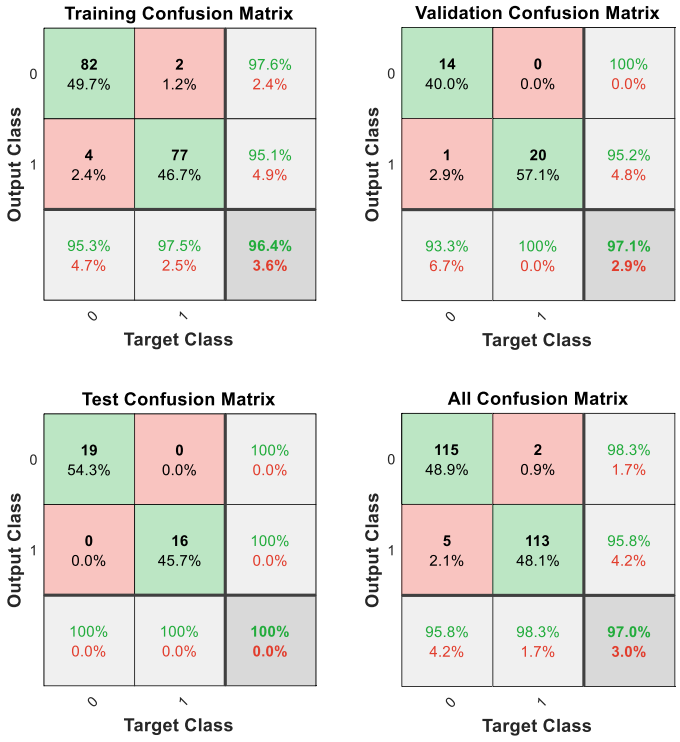

Fig. 18. Performance of CEEMDAM-PANN training with single hidden layer 3

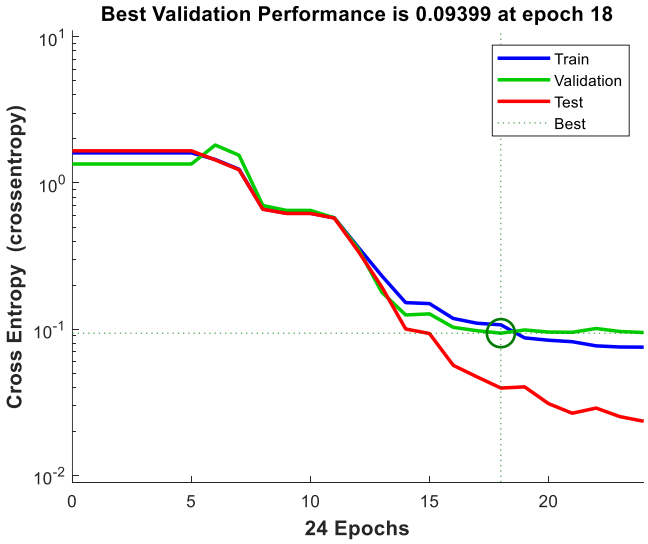

Fig. 19. Performance of CEEMDAN-PANN training with single hidden layers 3 neurons

3.7. 5-Fold Cross-Validation Performance

The model was evaluated using 5-fold cross-validation. *Figure 20* illustrates fold-wise training and testing accuracies. Table 4 summarizes performance metrics:

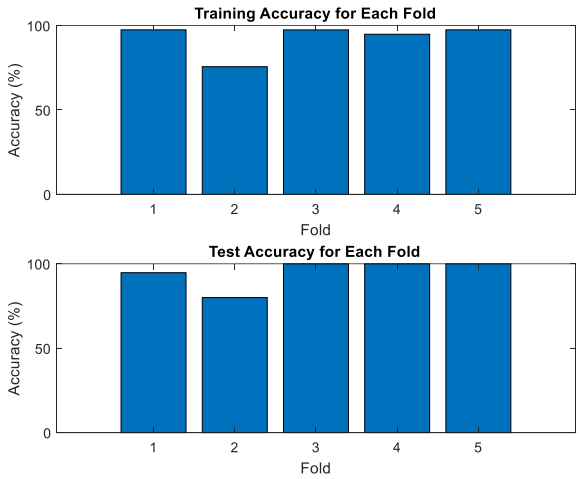

Fig. 19. Training and test accuracy of CEEMDAN-PANN on 5-fold cross validation

Table 2 cross validation Performance of the CEEMDAN-PANN islanding detection

Fold	Accuracy (%)	Precision	Recall	F1_score
1	94.83	0.97	0.93	0.95
2	98.61	1	0.97	0.98
3	94.92	0.97	0.93	0.95
4	96.61	0.97	0.97	0.97
5	98.31	1	0.97	0.98

These results indicate high robustness and generalization across different dataset subsets.

