Sukkur IBA University

SJET
Sukkur IBA Journal of Emerging Technologies

# Sukkur IBA Journal of Emerging Technologies

**Recognized in HEC Pakistan "Y" Category**

*P-ISSN: 2616-7069 E-ISSN: 2617-3115*

Volume: 5 | No.: 2 | Jul - Dec | 2022

**Sukkur IBA Journal of Emerging Technologies (SJET)** is a bi-annual research journal published by **Sukkur IBA University**, Pakistan. **SJET** is dedicated to serve as a key resource to provide applied engineering research associated with the Electrical, Electronics, and innovations in Energy at the global scale. This Journal publishes manuscripts that are well written by highlighting development in emerging technologies. This journal covers all branches of Engineering, Science & Emerging Technologies.



**Disclaimer:** The research material expressed in the **Sukkur IBA Journal of Emerging Technologies (SJET)** is the sole contribution of the authors. The research contribution of the authors does not reflect the management, advisory board, the editorial board, **Sukkur IBA University** press, and the organization to which the authors are affiliated. Manuscripts published in **SJET** shall be processed through double-blind peer-reviewed by the two experts of the field. The identities of the experts/reviewers shall remain anonymous to the authors. The Journal shall be published two issues in a year. Neither the **Sukkur IBA University** nor the **SJET** shall be responsible for errors or any consequences highlighted by the reader. The errors and deficiencies in terms of research in the manuscript can directly be reported to the authors.

==========================================================================

Publisher: **Sukkur IBA Journal of Emerging Technologies (SJET)**

**Office of Research, Innovation & Commercialization – ORIC**

**Sukkur IBA University -** Airport Road Sukkur-65200, Sindh Pakistan

Tel: (09271) 5644233 -37 Fax: (092 71) 5804425 Email: sjet@iba-suk.edu.pk URL: sjet.iba-suk.edu.pk

==========================================================================

***Mission Statement***

The mission of **Sukkur IBA Journal of Emerging Technologies (SJET)** is to provide a premier interdisciplinary platform to researchers, scientists, and practitioners from the field of engineering in particular, electrical, electronics, renewable, and emerging engineering fields for the dissemination of their finding and to contribute in the knowledge domain.

**Aims & Objectives**

**Sukkur IBA Journal of Emerging Technologies (SJET)** will publish and encourage the submission of critically reviewed manuscripts on cutting-edge research in the field of emerging engineering technologies.

The objectives of **SJET** are:

1. To bring new engineering ideas, research, and findings on a single platform.
2. To integrate interdisciplinary research for a technological sustainable solution.
3. To provide a scholarly platform to connect academia and industries for socio-economic development.

***Research Themes***

The research focused on but not limited to the following core thematic areas:

**Renewable Energy Sources and Storage:**

- Solar energy system fabrication and construction of advanced fuel cell technology
- Designing and analyzing smart hydro and wind energy systems
- Developing systems for biomass and bio-fuels
- Energy management and storage
- Energy devices and materials
- Energy harvesting for wireless and body sensor networks
- Energy efficiency and policies
- Energy devices and materials

**Power Systems and Smart Grids:**

- Power Quality Issues and solutions
- Microgrid systems and their Integration Problems
- Design control and management
- Energy management and Environmental issues
- Hybrid power system
- Distributed and co-generation systems
- Power market and power system economics

**Electrical Machines and Adjustable Speed Drives:**

- AC and DC machines and drives
- Sensor-less control
- Piezo and electrostatic actuators
- Machine design and equipment training
- Maintenance and fault diagnosis
- Bearing less driving technologies

==================================================================
Publisher: **Sukkur IBA Journal of Emerging Technologies (SJET)**
**Office of Research, Innovation & Commercialization – ORIC**
**Sukkur IBA University** - Airport Road Sukkur-65200, Sindh Pakistan
Tel: (09271) 5644233 -37 Fax: (092 71) 5804425 Email: sjet@iba-suk.edu.pk URL: sjet.iba-suk.edu.pk
==================================================================

**Power Electronics and its Application:**

- Hard-switching and soft-switching static converters
- Multi-level and matrix converters
- Emerging topologies
- Simulation and control power converters
- Power factor correctors
- Active filters and total harmonics distortions analysis
- Optoelectronics and photonic devices
- Power semiconductors, passive components, and packaging technologies
- Switch-mode power supplies and automotive
- Applications of power electronics in-home appliance

**High Voltage Engineering and Insulation Technology:**

- Micro-electromechanical system (MEMS)
- Power Integrated circuits (PIC)
- Power Engineering related Technologies
- Power system stability and control
- Power system transient modeling, simulation, and analysis
- Electromagnetic transient programs (EMTP)
- HVDC and FACTS applications

**Nanomaterials/Nanotechnology:**

- Sensors and Actuators
- Electronic Thin Films
- Nanogenerators
- Nanomaterials
- Nanotechnology optoelectronic sensors
- magnetic sensors
- thermal sensors
- mechanical sensors

**Communication and Signal Processing:**

- Communication & signal processing
- Radio frequency systems, microwave, and antenna design
- Analog and mixed-signal circuits
- Filter designing
- Satellite communication, mobile communication
- Cognitive and software design radio
- Analog and Mixed-Signal Circuits

**Biomedical Electronics:**

- Energy-efficient wireless body sensor networks
- Wireless power/energy transfer in e-health applications
- Green and battery-friendly wireless medical networks
- Renewable energy and energy harvesting for wireless and body sensor networks
- Telemedicine and medical IoT
- Medical video transmission
- Energy management for medical health applications
- Role of 5G in medical health applications

**Thermal and complex fluid dynamics:**

- Active and passive techniques for fluid flow manipulation

==============================================================================
Publisher: **Sukkur IBA Journal of Emerging Technologies (SJET)**
**Office of Research, Innovation & Commercialization – ORIC**
**Sukkur IBA University** - Airport Road Sukkur-65200, Sindh Pakistan
Tel: (09271) 5644233 -37 Fax: (092 71) 5804425 Email: sjet@iba-suk.edu.pk URL: sjet.iba-suk.edu.pk
==============================================================================

- Fluid flow process for industrial equipment’s
- Modeling of working fluids
- Experimental fluid dynamics
- Multifunctional heat exchangers/chemical reactors
- Energy-efficient combustion
- Environmental fluid flows

**Materials and their processing**

- Piezoelectric materials
- Polymers, metal oxides
- III, V and II, VI semiconductors
- Thick and thin films
- Optical glass fibers
- Amorphous
- Polycrystalline monocrystalline silicon, nanomaterials
- Synthesis of nanomaterials, composite materials
- Functional material
- Electronic thin films and integrated devices
- Engineering materials
- Solid and structural mechanics

## Patron's Message

**Sukkur IBA University** has been imparting education with its core values of merit, quality, and excellence since its inception. Sukkur IBA University has achieved numerous milestones in a very short span of time that hardly any other university has achieved in the history of Pakistan. The institute continuously being ranked as one of the best Institute in Pakistan by the Higher Education Commission (HEC). The distinct service of Sukkur IBA University is to serve rural areas of Sindh and also underprivileged areas of other provinces of Pakistan. Sukkur IBA University is committed to serve the targeted youth of Pakistan who is suffering from poverty and deprived of equal opportunity to seek quality education. Sukkur IBA University is successfully undertaking its mission and objectives that lead Pakistan towards socio-economic prosperity.

In continuation of endeavors to touch new horizons in the field of Engineering and Emerging Technologies, Sukkur IBA University publishes an international referred journal. Sukkur IBA University believes that research is an integral part of modern learnings ad development. **Sukkur IBA Journal of Emerging Technologies (SJET)** is the modest effort to contribute and promote the research environment within the university and Pakistan as a whole. SJET is a peer-reviewed and multidisciplinary research journal to publish findings and results of the latest and innovative research in the field. Following the tradition of Sukkur IBA University, SJET is also aimed at achieving international recognition and high impact research publication in the near future.

**Prof. Dr. Syed Mir Muhammad Shah**
**Vice-Chancellor**
**Sukkur IBA University**
Patron **SJET**

# Editorial

*Dear Readers,*

It is immense pleasure to present you the latest issue of the Sukkur IBA Journal of Emerging Technologies (SJET). Sukkur IBA University firmly believes in the research environment and has provided a platform for the intellectuals and researchers to share knowledge and new findings on emerging trends in various research areas to solve the difficult technical problems related to the technological advancements in response to the demands of the times. The SJET provided an interdisciplinary platform to the researchers' community to collaborate, co-innovate, and instigate efforts to break the technological barriers. This journal provides the opportunity to gain and present authentic and insightful scientific & technological information on the latest advances in the field of emerging technologies.

The SJET provides an invaluable source of information and enables the interested researchers to access the original information they are seeking. The manuscripts submitted in SJET have been followed by a double-blind peer-review process, which addresses key issues in the field of emerging engineering technologies. The SJET has endorsed high standards which are prerequisite for publishing high-quality research work. This journal manifests into an eco-system for the academician and engineers work together in the pursuit of excellence & innovation, that is why the editorial board of SJET is comprised of academic and industrial researchers from various advanced countries. The journal has been recognized by the higher education commission (HEC) of Pakistan under "Y" category. It has adopted an Open access policy without charging any publication fees that will certainly increase the readership by providing free access to a wider audience.

On behalf of the SJET, I welcome the submissions for the upcoming issue and looking forward to receiving your valuable feedback.

I hope this journal will make a difference in our perspective and choice of research.

Sincerely,

**Dr. Saeed Ahmed Khan**
*Chief Editor*
**SJET**

==================================================================
Publisher: **Sukkur IBA Journal of Emerging Technologies (SJET)**
**Office of Research, Innovation & Commercialization – ORIC**
**Sukkur IBA University** - Airport Road Sukkur-65200, Sindh Pakistan
Tel: (09271) 5644233 -37 Fax: (092 71) 5804425 Email: sjet@iba-suk.edu.pk URL: sjet.iba-suk.edu.pk
==================================================================

==================================================================
Publisher: **Sukkur IBA Journal of Emerging Technologies (SJET)**
**Office of Research, Innovation & Commercialization – ORIC**
**Sukkur IBA University** - Airport Road Sukkur-65200, Sindh Pakistan
Tel: (09271) 5644233 -37 Fax: (092 71) 5804425 Email: sjet@iba-suk.edu.pk URL: sjet.iba-suk.edu.pk
==================================================================

Vol. 5, No. 2 | July - December 2022

# Modeling of IEEE 9-Bus System with Load Flow and Short Circuit Analyses

Rehan Ali[1], Abdul Saboor Gul[1], Ramez Akbar Talani[2]

**Abstract:**
*Load Flow and Short Circuit Analyses have been performed on standard IEEE 9-Bus System. This paper is going to describe in detail how these analyses were performed and what results were achieved. In addition to the performance of these two analyses on the standard test system, they were also performed by removing a voltage- controlled generator from the test system. It is revealed that by removal of a generator, the magnitude of short- circuit fault current in a system would reduce as compared to the case where all generators were present in the system.*

**Keywords:** *IEEE 9- Bus system, Load flow analysis, Short circuit (SC) analysis*

## 1. Introduction

Before installing a power system in the real world, many tests and analyses are digitally performed on computer simulations in order to determine the exact requirements and necessary precautions for the installation of the actual system. As we know that within power systems, electric power flows from the point of generation to the required destination via different paths of network. This flow comprises of reactive and active power known as load flow. It is very important to investigate this load flow so as to plan and determine the steady operation of system. Such investigations employs systematic analytical approaches so as to examine the different voltages on bus, phase angle, as well as reactive and active power that is flowing within various branches of load such as generators and transformers when steady state is attained.

The important knowledge gained from such analysis gives the absolute values as well as phase angles of load bus voltage reactive powers and voltage phase angles at generator buses, real and reactive power flows on transmission lines together with power at the reference bus. In addition to analytical techniques, numerical methods are also employed to solve load flow equations, as they become nonlinear and requires solution by method of iteration. However, such numerical solutions mostly provide only approximate solution. For more than thirty years different numerical investigations have been carried out to analyze power flow problems. Out of them widely used methods are Fast decoupled methods the Gauss-Seidel and Newton-Raphson and [1-4].

Moreover, due to sudden and mammoth developments in industry the society requires electric systems which keep on enhancing in size and complexity. In such systems the power flow equations range to several thousand with such magnitude of equations it is not feasible for any numerical technique to provide solution which converges. Such

[1] Department of Electrical Engineering, NED University of Engineering and Technology, Karachi, Pakistan
[2] Faculty of Electrical Engineering, QUEST, Pakistan
Corresponding Author: rahimoon2223@gmail.com

difficulty compels electrical engineers to resort to more suitable and reliable techniques. Hence the issue is that industry is looking for technique that is more feasible for the analysis of power systems

Analytical calculations are appropriate when estimation of characteristics of smaller magnitude of circuits is being carried out, however for the more precise and accurate analysis it is imperative to employ specialized programs.

Digital computers are being used in calculation s of power flow equations since 1950s. With the rapid advancement in computing power all sort of load flow studies can now be conveniently carried out.

Load Flow Analysis is one of such analyses. It is important for planning, economic scheduling, operation, and the distribution of power between different sources. It is also required for many other analyses such as transient stability and contingency studies. This analysis is also required to determine the effects of new loads and new generating stations on the power system, so that the system can be extended. This analysis involves finding voltages, currents, real and reactive power flows at different points in a power system under normal/steady-state conditions. [2, 3]

Short-Circuit analysis is mainly required for the protection of the power system and its equipment Sizing. It is performed on different buses under faulty conditions which can cause a short-circuit in the system. It involves finding the huge short- circuit current that flows in the system in the event of a fault. This analysis is also a major part of the system's protection coordination study, which involves the sizing and placement of protective equipment (fuses, circuit breakers, etc.) in the system, in order to protect the system. [4-6].

### 1.1. Bus Classification

Any node where more than one component like electric generators and transmission lines are connected is known as bus. In electrical engineering the bus is concerned with four quantities which include potential difference and its phase angle, active and reactive power. [11,12].

Out of them 02 quantities are already known whereas remaining two has to be determined via solution of equation [13].

### 1.2. Slack Bus

The slack bus is a sort of reference bus used to satisfy the condition of power balance. It is mostly concerned with generation unit which is adjusted to employ the power balance condition [12].

### 1.3. Generator (PV) Bus

Generator bus is essentially concerned with the voltage control. It is connected with generating unit where power generated from the bus is controlled with the help of prime mover. Whereas the control over voltage so produced is made by varying the generator excitation. Mostly the limits are set for reactive power and it depends on the specifications of machine used.

### 1.4. Load (PQ) Bus

The load bus is not due to generator but we can obtain this from old data, measurements or from forecast. In this case real power provided to the system is taken as positive, whereas the electric power utilized within system is taken as negative. In this bus P and Q are known variable while |V| and δ are unknown variables. [8, 12].

### 1.5. Power Flow Analysis Methods

From this analysis we mean investigation of various variables of electrical system such as voltage, current, active power etc at various points in a system, once the system has attained steady state.This is carried out by solving simultaneous equations. These solutions form the platform for solving equations for performance of power system [4]

### 1.6. Gauss-Seidel Method

This method is developed based on the Gauss method. It is an iterative method used for solving set of nonlinear algebraic equations [14]. The method makes use of an initial guess for value of voltage, to obtain a calculated value of a particular variable. The initial guess value is replaced by a calculated value. The process is then repeated until the iteration solution converges. The convergence is quite sensitive to the starting values assumed. But this method suffers from poor convergence characteristics [15].

### 1.7. Fast Decoupled Method

This method is considered as speedy and efficient for obtaining the solutions of problems pertaining to power flow. This method is basically an extension of Newton-Raphson method. This method was first presented in 1974 by Stott and Alsac in 1974 [16-18]. Since this method is based on Newton-Raphson method so it provides much simplification in carrying out calculations, also reliable results can be obtained, rapid convergence and this method soon became mostly employed method in power flow problems.

In this method polar coordinates and some approximations are used so fast algorithms for power solutions are obtained. However, the use of certain approximations also results in non-convergence in some cases. For instance, in case of large resistance-to-reactance ratios as well as for the case where low voltages appears at some buses. In such cases the solutions do not converge nicely due to approximations, so one has to resort some assumptions so as to simplify Jacobian Matrix.

To overcome these difficulties arising due to non convergence of solutions various investigations have been carried out. For instance one of them is co n- convergence of systems with high R/X ratios, and others with low voltage buses.

### 1.8. Newton-Raphson Method

Sir Isaac Newton and Raphson formulated this method so it is named after them. This method employs an iterative method. By using this method we can approximate non linear simultaneous equation into linear ones with the use Taylor series. However, Taylor series is expanded to the first order approximations only.

This numerical technique is widely used in the calculation of power flow due to its powerful convergence characteristics as compared with similar techniques.

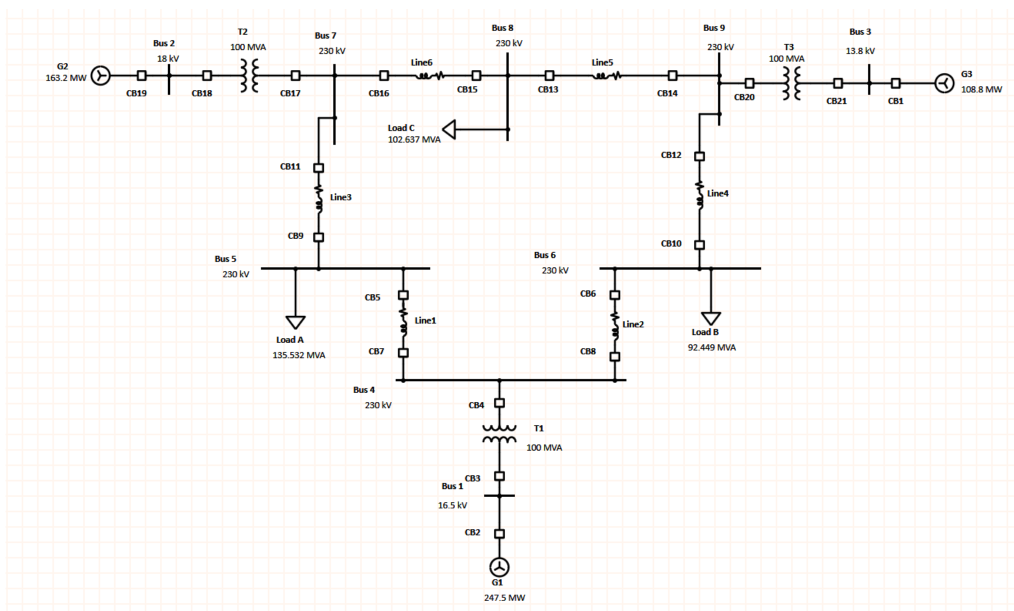



**Fig. 1.** IEEE 9-Bus System

## 2. Methodology

For this study, ETAP software has been used. The Load Flow and Short Circuit analyses are first performed on the standard IEEE 9-Bus system, and their results are recorded. The Load Flow has been evaluated by the software using Newton- Raphson iterative method. After this, one generator has been removed from this test system and the above mentioned two analyses are performed again. The results of these analyses on this new system are recorded and then compared with the results obtained from the standard system.

## 3. Test System Details

The test system for this study was the IEEE 9-Bus system. The SLD of this system is shown in figure-1. The system comprises of 3 generators, connected to 3 generator buses. These are buses 1, 2 and 3 respectively. After this, we have 3 step-up transformers, one after each generator bus, which step up the voltage to the transmission level. These transformers are connected to buses 4, 7 and 9 respectively. These buses are further connected to 3 load buses, namely buses 5, 6 and 8. The system has 1 swing generator operating at a voltage of 1.04p.u. and 2 voltage-controlled generators operating at a voltage of 1.025p.u. The system consists of 3 load buses interconnected through 6 transmission lines. The transmission lines are modeled with vertical configuration of conductors and a conductor spacing of 10ft, with a single ground wire. Gymnastic conductor is used. The loads are modeled with a requirement of 230kv and varying real and reactive powers. Also, circuit breakers are connected on each end of the transmission lines and generator buses for protection from overloading or faults.

The Load Flow and Short Circuit analyses are performed on this standard system first. After this, generator 3 is removed from the system, and then the above 2 analyses are repeated for this modified system and the results are compared.

## 4. Simulation and Discussion

In this section, we will describe the simulations performed in this study and discuss their results.