3.8. Classification Results

The CEEMDAN–PANN classifier achieved: 100% detection for islanding events and 96.67% detection for non-islanding events (Table 3, *Figure 21*).

Table 3 CEEMDAN-PANN classification result with IMFs

classes	No of cases	Correct detection	accuracy
Non-islanding	30	29	96.67
islanding	29	29	100

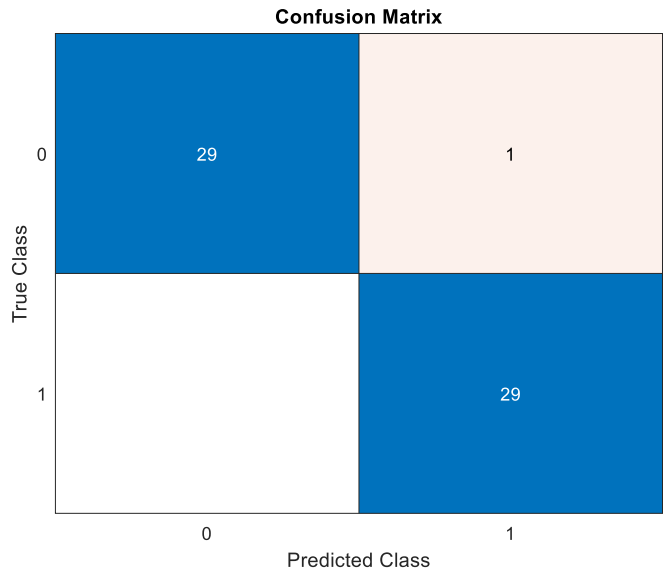

Fig. 20. CEEMDAN-PANN confusion matrix for islanding condition

3.9. Comparison with Existing Methods

The proposed method outperformed recent approaches in both accuracy and detection speed (Table 5):

Table 4 comparison with existing methods

Reference	**Signal processing**	**classifier**	**Accuracy (%)**
[19]	Wavelet transform	ANN	91.43
[20]	Wavelet transform	Decision tree	97.9
[21]	S-transform	KNN	98.1
[22]	Wavelet-singular spectrum entropy	Deep learning	98.4
Proposed method	CEEMDAN	Pattern artificial neural network	98.6

Detection time of 0.1806 seconds, suitable for real-time applications

4. CONCLUSION

This study presents a robust CEEMDAN–PANN-based islanding detection method for solar PV-distributed generation systems. By leveraging negative sequence voltage, selecting significant IMFs, and extracting max, standard deviation, and entropy features, the approach achieved high detection accuracy (98.6%) and fast response (0.1806 s).

Compared to conventional methods, the proposed framework reduces the non-detection zone and ensures reliable, real-time identification of islanding events under power mismatches, load switching, and noisy environments. Future work will focus on field implementation and optimization for evolving smart grid scenarios.

REFERENCES

[1] A. K. Hamid, N. T. Mbungu, A. Elnady, R. C. Bansal, A. A. Ismail, and M. A. AlShabi, "A systematic review of grid-connected photovoltaic and photovoltaic/thermal systems: Benefits, challenges and mitigation," *Energy Environ.*, vol. 34, no. 7, pp. 2775–2814, 2023, doi: 10.1177/0958305X221117617.

[2] S. K. G. Manikonda and D. N. Gaonkar, "Comprehensive review of IDMs in DG systems," *IET Smart Grid*, vol. 2, no. 1, pp. 11–24, 2019, doi: 10.1049/iet-stg.2018.0096.

[3] A. Serrano-Fontova, J. A. Martinez, P. Casals-Torrens, and R. Bosch, "A robust islanding detection method with zero-non-detection zone for distribution systems with DG," *Int. J. Electr. Power Energy Syst.*, vol. 133, p. 107247, 2021, doi: 10.1016/j.ijepes.2021.107247.

[4] P. Systems, "IEEE standards," *IEEE Spectr.*, vol. 13, no. 5, pp. 108–108, 2013, doi: 10.1109/mspec.1976.6369202.

[5] R. Bakhshi-Jafarabadi, J. Sadeh, A. Serrano-Fontova, and E. Rakhshani, "Review on islanding detection methods for grid-connected photovoltaic systems, existing limitations and future insights," *IET Renew. Power Gener.*, vol. 16, no. 15, pp. 3406–3421, 2022, doi: 10.1049/rpg2.12554.

[6] S. Perlenfein, M. Ropp, J. Neely, S. Gonzalez, and L. Rashkin, "Subharmonic power line carrier (PLC) based island detection," *Conf. Proc. - IEEE Appl. Power Electron. Conf. Expo. - APEC*, vol. 2015-May, no. May, pp. 2230–2236, 2015, doi: 10.1109/APEC.2015.7104659.