### 4.1. Load Flow Analysis on standard IEEE 9-Bus System

| Bus | Voltage | | | Generation | | Load | | Load Flow | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | kV | % Mag | Ang. | MW | Mvar | MW | Mvar | ID | MW | Mvar | Amp | %PF |
| *Bus 1 | 16.500 | 104.000 | 0.0 | 65.725 | 63.710 | 0.000 | 0.000 | Bus 4 | 65.725 | 63.710 | 3079.7 | 71.8 |
| *Bus 2 | 18.000 | 102.500 | 3.6 | 163.000 | 40.218 | 0.000 | 0.000 | Bus 7 | 163.000 | 40.218 | 5253.7 | 97.1 |
| *Bus 3 | 13.800 | 102.500 | 0.7 | 85.000 | 36.403 | 0.000 | 0.000 | Bus 9 | 85.000 | 36.403 | 3774.2 | 91.9 |
| Bus 4 | 230.000 | 100.534 | -2.1 | 0.000 | 0.000 | 0.000 | 0.000 | Bus 5 | 32.707 | 38.097 | 125.4 | 65.1 |
| | | | | | | | | Bus 6 | 33.014 | 21.151 | 97.9 | 84.2 |
| | | | | | | | | Bus 1 | -65.721 | -59.248 | 220.9 | 74.3 |
| Bus 5 | 230.000 | 100.523 | -2.1 | 0.000 | 0.000 | 127.160 | 50.855 | Bus 4 | -32.705 | -38.153 | 125.5 | 65.1 |
| | | | | | | | | Bus 7 | -94.455 | -12.702 | 238.0 | 99.1 |
| Bus 6 | 230.000 | 100.527 | -2.1 | 0.000 | 0.000 | 88.631 | 29.544 | Bus 4 | -33.013 | -21.209 | 98.0 | 84.1 |
| | | | | | | | | Bus 9 | -55.618 | -8.334 | 140.4 | 98.9 |
| Bus 7 | 230.000 | 100.530 | -2.1 | 0.000 | 0.000 | 0.000 | 0.000 | Bus 5 | 94.459 | 12.662 | 238.0 | 99.1 |
| | | | | | | | | Bus 8 | 68.524 | 10.788 | 173.2 | 98.8 |
| | | | | | | | | Bus 2 | -162.983 | -23.450 | 411.2 | 99.0 |
| Bus 8 | 230.000 | 100.524 | -2.1 | 0.000 | 0.000 | 97.897 | 34.250 | Bus 9 | -29.375 | -23.412 | 93.8 | 78.2 |
| | | | | | | | | Bus 7 | -68.522 | -10.838 | 173.2 | 98.8 |
| Bus 9 | 230.000 | 100.531 | -2.1 | 0.000 | 0.000 | 0.000 | 0.000 | Bus 6 | 55.619 | 8.280 | 140.4 | 98.9 |
| | | | | | | | | Bus 8 | 29.376 | 23.354 | 93.7 | 78.3 |
| | | | | | | | | Bus 3 | -84.995 | -31.634 | 226.5 | 93.7 |

**Fig. 2.** Load Flow Report of IEEE 9-Bus System

As we have studied theoretically, the real power flows when there is a difference between the magnitudes of voltages between 2 buses, and reactive power flows when there is a difference between the voltage phase angles of 2 buses. This is proven here in the simulation as well, as we can see figure-2.

### 4.2. Short Circuit Analysis on standard IEEE 9-Bus System

Whenever a short-circuit fault occurs, it has 2 components: AC component and DC component. The AC component has further 3 stages; sub transient, transient and steady state. At the instance of fault occurrence, the sub transient current exists. It has the highest magnitude and the shortest time duration (0.5 cycles).

#### 4.2.1. Sub-Transient

Sub transient current rapidly decreases to the transient stage. It is relatively longer (from 1.5 cycles to 4 cycles approx.). This transient current further reduces to the steady-state value, which occurs after about 30 cycles, and its magnitude is sustained throughout the rest of the fault. As shown in figure 3.

NOTE: The steady-state current mentioned in figure is still the fault current, as its magnitude is still many times greater than the nominal current.

1/2 Cycle - 3-Phase, LG, LL, & LLG Fault Currents

Prefault Voltage = 100 % of the Bus Nominal Voltage

| Bus | | 3-Phase Fault | | | Line-to-Ground Fault | | | Line-to-Line Fault | | | 'Line-to-Line-to-Ground | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | kV | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. |
| Bus 1 | 16.500 | 1.903 | -58.126 | 58.158 | 5.739 | -76.221 | 76.436 | 55.589 | 3.208 | 55.682 | -56.165 | 39.901 | 68.896 |
| Bus 2 | 18.000 | 2.101 | -53.339 | 53.381 | 4.504 | -68.620 | 68.768 | 47.512 | 2.858 | 47.598 | -49.166 | 42.253 | 64.828 |
| Bus 3 | 13.800 | 2.009 | -58.226 | 58.261 | 4.246 | -74.556 | 74.677 | 51.843 | 2.730 | 51.915 | -53.250 | 45.873 | 70.284 |

**Fig. 3.** Short Circuit Fault Sub-Transient Component

In the above-mentioned results, we can see the sub transient magnitudes of symmetrical and unsymmetrical fault currents (in kA) which flow through the 3 generator buses in the event of a fault. The current is highest during this stage as the generator's impedance is lowest here.

#### 4.2.2. Transient

30 Cycle - 3-Phase, LG, LL, & LLG Fault Currents

Prefault Voltage = 100 % of the Bus Nominal Voltage

| Bus | | 3-Phase Fault | | | Line-to-Ground Fault | | | Line-to-Line Fault | | | *Line-to-Line-to-Ground | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | kV | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. |
| Bus 1 | 16.500 | 1.072 | -44.505 | 44.518 | 4.431 | -67.263 | 67.409 | 47.569 | 2.318 | 47.625 | -45.094 | 32.785 | 55.752 |
| Bus 2 | 18.000 | 1.073 | -39.918 | 39.933 | 3.362 | -59.997 | 60.091 | 40.522 | 2.015 | 40.572 | -38.465 | 33.910 | 51.278 |
| Bus 3 | 13.800 | 1.077 | -45.140 | 45.153 | 3.268 | -66.365 | 66.446 | 45.131 | 1.992 | 45.175 | -43.012 | 37.841 | 57.289 |

**Fig. 4.** Short Circuit Fault Transient Component

Comparing the results of the sub transient and transient fault currents, we see that both of these are identical. This is due to the fact that ETAP calculates both these currents using the sub transient impedance values, but in reality, the impedances of both these currents are different, and therefore, there magnitudes are different as well. As shown in Figure 4

This space is intentionally left blank to adjust the figers

#### 4.2.3. Steady-State

30 Cycle - 3-Phase, LG, LL, & LLG Fault Currents

Prefault Voltage = 100 % of the Bus Nominal Voltage

| Bus | | 3-Phase Fault | | | Line-to-Ground Fault | | | Line-to-Line Fault | | | *Line-to-Line-to-Ground | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | kV | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. |
| Bus 1 | 16.500 | 1.072 | -44.505 | 44.518 | 4.431 | -67.263 | 67.409 | 47.569 | 2.318 | 47.625 | -45.094 | 32.785 | 55.752 |
| Bus 2 | 18.000 | 1.073 | -39.918 | 39.933 | 3.362 | -59.997 | 60.091 | 40.522 | 2.015 | 40.572 | -38.465 | 33.910 | 51.278 |
| Bus 3 | 13.800 | 1.077 | -45.140 | 45.153 | 3.268 | -66.365 | 66.446 | 45.131 | 1.992 | 45.175 | -43.012 | 37.841 | 57.289 |

**Fig. 5.** Short Circuit Fault Steady State Component

Comparing the steady-state current values with the sub transient and transient current values, we can see that its magnitude is lower than the other two fault currents. This is because the impedance of the generator during the steady-state stage becomes equal to the synchronous reactance. The fault current will not decrease any further than this value. Load Flow Analysis on standard IEEE 9-Bus System after removing one PV Generator. As shown in figure 5.

The voltage-controlled generator maintains the voltage throughout the system. Removing it causes a voltage drop at all the buses (except the generator buses). The load supplied by this generator is now being fed by the remaining 2 generators. Therefore, the power flow at the remaining 2 generator buses increases.

### 4.3. Short Circuit Analysis on standard IEEE 9-Bus System after removing one PV Generator

The short-circuit analysis is repeated after removing one of the PV generators. We see that the magnitude of the overall fault current reduces, when compared to the original test system's fault current. This is due to the obvious fact that now there is one less generator to feed the fault. Shown in figure 6.

| Bus | | Voltage | | Generation | | Load | | Load Flow | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | kV | % Mag | Ang. | MW | Mvar | MW | Mvar | ID | MW | Mvar | Amp | %PF |
| *Bus 1 | 16.500 | 104.000 | 0.0 | 144.168 | 87.251 | 0.000 | 0.000 | Bus 4 | 144.168 | 87.251 | 5669.7 | 85.6 |
| *Bus 2 | 18.000 | 102.500 | 1.1 | 163.000 | 57.607 | 0.000 | 0.000 | Bus 7 | 163.000 | 57.607 | 5409.9 | 94.3 |
| Bus 4 | 230.000 | 99.480 | -4.6 | 0.000 | 0.000 | 0.000 | 0.000 | Bus 5 | 57.490 | 36.383 | 171.7 | 84.5 |
| | | | | | | | | Bus 6 | 86.663 | 35.745 | 236.6 | 92.4 |
| | | | | | | | | Bus 1 | -144.153 | -72.128 | 406.7 | 89.4 |
| Bus 5 | 230.000 | 99.468 | -4.6 | 0.000 | 0.000 | 124.507 | 49.794 | Bus 4 | -57.488 | -36.432 | 171.8 | 84.5 |
| | | | | | | | | Bus 7 | -67.019 | -13.361 | 172.5 | 98.1 |
| Bus 6 | 230.000 | 99.467 | -4.6 | 0.000 | 0.000 | 86.773 | 28.924 | Bus 4 | -86.659 | -35.783 | 236.6 | 92.4 |
| | | | | | | | | Bus 9 | -0.114 | 6.859 | 17.3 | -1.7 |
| Bus 7 | 230.000 | 99.475 | -4.6 | 0.000 | 0.000 | 0.000 | 0.000 | Bus 5 | 67.021 | 13.312 | 172.4 | 98.1 |
| | | | | | | | | Bus 8 | 95.961 | 26.515 | 251.2 | 96.4 |
| | | | | | | | | Bus 2 | -162.982 | -39.828 | 423.4 | 97.1 |
| Bus 8 | 230.000 | 99.464 | -4.6 | 0.000 | 0.000 | 95.843 | 33.531 | Bus 9 | 0.114 | -6.980 | 17.6 | -1.6 |
| | | | | | | | | Bus 7 | -95.957 | -26.551 | 251.3 | 96.4 |
| Bus 9 | 230.000 | 99.465 | -4.6 | 0.000 | 0.000 | 0.000 | 0.000 | Bus 6 | 0.114 | -6.920 | 17.5 | -1.6 |
| | | | | | | | | Bus 8 | -0.114 | 6.920 | 17.5 | -1.6 |

**Fig. 6.** Short Circuit Analysis after Removing One PV Generator

#### 4.3.1. Sub-Transient

when compared to the original test system's fault current. This is due to the obvious fact that now there is one less generator to feed the fault. As shown in figure 7.

1/2 Cycle - 3-Phase, LG, LL, & LLG Fault Currents

Prefault Voltage = 100 % of the Bus Nominal Voltage

| Bus | | 3-Phase Fault | | | Line-to-Ground Fault | | | Line-to-Line Fault | | | Line-to-Line-to-Ground | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | kV | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. |
| Bus 1 | 16.500 | 1.775 | -50.525 | 50.556 | 5.278 | -68.754 | 68.957 | 48.888 | 3.013 | 48.981 | -49.026 | 36.946 | 61.388 |
| Bus 2 | 18.000 | 1.977 | -46.802 | 46.844 | 4.260 | -61.903 | 62.049 | 42.043 | 2.712 | 42.130 | -43.360 | 39.040 | 58.345 |

**Fig. 7.** Short Circuit Fault Sub-Transient Component after Removing One PV Generator

### 4.3.2. Transient

Short circuit analysis of transient component on a 9-Bus system after removing one generator is shown in figure 8.