[7] O. Tshenyego, R. Samikannu, and B. Mtengi, "Wide area monitoring, protection, and control application in islanding detection for grid integrated distributed generation: A review," *Meas. Control (United Kingdom)*, vol. 54, no. 5–6, pp. 585–617, 2021, doi: 10.1177/0020294021989768.

[8] N. Gupta, R. Dogra, R. Garg, and P. Kumar, "Review of islanding detection schemes for utility interactive solar photovoltaic systems," *Int. J. Green Energy*, vol. 19, no. 3, pp. 242–253, 2022, doi: 10.1080/15435075.2021.1941048.

[9] N. A. Larik, M. F. Tahir, Z. M. S. Elbarbary, M. Z. Yousaf, and M. A. Khan, "A comprehensive literature review of conventional and modern islanding detection methods," *Energy Strateg. Rev.*, vol. 44, no. November, p. 101007, 2022, doi: 10.1016/j.esr.2022.101007.

[10] M. S. Thomas and P. P. Terang, "Islanding detection using decision tree approach," *2010 Jt. Int. Conf. Power Electron. Drives Energy Syst. PEDES 2010 2010 Power India*, pp. 1–6, 2010, doi: 10.1109/PEDES.2010.5712394.

[11] A. Ezzat, B. E. Elnaghi, and A. A. Abdelsalam, "Microgrids islanding detection using Fourier transform and machine learning algorithm," *Electr. Power Syst. Res.*, vol. 196, no. February, p. 107224, 2021, doi: 10.1016/j.epsr.2021.107224.

[12] P. K. Dash, M. Padhee, and T. K. Panigrahi, "Electrical Power and Energy Systems A hybrid time – frequency approach based fuzzy logic system for power island detection in grid connected distributed generation," *Int. J. Electr. Power Energy Syst.*, vol. 42, no. 1, pp. 453–464, 2012, doi: 10.1016/j.ijepes.2012.04.003.

[13] Himanshu Jyoti Saikia, "Development of a Passive Islanding Detection Method based Fuzzy Controller for a Grid Connected Distributed Generator.," *Int. J. Eng. Res.*, vol. V5, no. 12, pp. 308–311, 2016, doi: 10.17577/ijertv5is120242.

[14] M. I. Dieste-Velasco, "Application of a pattern-recognition neural network for detecting analog electronic circuit faults," *Mathematics*, vol. 9, no. 24, 2021, doi: 10.3390/math9243247.

[15] S. Kujabi, E. A. Frimpong, and F. B. Effah, "Energy-Efficient Islanding Detection Using CEEMDAN and Neural Network Integration in Photovoltaic Distribution System." [Online]. Available: https://www.ijeca.info

[16] L. Zhao, Z. Li, J. Zhang, and B. Teng, "An Integrated Complete Ensemble Empirical Mode Decomposition with Adaptive Noise to Optimize LSTM for Significant Wave Height Forecasting," *J. Mar. Sci. Eng.*, vol. 11, no. 2, 2023, doi: 10.3390/jmse11020435.

[17] S. Paul, "Islanding Detection in Grid-Connected 100 KW Photovoltaic System Using Wavelet Transform," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 5, no. 5, pp. 92–98, 2016.

[18] P. Jaiswal, S. K. Srivastava, and K. B. Sahay, "Modeling and simulation of proposed 100 KW solar PV array power plant for MMMUT Gorakhpur," *Int. Conf. Emerg. Trends Electr. Electron. Sustain. Energy Syst. ICETEESES 2016*, pp. 261–266, 2016, doi: 10.1109/ICETEESES.2016.7581391.

[19] Z. Guan and Y. Liao, "A New islanding detection method based on wavelet-Transform and ann for micro-grid including inverter assisted distributed generator," *Int. J. Emerg. Electr. Power Syst.*, vol. 20, no. 5, pp. 1–10, 2019, doi: 10.1515/ijeeps-2019-0074.

[20] M. A. Khan, A. Haque, and V. S. B. Kurukuru, "Machine Learning Based Islanding Detection for Grid Connected Photovoltaic System," *2019 Int. Conf. Power Electron. Control Autom. ICPECA 2019 - Proc.*, vol. 2019-Novem, no. 1, 2019, doi: 10.1109/ICPECA47973.2019.8975614.

[21] M. Mishra, C. K. Patra, P. K. Muni, D. A. Gadanayak, and T. Parida, "Islanding detection in distributed generation system based on optimized KNN utilizing S-transform based features," *2023 Int. Conf. Adv. Power, Signal, Inf. Technol.*, pp. 41–47, 2023, doi: 10.1109/APSIT58554.2023.10201758.

[22] X. Kong, X. Xu, Z. Yan, S. Chen, H. Yang, and D. Han, "Deep learning hybrid method for islanding detection in distributed generation," *Appl. Energy*, vol. 210, no. August 2017, pp. 776–785, 2018, doi: 10.1016/j.apenergy.2017.08.014.