1.5-4 Cycle - 3-Phase, LG, LL, & LLG Fault Currents

Prefault Voltage = 100 % of the Bus Nominal Voltage

| Bus | | 3-Phase Fault | | | Line-to-Ground Fault | | | Line-to-Line Fault | | | Line-to-Line-to-Ground | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | kV | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. |
| Bus 1 | 16.500 | 1.775 | -50.525 | 50.556 | 5.278 | -68.754 | 68.957 | 48.888 | 3.013 | 48.981 | -49.026 | 36.946 | 61.388 |
| Bus 2 | 18.000 | 1.977 | -46.802 | 46.844 | 4.260 | -61.903 | 62.049 | 42.043 | 2.712 | 42.130 | -43.360 | 39.040 | 58.345 |

**Fig. 8.** Short Circuit Fault Transient Component after Removing One PV Generator

### 4.3.3. Steady-State

Short circuit analysis of the steady state component on a 9-Bus system after removal of a single generator is shown in figure 9.

30 Cycle - 3-Phase, LG, LL, & LLG Fault Currents

Prefault Voltage = 100 % of the Bus Nominal Voltage

| Bus | | 3-Phase Fault | | | Line-to-Ground Fault | | | Line-to-Line Fault | | | Line-to-Line-to-Ground | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | kV | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. | Real | Imag. | Mag. |
| Bus 1 | 16.500 | 0.972 | -37.859 | 37.872 | 3.958 | -59.730 | 59.861 | 41.205 | 2.122 | 41.260 | -38.452 | 29.700 | 48.586 |
| Bus 2 | 18.000 | 0.982 | -34.213 | 34.227 | 3.092 | -53.286 | 53.376 | 35.319 | 1.866 | 35.369 | -33.060 | 30.597 | 45.046 |

**Fig. 9.** Short Circuit Fault Steady-State Component after Removing One PV Generator

## 4.4. Comparison of the system before and after removing the PV Bus

### 4.4.1. Load Flow Comparison

TABLE I. LOAD FLOW COMPARISON BEFORE AND AFTER REMOVING PV BUS

| | Before removing PV Bus | After removing PV Bus |
|---|---|---|
| *Voltage magnitude at Bus 1 | 104% | 104% |
| *Voltage magnitude at Bus 2 | 102.5% | 102.5% |
| *Voltage magnitude at Bus 3 | 102.5% | - |
| Voltage magnitude at Bus 4 | 100.534% | 99.480% |
| Voltage magnitude at Bus 5 | 100.523% | 99.468% |
| Voltage magnitude at Bus 6 | 100.527% | 99.467% |
| Voltage magnitude at Bus 7 | 100.530% | 99.475% |
| Voltage magnitude at Bus 8 | 100.524% | 99.464% |

| | | |
|---|---|---|
| **Voltage magnitude at Bus 9** | 100.531% | 99.465% |

Here we can see that after removing the PV generator from bus 3, the voltages at all the buses (except the generator buses) drop, as has been discussed above. Shown in table-I

### 4.4.2. Short Circuit Comparison:

NOTE: The steady-state fault currents in the table II below are 3-phase fault currents. The results will remain same for other types of faults. Shown in table-II

TABLE II. SHORT CIRCUIT COMPARISON BEFORE AND AFTER REMOVING PV BUS

| | **Before removing PV Bus** | **After removing PV Bus** |
|---|---|---|
| ***Steady-State Current at Bus 1** | 44.518kA | 37.872kA |
| ***Steady-State Current at Bus 2** | 39.933kA | 34.227kA |
| ***Steady-State Current at Bus 3** | 45.153kA | - |

Here we can see that the fault current reduces after removing a PV generator from the system, as there is one less generator to feed the fault here, as has been discussed above as well.

## *5.* Conclusion

In this paper a numerical study has been carried out for load flow and short circuit analysis on standard IEEE 9-bus system. From the above study we conclude that, by removal of a voltage-controlled generator from our system, the magnitude and phase angles of voltages will reduce for all the buses other than the remaining generator buses. Since this was a particularly small test system, the voltage drop was minimum and did not affect the system greatly. But, had it been a large practical system, then such a removal of a voltage-controlled generator would have damaging effects on the transmission system and can also damage the connected loads due to under-voltage. Moreover, it is also found that by removal of a generator, the magnitude of short- circuit fault current in a system would reduce as compared to the case where all generators were present in the system.

## REFERENCES


[1] Mageshvaran, R., Raglend, I.J., Yuvaraj, V., Rizwankhan, P.G., Vijayakumar, T. and Sudheera (2008) Implementation of Non-Traditional Optimization Techniques (PSO, CPSO, HDE) for the Optimal Load Flow Solution. TENCON2008- 2008 IEEE Region 10 Conference, 19-21 November 2008.

[2] Elgerd, O.L. (2012) Electric Energy Systems Theory: An Introduction. 2nd Edition, Mc-Graw-Hill.

[3] Kothari, I.J. and Nagrath, D.P. (2007) Modern Power System Analysis. 3rd Edition, New York.

[4] Keyhani, A., Abur, A. and Hao, S. (1989) Evaluation of Power Flow Techniques for Personal Computers. IEEE Transactions on Power Systems, 4, 817-826.

[5] Hale, H.W. and Goodrich, R.W. (1959) Digital Computation or Power Flow—Some New Aspects. Power Apparatus and Systems, Part III. Transactions of the American Institute of Electrical Engineers, 78, 919-923.

[6] Sato, N. and Tinney, W.F. (1963) Techniques for Exploiting the Sparsity or the Network Admittance Matrix. IEEE Transactions on Power Apparatus and Systems, 82, 944-950.

[7] Aroop, B., Satyajit, B. and Sanjib, H. (2014) Power Flow Analysis on IEEE 57 bus System Using Mathlab. International Journal of Engineering Research & Technology (IJERT), 3.

[8] Milano, F. (2009) Continuous Newton's Method for Power Flow Analysis. IEEE Transactions on Power Systems, 24,50-57.

[9] Grainger, J.J. and Stevenson, W.D. (1994) Power System Analysis. McGraw-Hill, New York.

[10] Tinney, W.F. and Hart, C.E. (1967) Power Flow Solution by Newton's Method. IEEE Transactions on Power Apparatus and Systems, PAS-86, 1449-1460.

[11] Bhakti, N. and Rajani, N. (2014) Steady State Analysis of IEEE-6 Bus System Using PSAT Power Tool Box. International Journal of Engineering Science and Innovation Technology (IJESIT), 3.

[12] Hadi, S. (2010) Power System Analysis. 3rd Edition, PSA Publishing, North York.

[13] Kabisama, H.W. Electrical Power Engineering. McGraw-Hill, New York.

[14] Gilbert, G.M., Bouchard, D.E. and Chikhani, A.Y. (1998) A Comparison of Load Flow Analysis Using Dist Flow, Gauss-Seidel, and Optimal Load Flow Algorithms. Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering, Waterloo, Ontario, 24-28 May 1998, 850-853.

[15] Glover, J.D. and Sarma, M.S. (2002) Power System Analysis and Design. 3rd Edition, Brooks/Cole, Pacific Grove.

[16] Stott, B. and Alsac, O. (1974) Fast Decoupled Load Flow. IEEE Transactions on Power Apparatus and Systems, PAS-93, 859-869.

[17] Stott, B. (1974) Review of Load-Flow Calculation Methods. Proceedings of the IEEE, 62, 916-929.

Adejumobi, I.A., et al. (2014) Numerical Methods in Load Flow Analysis: An Application to Nigeria Grid System. International Journal of Electrical and Electronics Engineering.

Vol. 5, No. 2 | July – December 2022

# Machine Learning-Based Fake News Detection with Amalgamated Feature Extraction Method

Muhammad Bux Alvi[1]*, Majdah Alvi[1], Rehan Ali Shah[1], Adnan Akhter[1], Mubashira Munir[1], Rakesh Kumar[2], Kavita Tabbassum[3]

**Abstract:**

*Product fake reviews are increasing as the trend is changing toward online sales and purchases. Fake review detection is critical and challenging for both researchers and online retailers. As new techniques are introduced to catch the non-organic reviewer, so are their intruding approaches. In this paper, different features are amalgamated along with sentiment scores to design a model that checks the model performance under different classifiers. For this purpose, six supervised learning algorithms are utilized to build the fake review detection models, using LIWC, unigrams, and sentiment score features. Results show that the amalgamation of selected features is a better approach to counterfeit review detection, achieving an accuracy score of 88.76%, which is promising when compared to similar other work.*

**Keywords:** *fake reviews, machine learning, amalgamated features, LIWC, sentiment score*

## 1 Introduction

A fake review is a false judgment or an opinionated text on a product or a service. Reviews can significantly affect the decision of buyers while shopping online. According to “Statista” statistics, e-commerce sales increase 6% in America from 2013 to 2020 [1]. As online purchase increases, so is the competition of online retailer giants. Therefore, the retailers and manufacturers take these reviews on a serious note. Fake reviewers capitalize on this opportunity to artificially devalue or promote products and services [2][3]. Hence, fake review prediction becomes a critical research area as online purchases increase. With the explosive growth of online businesses, the quantity and importance of reviews continue to increase. Fake reviews severely threaten researchers [4] and online retailers [5]. Reviews can be positive to increase purchases on an online platform by manipulating users with fake customer reviews. Conversely, it can be a negative review to distract purchasers. It is estimated that 80% of users believe in posted product reviews before purchasing any product [6]. Negative fake reviews are used to defame competitor’s reputations. People who post such fake reviews are usually freelancers, and companies hire their services for writing fake reviews. Giant retailers like Amazon find these fake reviews of severe threat to their reputation and filed a complaint against review spamming [7].

Fake review prediction can be performed manually or automatically. Research has been carried out on manual opinion spam prediction for several years [8]. Early methods of fake review prediction were rudimentary. Many texts

[1]Department of Computer Systems Engineering, Faculty of Engineering, The Islamia University of Bahawalpur, Pakistan
[2]Freelancer and Researcher
[3]Department Information Technology Center, Sindh Agriculture University Tandojam, Sindh, Pakistan
Corresponding Author: mbalvi@iub.edu.pk

analysis-based approaches are found in the literature [9]. Based on the research, commercial platforms developed opinion spam filtering systems to detect deceptive reviews.

Nevertheless, these systems make the fake reviewers enhance their review quality and deceive the detecting systems [10]. As time elapsed, those traditional approaches would not work efficiently because the fake reviewers started behaving like regular users.

Therefore, the trend of manual fake review prediction changed from text-based analysis to pattern and feature analysis like time [11], topics [12], ranking pattern [13], activity volume [14], and geolocation [15]. However, manual methods are slow, expensive, and of low accuracy. Automated methods based on machine learning could also identify the opinion spams and spammers by analyzing the review features. Text mining and Natural Language Processing (NLP) work together to generate the concept of content mining, and review spam detection comes under this concept. Additional review characteristics like review timings, reviewer id, and deviation trend of the review from other reviews of the same category are also considered in spam review detection. Jindal et al.[16] used the machine learning technique and showed that the amalgam of features is more robust than a single feature for fake review prediction. Li et al. in [17] showed that combining a bag of words (BOW) with more general features performs better than BOW alone. Mukherjee et al. [18] used machine learning with abnormal behavioral features of the reviewers and depicted that this technique was better than the linguistic features-based technique.

The significant contribution of this research is to develop a fake reviews detection model that uses machine learning techniques that will employ a heuristic optimization algorithm for affecting features and test its reliability and robustness against existing techniques. Such a model, when employed, can benefit retailers and giant business companies to shield their businesses against fake reviews and reviewers.

## 2 Literature Review

Advancements have been made in fake review detection by introducing new techniques and methods by researchers. These techniques play their role in improving accuracy and performance. So far, reviews are marked as spam based on either review spam detection or reviewer spam detection. Both techniques are helpful in fake review detection. Prior deals with content mining and natural language processing (NLP), whereas later technique applied on reviewer id and his behavior. Jindal et al. [16] is the first researcher who studied opinion spamming using supervised learning. The author divided the reviews into three categories (fake opinions, the brand only reviews, and non-review) and detected opinion spamming by finding duplicate reviews using the "w-shingling" method. The author used a dataset from Amazon with more than 5 million product reviews, applied his devised technique with a logistic regression algorithm, and achieved an AUC of 78%. Lim, Nguyen, Jindal, Liu, and Lauw [19] proposed a behavioral methodology for revealing spammers for review. They tried to figure out some spammer habits like targeting goods and tried to optimize their effect.

Moreover, they suggested a model focused on specific patterns to identify rating spammers. Ott et al. [20][21] created a data set for analysis in review spam detection. The data set comprised positive opinion spam with truthful reviews and negative opinion spam with real reviews. The author applied the n-gram and linguistic features to find fake reviews under a supervised learning mechanism, and the results were verified with human performance. In their research, Feng et al. [4] framed a model based on the normal distribution of opinion to detect fake reviews. In their view, a product or a service review involved this concept of normal distribution of opinion. Shojaee et al. [9] suggested a novel technique for fake review detection by combining Lexical and Synthetic features. Elmurngi and Gherbi [22] proposed a text classification and sentimental analysis approach for different machine learning algorithms with stop words and without stop words. They also applied a decision tree algorithm to improve their results. Shah, Ahsan, Kafi, Nahian, and Hossain [23] combined Supervised & Active learning and created a model to detect spamming. Both fictitious and real-life data were used for spam analysis.

## 3 Proposed Approach

This section describes the proposed method to accomplish the task of fake review detection. This research uses two features as classification criteria with a sentiment score feature (an additional feature). These individual features and combinations are used to train various classifiers and tested against evaluation metrics. Reviews are classified as fake or not fake. This study uses six classification algorithms: Naive Bayes, decision tree, instance-based KNN, support vector machine (SVM), logistic Regression, and Random Forest. Training data is 80%, and 20% of data is set aside for testing purposes with a 5-fold cross-validation technique. Figure 1 presents the adopted research method for this work.

### 3.1 Data Acquisition and Pre-processing

The data set selected for this research contains 1600 reviews combined from two data sets (hotel review data sets). The data sets were created by Myle et al. and are available from [20][21]. The data set contains eight hundred truthful reviews, of which four hundred are positive, and four hundred are negative. Similarly, 800 spam reviews are also included in this data set, of which half are positive, and half are negative. The preprocessing of the data set significantly affects the accuracy of results [24][25][26]. Furthermore, preprocessing curbs feature vector space. Therefore, preprocessing techniques like missing values management, tokenization, stop words removal, and generating n-gram are implemented on the data set to obtain cleaner data set.

### 3.2 Features

Features are pieces (s) of text that have semantic significance. In the text data systems, features highly influence the effectiveness of the developed model.

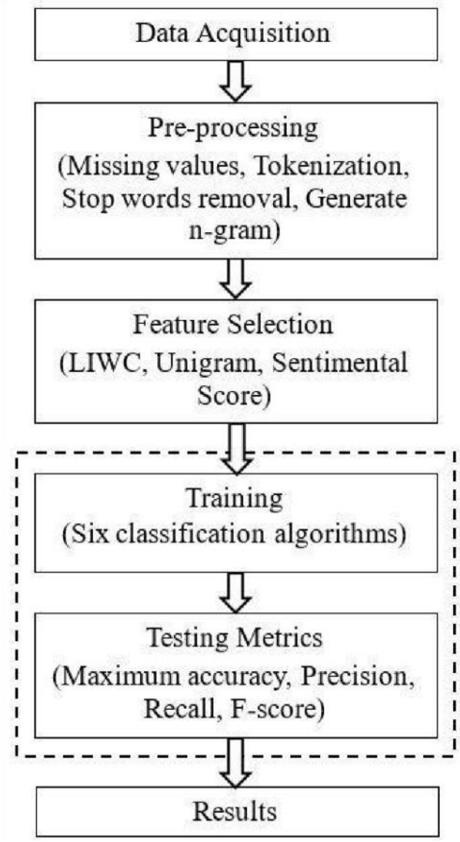



**Fig. 1.** Proposed Machine Learning Approach with Amalgamated Features for Fake Reviews Detection

#### 3.2.1 N-Grams

In this feature extraction method, n-adjacent tokens are picked as a feature from review contents. It is denoted as unigram if one adjacent word is selected, bigram if two adjacent words are selected, and trigram with three adjacent words at a time. These features can effectively help model all the content within the text. In this research work, unigram is used as a feature.

TABLE I. STATE OF ART SPAM REVIEW DETECTION TECHNIQUES

| Reference | Year | Data set | Learning type | Techniques/ Algorithm | Results | Limitations |
|---|---|---|---|---|---|---|
| [4] Feng | 2012 | Ott et al. data set with modification | Supervised Learning | LIBSVM classifier/ Term frequency | Accuracy 72.5% | Specific kind of dataset |
| [9] Shojaee | 2013 | Ott et al. data set | Supervised Learning | SVM/ Naïve Bayes/ Stylometric Feature | F-measure 84% | Limited to a specific domain |
| [13] Jindal N, Liu B | 2007 | Data set of the manufactured product only | Supervised Learning | Logistic Regression | average AUC 78% | Lack of accuracy of a real-world data set |
| [18] Lim, Ee-Peng | 2010 | Amazon Data set | Supervised Learning | Behavioral features of Spammer | Accuracy 78% | Limit set of data for supervised learning |
| [20] Jeffrey T. at el | 2013 | Ott et al. data set | Supervised Learning | Support Vector Machine (SVM) | Accuracy 86 % | Human judgments can be imperfect and biased. |
| [21] Elmurngi E. | 2017 | Movie review data set | Supervised Learning | DT(DT-J48)/ SVM/KNN | Accuracy 81.75% | Feature selection methods are not used |
| [22] Ahsan, Nahian, Kafi, Hossain and Shah | 2017 | Ott dataset | Active/ Supervised Learning | Hybrid classifier using NB/ SVC /DT /Maximum Entropy | Accuracy 95% | Small scale dataset is used for a specific domain |

### 3.2.2 LIWC

The Linguistic Inquiry and Word Count (LIWC) is a text analysis method. This method can analyze eighty different features, for example, psychological concerns like emotion, text functional aspects, and personal and perception concerns like religion [27].

### 3.2.3 Sentiment Score

It has been observed that spammers with negative reviews generally use more negative words like “bad” and “dissatisfied”. This way, the degree of negative sentiment is increased compared to a non-spam negative review. Likewise, spammers with positive reviews generally use positive terms such as “good”, “great”, “nice”, and “gorgeous”. Therefore, reviewers show more positive sentiment than a non-spam positive review. The sentiment score of a review can be calculated by the following formula [28].

$$SC(rt) = \sum(-1)^n \frac{S(W_i)}{Distance\ (fet_i\ W_i)} \quad (1)$$

where "rt" is review text, "S(Wi)" is the sentiment polarity of word Wi (+1 or -1), "n" denotes the total number of negation-words in a feature with default = 0, "fet" refers to a feature in a review sentence and "distance (fet, Wi)" is the distance between feature and word.

## 3.3 Classification Algorithms

Six various classification algorithms are used in this paper in order to determine the effect of different features and their combinations on classification accuracy and performance.

### 3.3.1 Naïve Bayes (NB)

NB is based on the Bayes theorem [29]. It is a probabilistic multiclass classification algorithm assuming features independency to foresee the output class. Equation 2 checks the probability of the feature-set being categorized into a particular class:

$$P(x) = P(x_1)P(x_2)P(x_3)P(x_n) \quad (2)$$

where "$x = (x_1, x_2, \ldots, x_n)$" are a set of features. Individual probabilistic classification of a feature may be calculated as given in equation 3:

$$P(x) = \frac{p(C_k)p(x|C_k)}{p(x)} \quad (3)$$

### 3.3.2 Decision Tree (DT)

The working principle of DT is based on a hierarchical breakdown of the data set used for training. In this classifier, features are used for labeling tree nodes, and the branches between them are given the weight representing the occurrence of feature in the test data; finally, class names are assigned to the leaf. The data set is divided into the presence or absence of features. The data set is divided recursively until the leaf nodes are reached.

Entropy Formula:

$$Entropy = -\sum_{j=1}^{m} P_{ij} log2P_{ij} \quad (4)$$

### 3.3.3 Random Forest (RF)

RF is a voting method where many decision trees are grown simultaneously. The input features are fed to individual trees in the forest. The final classification is based on the overall most votes from all trees in the forest [30]. The mathematical form of random forest to calculate mean square error is:

$$f^{\wedge} = \sum_{s=1}^{S} \frac{1}{S}(fs - ys)^2 \quad (5)$$

Where "$S$" denotes the number of data points, "$fs$" is the value returned by the model, and "$ys$" is the actual value of data points.

### 3.3.4 Support Vector Machine (SVM)

SVM is a classification algorithm that finds the maximum margin hyperplane to classify the "$i^{th}$" vector. Optimal "$y_i$" ($y_i$ denotes the target), "$X_i$" hyperplane is found by linear features between two classes (0 or 1).

### 3.3.5 K-nearest neighbor (KNN)

KNN is an instance-based algorithm that assumes that similar things exist in close proximity. In this technique, the feature is classified by the plurality vote of its neighbors by calculating their distances. It uses Euclidean distance formula to compute the distance between the points, which is mathematically represented as:

$$D = \sqrt{\sum_{i=1}^{m}}(x_i - y_i)^2 \quad (6)$$

### 3.3.6 Logistic Regression

Logistic Regression is a model-based algorithm often used when the dependent variable is dichotomous in nature. However, it can be tuned to be used with multiclass classification tasks as well. Logistic Regression describes the data set and defines the relationship between one dependent binary variable with one or more independent variables.

### 3.4 Testing Metrics

Accuracy, precision, recall, and f-score are used to evaluate model performance. These metrics can be defined as:

$$Accuracy = \frac{TuP + TuN}{TuP + TuN + FaP + FaN} \quad (7)$$

$$Precision = \frac{TuP}{TuP + FaP} \quad (8)$$

$$Recall = \frac{TuP}{TuP + FaN} \quad (9)$$

$$F - Score = \frac{2 * (recall * precision)}{(recall + precision)} \quad (10)$$

where "TuN", "TuP", "FaN", and "FaP" are true negative, true positive, false negative, and false positive respectively.

## 4 Experimental Results, Discussion and Evaluation

This section describes the experimental results, discusses the results, and evaluates the developed model quantitatively. Six machine learning algorithms (Naïve Bayes, decision tree, Random Forest, SVM, K-nearest neighbor, and Logistic Regression) were used to develop the model using three feature extraction techniques (LIWC, n-gram (unigrams), and sentiment score).

The results of individual feature and their combinations are shown in table 2. An accuracy of 63.22% is achieved when LIWC is used alone, but when combined with a sentiment score, accuracy increases to 70.35%. Classification model using unigram feature alone gives an accuracy of 73.55%, but if combined with sentiment score, it increases to 80.34%. Maximum accuracy of 88.76% is attained by combining LIWC, unigram, and sentiment scores. Eventually, this study supports and proves the initial hypothesis of getting improved results by using amalgamated features with machine learning algorithm-based classification models.

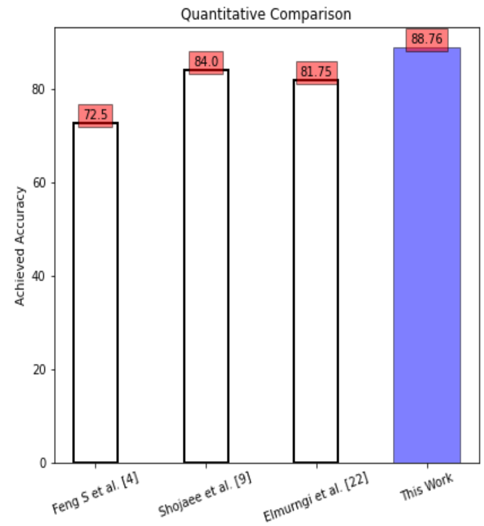



**Fig. 2.** Quantitative Comparison with Previous Similar Work

### 4.1 Qualitative Evaluation

To strengthen the hypothesis, Figure 2 shows the quantitative comparison of previous work [4][9][22] and the proposed method in this work to detect fake reviews using the Hotel reviews dataset. It indicates that the undertaken work supersedes the other work. Since a balanced dataset is used, the accuracy score measure is good enough for quantitative performance comparison of the developed machine learning algorithm-based models with previous work. This improvement is, at least, 4-5 points when compared with previous similar work.

## 5 Conclusin

This work was an effort to determine the effective combination of features that performs well for fake reviews detection. The study work used n-gram, LIWC, and sentiment score features for training purpose. Different classifiers are trained on these features. The classification algorithms we chose are described in section 3.3. After experimental work, logistic Regression outperformed all other machine learning algorithm-based models. As far as features performance is concerned, unigram proved to be better when applied in separation than LIWC. However, the combination of both (unigram + LIWC) with sentiment score performed more adequately, giving maximum accuracy of 88.76%. This result is better than some other techniques described in section 2. For future work, it is suggested to use semi-supervised learning to check the accuracy and performance of unigram and LIWC features on the fake review detection method. In this way, possible performance enhancement will be measured. At the same time, the limitation of the labeled data set for supervised learning will also be resolved.

TABLE II. PERFORMANCE EVALUATION OF DIFFERENT FEATURE EXTRACTION METHODS

| Approaches (Features) | Maximum accuracy (%) | Precision | Recall | F-score |
|---|---|---|---|---|
| **LIWC** | 63.22 | 58.00 | 64.43 | 61.05 |
| **Sentiment score, LIWC** | 70.35 | 62.50 | 73.88 | 67.71 |
| **Unigram** | 73.55 | 78.00 | 74.50 | 76.21 |
| **Sentiment score, Unigram** | 80.34 | 91.50 | 77.59 | 83.97 |
| **Sentiment score, LIWC, Unigram** | 88.76 | 92.00 | 82.61 | 87.05 |

## REFERENCES


[1] "• U.S. e-commerce share of retail sales 2021-2025 | Statista." https://www.statista-.com/statistics/379112/e-commerce-share-of-retail-sales-in-us/.

[2] F. Li, M. Huang, Y. Yang, and X. Zhu, "Learning to identify review spam," IJCAI Int. Jt. Conf. Artif. Intell., pp. 2488–2493, 2011, doi: 10.5591/978-1-57735-516-8/IJCAI11-414.

[3] R. Y. K. Lau, S. Y. Liao, R. Chi-Wai Kwok, K. Xu, Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detection," ACM Trans. Manag. Inf. Syst., vol. 2, no. 4, Dec. 2011, doi: 10.1145/2070710.2070716.

[4] S. Feng, L. Xing, A. Gogar, and Y. Choi, "Distributional footprints of deceptive product reviews," ICWSM 2012 - Proc. 6th Int. AAAI Conf. Weblogs Soc. Media, pp. 98–105, 2012.

[5] Sussin, J., and E. Thompson, "The consequences of fake fans,'Likes' and reviews on social networks," Gart. Res., vol. 2091515, 2012.

[6] "Amazon sues to block alleged fake reviews on its website | Reuters." https://www.reuters.com/article/us-amazon-com-lawsuit-fake-reviews-idUSKBN0N02LP20150410 .

[7] "Local Consumer Review Survey 2022: Customer Reviews and Behavior." https://www.brightlocal.com/research/local-consumer-review-survey/ .

[8] N. Spirin and J. Han, "Survey on Web Spam Detection : Principles and Algorithms," vol. 13, no. 2, pp. 50–64.

[9] N. M. S. and S. N. Somayeh Shojaee, Masrah Azrifah Azmi Muradt, Azreen Bin Azman, "Detecting Deceptive Reviews Using Lexical and Syntactic Features," pp. 219–223, 2013.

[10] Y. Yao, B. Viswanath, J. Cryan, H. Zheng, and B. Y. Zhao, "Automated crowdturfing attacks and defenses in online review systems," Proc. ACM Conf. Comput. Commun. Secur., pp. 1143–1158, 2017, doi: 10.1145/3133956.3133990.

[11] K. C. Santosh and A. Mukherjee, “On the temporal dynamics of opinion spamming: Case studies on yelp,” 25th Int. World Wide Web Conf. WWW 2016, pp. 369–379, 2016, doi: 10.1145/2872427.2883087.

[12] S. Nilizadeh et al., “Poised: Spotting twitter spam off the beaten paths,” dl.acm.org, pp. 1159–1174, Oct. 2017, doi: 10.1145/3133956.3134055.

[13] H. Chen, “Toward Detecting Collusive Ranking Manipulation Attackers in Mobile App Markets,” pp. 58–70, 2017.

[14] D. Y. T. Chino, A. F. Costa, A. J. M. Traina, and C. Faloutsos, “VOLTIME: Unsupervised anomaly detection on users’ online activity volume,” Proc. 17th SIAM Int. Conf. Data Mining, SDM 2017, pp. 108–116, 2017, doi: 10.1137/1.9781611974973.13.

[15] R. Deng, N. Ruan, R. Jin, Y. Lu, and W. Jia, “SpamTracer : Manual Fake Review Detection for O2O Commercial Platforms by Using Geolocation Features,” pp. 1–20.

[16] N. Jindal and B. Liu, “Review spam detection,” 16th Int. World Wide Web Conf. WWW2007, pp. 1189–1190, 2007, doi: 10.1145/1242572.1242759.

[17] J. Li, M. Ott, C. Cardie, and E. Hovy, “Towards a General Rule for Identifying Deceptive Opinion Spam,” pp. 1566–1576, 2014.

[18] A. Mukherjee, V. Venkataraman, … B. L.-S. international A., and U. 2013, “What Yelp fake review filter might be doing?,” Proc. Int. AAAI Conf. Web Soc. Media, vol. 7, no. 1, pp. 409–418, 2011.

[19] B. Liu and H. W. Lauw, “Detecting Product Review Spammers using Rating Behaviors,” pp. 939–948, 2010.

[20] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” ACL-HLT 2011 - Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol., vol. 1, pp. 309–319, 2011.

[21] M. Ott, C. Cardie, and J. T. Hancock, “Negative deceptive opinion spam,” NAACL HLT 2013 - 2013 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Main Conf., no. June, pp. 497–501, 2013.

[22] E. Elmurngi and A. Gherbi, “An empirical study on detecting fake reviews using machine learning techniques,” 7th Int. Conf. Innov. Comput. Technol. INTECH 2017, no. Intech, pp. 107–114, 2017, doi: 10.1109/INTECH.2017.8102442.

[23] M. N. I. Ahsan, T. Nahian, A. A. Kafi, M. I. Hossain, and F. M. Shah, “An ensemble approach to detect review spam using hybrid machine learning technique,” 19th Int. Conf. Comput. Inf. Technol. ICCIT 2016, pp. 388–394, 2017, doi: 10.1109/ICCITECHN.2016.7860229.

[24] W. Etaiwi and G. Naymat, “The Impact of applying Different Preprocessing Steps on Review Spam Detection,” Procedia Comput. Sci., vol. 113, pp. 273–279, 2017, doi: 10.1016/j.procs.2017.08.368.

[25] M. B. Alvi, N. A. Mahoto, M. A. Unar, and M. A. Shaikh, “An Effective Framework for Tweet Level Sentiment Classification using Recursive Text Preprocessing Approach,” no. July, 2019, doi: 10.14569/IJACSA.2019.0100674.

[26] M. B. Alvi, N. A. Mahoto, M. Alvi, M. A. Unar, and M. Akram Shaikh, “Hybrid classification model for twitter data-A recursive preprocessing approach,” 5th Int. Multi-Topic ICT Conf. Technol. Futur. Gener. IMTIC 2018 - Proc., 2018, doi: 10.1109/IMTIC.2018.8467221.

[27] C. G. Harris, “Detecting deceptive opinion spam using human computation,” AAAI Work. - Tech. Rep., vol. WS-12-08, pp. 87–93, 2012.

[28] P. Cavallo et al., “Journal of Software,” vol. 9, no. 8, 2018.

[29] M. Ben-bassat, K. L. Klove, and M. A. X. H. Weil, “CALO ( x = ALO ( x ) e,” vol. 2, no. 3, pp. 261–266, 1980.

[30] A. Akhter, M. B. Alvi, and M. Alvi, “Forecasting Multan estate prices using optimized regression techniques,” Univ. Sindh J. Inf. Commun. Technol. , vol. 5, no. 4 SE-Computer Science, Apr. 2022, https://sujo.usindh.edu.pk/index.php/USJICT/article/view/4340.

[31] G. Mujtaba and E. S. Ryu, “Client-Driven Personalized Trailer Framework Using Thumbnail Containers,” IEEE Access, vol. 8, pp. 60417–60427, 2020, doi: 10.1109/ACCESS.2020.2982992

Vol. 5, No. 2 | July – December 2022

# Wireless Sensor Networks-based Smart Agriculture: Sensing Technologies, Application and Future Directions

Mushtaque Ahmed Rahu[1], Sarang Karim[2]*, Rehan Shams[3]*,
Ayaz Ahmed Hoshu[4], Abdul Fattah Chandio[1]

**Abstract:**

*With the advent of the latest sensing technologies and Wireless Sensor Networks (WSN), agricultural tasks can be performed quickly, adequately and precisely. These practices are termed Smart agriculture. In this paper, we discuss various sensing technologies that enable smart agriculture. Later, a system based on WSN has been designed to monitor agricultural parameters. The proposed system has been deployed in a Wheat field. This work aims to increase the quality and productivity of the Wheat crops and minimize the extensive field visits of the farmers. This system enables precision agriculture by periodically measuring the three most key parameters (temperature, light, and water level) for achieving a remarkable increase in quality, productivity and growth of the Wheat crops. Thus, this system helps the agriculturists, landowners and research experts to monitor these parameters at the base station without going to the field site. A GUI tool is also designed to display the measured data and stored it in the database accordingly. While designing this system; IRIS mote, MDA100 data acquisition board, and MIB520 USB interface board are employed. We use TinyOS operating system for the development of codes for wireless nodes and the GUI tool is designed in Microsoft Visual Studio. ZigBee IEEE 802.15.4 protocol and direct topology are used for the communication of nodes with the base station. In last, we also discuss future research directions.*

**Keywords:** *Wireless sensor networks, smart agriculture, precision agriculture, sensing technologies, Wheat crops.*

Pakistan, being an agricultural country stabilizes its economy through agricultural projects. The economy of most of the population depends on the outputs gained from the agriculture sector. It fulfils the major ingredient of food for mankind and other living organisms on earth. Modern technology can bring a remarkable increase in the production and quality rate of crops. Through modern techniques and technologies not only the human efforts can be reduced but agricultural expenditures from sowing to harvesting can also be minimized. Two basic needs of humans can only be fulfilled from

[1]Department of Electronic Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah 67450, Pakistan.
[2]Department of Telecommunication Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah 67450, Pakistan.
[3]Department of Telecommunication Engineering, Sir Syed University of Engineering and Technology, Karachi 75300 Pakistan.
[4]Department of Electronic Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Larkana Campus, Larkana 77150, Pakistan.
Corresponding Author: r.shams@ssuet.edu.pk and sarangkarim@quest.edu.pk

agricultural resources, which are food and clothes. Hence, modern techniques and technologies always do come up with the best solution for higher productivity and quality of cash crops like wheat, rice, cotton, sugar-cane and vegetables [1]. For enhanced productivity with low cost, a Wireless Sensor Network (WSN) provides a variety of solutions for real-world challenges. WSN has proved its significance very effectively in various other applications like structural and health monitoring of highways and buildings, habitats and irrigation [2]. Meanwhile, WSN can also be equally beneficial in precision agriculture for monitoring the quality parameters of the crops and helping the farmers and landowners in cultivation procedures as well [3].

A management framework in which suitable strategies and plans based on information and communication technologies (ICT) are made to handle the agricultural practices is called precision agriculture or precision farming [4], [5]. The impacts of precision agriculture on improving crop profitability are addressed in [6]. The principle aim of precision agriculture is to increase crop productivity and quality and reduce the efforts of the farmers using ICT systems and algorithms [7]. Precision agriculture is somehow related to data monitoring of an agricultural field [8]. Typical attributes of precision agriculture are [7], [9]:

- Monitoring of agricultural environmental parameters.
- Appropriate set point for data collection.
- Data transmission from source (field) to destination (sink).
- Control actions and decision-making as per sensed data.

In this paper, we portray novel deployment architecture of smart agriculture, which depicts the shape of future digital agriculture. The architecture is comprised of various key technologies, for example, renewable energy sources, sensor classifications, and other emerging technologies. We also provide a detailed discussion on different sensors and sensing technologies is carried out. Later, we perform a case study on the wheat crop. In which, we design a WSN-based system to monitor the three most key parameters (temperature, light, and water level). In last, we discuss the research gaps in this study, which can be implemented in the forthcoming paper. The readers and researchers can also integrate our suggested research gaps in their work.

The remaining sections of the paper are described as follows: Section 2 describes the previous research work based on WSN that had been carried out by different researchers. Section 3 gives the details about the term smart agriculture followed by different sensors and sensing technologies. In section 4, we describe the general block diagram of the system followed by the complete picture of the sensor node with IRIS mote and deployment scenario. Section 5 gives the software implementation description, which is also divided into two sub-sections: implantation in TinyOS and implementation in MS Visual Studio. The deployment results are discussed in section 6. Section 7 is based on general discussions, for example, the requirement of Wheat crops, different users, and future directions. Conclusion is given in section 8.

Exhaust research work on WSN-based precision agriculture has been carried out by different researchers. Most of the research work on precision agriculture is based on various application scenarios like agricultural monitoring, data monitoring, greenhouses, pest detection, soil monitoring and irrigation management [5], [8], [12]–[18] by incorporating different methodologies, techniques, tools and platforms [19]–[21] along with basic guidelines [22]. In [23], a WSN-based system was developed for monitoring the fluoride-affected area. Their system enables the users to access the status of fluoride sensors at the remote station on their cell phones via the Internet. Government organizations and ordinary people can also make use of their systems to monitor the affected areas. Nowadays, mobile devices such as robots and drones are engaged in agricultural monitoring.

In [14], WSN-based aerial robots were employed. Aerial robots were dedicated to vineyards for frost monitoring. The core task of the aerial robot is to provide dynamic mobility to the nodes in vineyards for sensing and creating a communication link between the base station and scrubby clusters sited at disjointed points. Their system is very suitable for such areas where wireless sensor networks have some limitations for such characteristics. By using the dedicated communication channel, data can be routed even at long distances. In [13], an instrumentation setup has been made for monitoring the critical inputs of Wheat crops such as water and Nitrogen. Their system was employed for collecting field data continuously. Besides, agricultural environmental monitoring, a suitable irrigation plan is also needed for precision agriculture [9]. Providing an adequate amount of water in meantime plays a dramatic role in improving crop quality and quantity. In addition, some basic guidelines must be considered before the deployment of WSN in an agricultural field and useful descriptions, such as SOPs (Standard-Operating-Procedures) of a crop must be reviewed for proper crop monitoring and pre-and-post processing actions.

In [22], authors have reviewed different existing components of WSNs for precision agriculture, viz. mote platforms, sensor types, operating systems, communication issues, maintenance, power supply, etc. After a deep review of these data, basic guidelines have been proposed for the deployment of WSNs in any application scenario related to agriculture. In [16], authors have performed real-time deployment for monitoring the agricultural land. They have designed a WSN system based on IRIS motes to monitor the humidity, temperature and light intensity. TinyOS and MS Visual Studio were used to program the IRIS motes and a GUI tool for displaying the results respectively.

In [17], authors have carried out two operations for agricultural monitoring. Firstly, they designed an irrigation system for watering agricultural land automatically. Then after, they designed a system to sense the key parameters, Potassium (K), Nitrogen (N) and Phosphorus (P) for improving crop quality and quantity. Their system was managed via an ARM 7 Processor and the monitored data were sent to the remote station with the aid of IoT (internet of things). In WSN algorithms can improve the data packet transmission efficiency by up to 25% and also helps to prolong the lifetime of Wireless Sensor Networks to achieve efficient data transmission [49].

In [50], authors provided a survey of tending resource-efficient and secure techniques used with distributed estimation algorithms over WSN. In [39], authors have performed a short survey on the implementations and usage of artificial intelligence in smart agriculture. In addition, various machine learning techniques for smart agriculture are also discussed. Some other survey papers related to smart agriculture can be found in [41], [42], in which, the authors have discussed different challenges and proposed some solutions to be faced in smart agriculture.

Fig. 1 portrays the smart agriculture deployment architecture equipped with modern techniques, paradigms, and technologies in order to provide strong technical support in speeding the agricultural transformation and development [30]. The smart agriculture system collects the key information from the cultivated field via sensing and communication devices [31]. The key information includes soil pH, soil moisture, humidity, temperature, water level, and so on (more classifications of agricultural sensors are given in Table.1). Agricultural production increases by continuous monitoring of these parameters [31].

The smart agriculture system requires immense quantities and different types of sensor nodes to cover the whole field. These sensor nodes are tiny in size with limited battery power [32]. So, whenever the battery of these nodes is depleted, they halt the sensing operation, and the network lifespan is shortened. Hence, it is essential to provide a continuous supply of energy to the smart agriculture system, for which it would be better to opt and energy harvesting techniques for avoiding the battery depletion of these low powered sensor nodes.

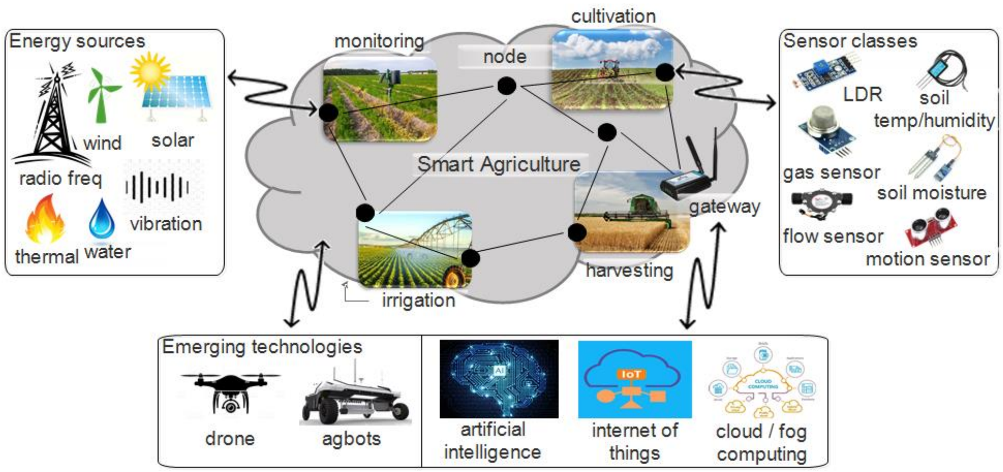



**Fig. 1.** Deployment architecture of smart agriculture.

There are various energy harvesting techniques, such as solar, wind, thermal, vibrations, radio frequency (RF), water, and so on [33]–[35], [40]. The on-board field information collection and consequent actions are the great challenges in agriculture. The data collected from a large field or crop monitoring requires extensive time and effort. For example, performing soil analysis with crop and environmental monitoring requires regular field trips or even multiple times for certain crops. This is highly exhaustive as it requires too much time and manpower along with expenditures to collect the sample and on-field data.

Thus, thanks to the sensing technologies, which collect the field data and monitor the crops by providing comfort to the farmers. There are several sensing devices and kits in hundreds of thousands and are available in different taxonomies. These sensor taxonomies include [38]: i) Measurement, ii) Technology, iii) Material, iv) Operating principle, v) Conversion techniques and vi) Application areas. We classify the agricultural sensors into these groups i) Chemical, ii) Physical, and iii) Mechanical. Table 1 includes an additional breakdown of these sensors.

### 3.1.1 Sensing Technologies

Sensing technologies experience explosive creativity, activities, exciting applications, and innovations in the agricultural sector. Various technological firms and entrepreneurs show their diversity and willingness to enter a gigantic field of agriculture. Here we discuss some sensing technologies along with their extensive products and their sensory role in the agricultural sector. Table 1 provides a review on different agricultural sensing technologies.

TABLE I. AGRICULTURAL SENSORS CLASSIFICATION

| Class | Type | Classification |
|---|---|---|
| Chemical sensors | pH sensor | Soil and water quality monitoring |
| | Biosensor | Glucose and acids |
| | Gas sensor | Pollution and air quality |

| | | |
|---|---|---|
| Physical sensors | Temperature sensor | Soil, plant, crops, and environment |
| | Humidity sensor | Soil and environment |
| | Watermark sensor | Soil humidity |
| | Rain sensor | Environmental monitoring |
| | Electrical conductivity | Soil monitoring |
| | Leaf wetness sensor | Tress, crops, and plants |
| | Terrestrial sensor | Weather and environmental monitoring |
| | Color sensor | Nutrient monitoring |
| | Passive Infra-red | Environmental monitoring |
| | Underwater sensor | Salinity, solvents, and quality |
| | Underground sensor | Soil compaction and moisture |
| | Solar radiation | Crops and plants |
| Mechanical sensors | Pest detector | Pest and bug detection |
| | Pressure sensor | Soil compaction |
| | Vibration sensor | Soil and atmosphere |
| | Wind sensor | Speed and direction of air |
| | Motion sensor | Environmental monitoring |
| | Water flow sensor | Irrigation |
| | Water level sensor | Ground and underground |

#### 3.1.2 Libelium Smart Agriculture Xtreme [58]

This smart agriculture tool kit offers numerous onboard sensors developed by Libelium. This kit allows various communication protocols, e.g., LoRaWAN (Long Range Wide Area Networks), Wi-Fi (Wireless Fidelity), ZigBee, Sigfox, and 4G. This Smart Agriculture Xtreme seeks a variety of applications based on IoT, for example, precision farming, greenhouses, weather station, and irrigation. The main advantage of Libelium kits is that they are solar powered.

#### 3.1.3 Smart Pot [51]

Parrot Pot has introduced a smart pot, especially for indoor farming within an in-situ Bluetooth communication module. This smart pot is also a suitable ingredient for greenhouses and urban farming. Various sensors are embedded in this smart pot, for example, sunlight, soil moisture, temperature, and fertilizer levels, which help in continuous monitoring of the plants.

#### 3.1.4 Open Garden platform [52]

Open Garden platform has capabilities to monitor three different setups: 1) indoor farming (greenhouse and vertical farming) 2) outdoor farming (crops, gardens, etc.), and 3) hydroponics (water sensors). The Open Garden platform can be equipped with multicommunication modules, such as 3G, Wi-Fi, or GPRS module. The sensed data can be propagated via any of these communication protocols. The indoor/outdoor setup kits include temperature, humidity, light levels, drip (indoor) / sprinkler (outdoor) water pumps, and so on. The hydroponics kit includes many water sensors: moisture, humidity, conductivity, pH, growing light, oxygen pump, etc.

#### 3.1.5 X-ray Computed Tomography [54]

X-ray Computed Tomography (CT) has the capabilities to examine water content, conductivity, and configuration of the soil by using images. In [56], authors have used X-ray CT images to quantify the hydraulic conductivity of bulk soil. Another work using CT images was reported in [57] to air- and water-filled pore space and structure of the soil. In [55], a novel work on X-ray CT and image processing was reported to calculate the root-water absorption in the soil.

### 3.1.6 The Arable Mark [59]

The Arable Mark has designed an irrigation management kit for smart agriculture. Farmers can use this tool to irrigate their crops properly and precisely. This tool acquires the field data, such as soil moisture, crop water scarcity, and precipitation in order to provide the required water to crops. This helps in avoiding over- and under-irrigation.

### 3.1.7 GreenIQ [61]

The Eastern Peak provides an agricultural field device for irrigation purposes and is named GreenIQ. This is suitable for greenhouses and house gardens. It irrigates the gardens using a sprinkler controller.

### 3.1.8 Grofit climate monitoring system [60]

The Grofit introduced this system for agricultural climate monitoring, it offers a Bluetooth based monitoring system with a propagation area of up to 200 meters. This can measure air humidity, temperature, and solar radiation. Grofit also provides a cloud to access information related to their devices.

### 3.1.9 MeteoHelix IoT Pro [63]

The allMeteo designed MeteoHelix IoT (Internet of Things) Pro hardware, especially for agricultural environmental monitoring based on IoT and WSN. It offers a broad number of sensing parameters, for example, solar radiation, rain gauge, pressure, pollution, dew, humidity, temperature, and so on. This is only suited for terrestrial monitoring. This greatly helps in solving the meteorological solutions.

### 3.1.10 SKY-LoRa Weather Station [64]

Farmers can use this weather station kit to forecast any uncertain environmental and weather occurrences, such as rain and wind. This tool kit uses the LoRa communication protocol to communicate with the master sensor node. Its communication coverage is approximately 600 m.

### 3.1.11 EC1 Speed controller [65]

This device performs its operation upon monitoring the atmospheric conditions and then toggling the condition of different devices. This controller helps to control various agricultural field devices and equipment by turning on/off the machines.

The main system is elaborated in Fig. 2. Numerous devices come together to make the main system. These devices include IRIS motes [10], various sensors and a display screen to monitor the data. In this research, our nodes are connected wirelessly in a way of direct topology to communicate with the sink located at the base station. Each IRIS


**Fig. 2.** Block diagram of the system

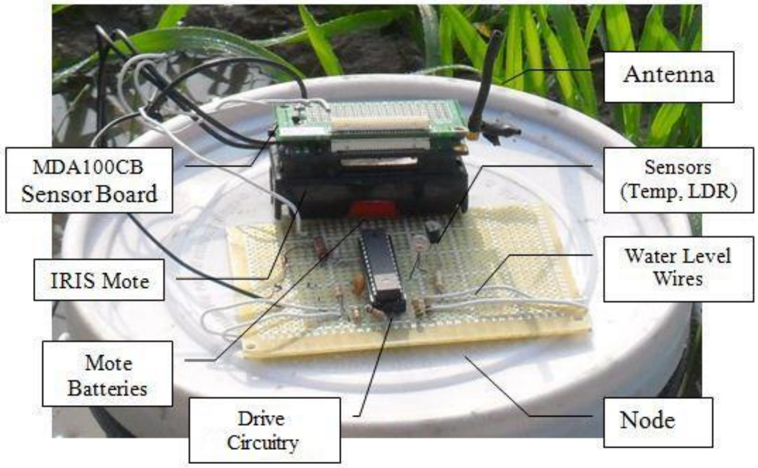



**Fig. 3.** Complete picture of a sensor node

mote is connected to a sensor board. The main control station collects sensed information from individual nodes, which is later on directed to the sink node. MIB520CB USB board [11] is used to connect the sink node with the computer via a serial USB port. GUI interface is provided for the visualization of measured data to users.

### 4.1 Sensor Node with IRIS

Fig. 3 shows the complete picture of the sensor node with its peripherals. Its peripherals include IRIS motes [10], MDA100CB Board [25], Temperature sensor [26], LDR sensor, water level sensor and drive circuitry. For this work, three sensor nodes are used, which are wirelessly connected via direct topology with a sink or base station with the aid of IRIS motes. Sensor nodes can acquire temperature, LDR and water level data of irrigated Wheat fields. The base station is connected to a computer, where acquired data is monitored and stored in a database for post-processing aspects. Batteries keep the sensor nodes alive till the work is done.

### 4.2 Deployment of nodes

We performed the deployment of sensor nodes in an agricultural field of the Wheat crop (see Fig. 4). Fig. 4 shows the real-time deployments of the sensor nodes. The sink was about 50-60m away from the sensing field, while sensor nodes were placed at a distance of about 20-30m from each other.

**Fig. 4.** Physical deployment site of sensor nodes

We deployed various sensor nodes consisting of temperature, LDR and water level sensors to sense the agricultural data of Wheat crops. Experiments were conducted for 10 (ten) weeks, consecutively to obtain the results from the sensing field. The nodes transmit the sensed data via direct topology [15], [28]. The sink node is directly connected with all nodes for access to the sensed data.

### 5.1 Implementation in TinyOS

A small operating system named TinyOS [24] is used to program the IRIS motes by interfacing the MIB520 interface board with a computer via serial communication. The code for IRIS motes is developed in the Nested C language. Despite being a small Linux-based operating system TinyOS comes with various built-in applications, which makes the programming of IRIS motes much easier. We programmed four IRIS motes, one for the base station (as a sink) and the other three for the sensing field (as a sensor node). Each IRIS possesses a unique ID.

The programming flowchart for sink and sensor nodes is depicted in Fig. 5. The sink node is serially connected to a computer/ laptop with the help of MIB520CB programmable board. The data is displayed in a GUI tool and in the meantime stored in a database so that users can easily observe the measured data. Whereas, sensor nodes are deployed at the field site (as shown in Fig. 4). Once the sink nodes receive the data from sensor nodes, it will display that measured data on the GUI tool. The sensor nodes measure the temperature, LDR, and water level of the Wheat field and then each sensor node transmits the measured data to the sink node using direct topology. Sink nodes and sensor nodes can communicate with each other via a wireless medium.

### 5.2 Implementation in MS Visual Studio

The main GUI tool (refer to Fig. 6) is developed in Microsoft Visual Studio [29]. The purpose of designing the GUI tool is to display the measured data of each sensor of each node. In the GUI tool, all three sensor nodes are recognized with node ID and three sensors (temperature in degrees Celsius, LDR in light/dark (in form of LUX in the database) and water level in inches). The data received by the sink node from the sensor nodes are sent to a computer for display on the GUI tool through serial communication via a USB port. At first, GUI is loaded, in which default values are displayed then for receiving the measured data, the sink node must be interfaced serially via MIB520 USB interface board with PC by initializing the port on which it is connected. After the initialization process, the measured data of each node is displayed on the GUI tool and updated into the database accordingly. The database can be imported into the data log window by clicking the button to display the data log available on the GUI tool. The database can also be viewed in MS Access: an office application.

This system has been deployed in the irrigated (shallow-water based) Wheat field for monitoring three different agricultural parameters (temperature, light and water level). The deployment results are depicted in Fig. 6. Fig. 6a. shows the temperature measured by three different sensor nodes located at different points in the Wheat field (refer Fig. 4) for consecutive 10 (ten) weeks.

### 6.1 Temperature Measurement

Fig. 6b shows the comparison between the two temperatures i.e. one is the actual temperature (i.e., standard temperature) of the deployment location and the other is the measured temperature by the sensor nodes. It is obvious that the average temperature as measured by the sensor nodes should be equal to the actual temperature. Well, a minor difference in readings is also acknowledged, this difference might be accorded due to different issues, such as power consumption and the distance between the sink node and sensor nodes. It is seen that the average difference between the two mentioned temperatures is 3-4 degree Celsius. This difference can be reduced and more accurate and reliable results can be obtained if the temperature sensors are calibrated properly and by minimizing the interferences, network issues, etc.

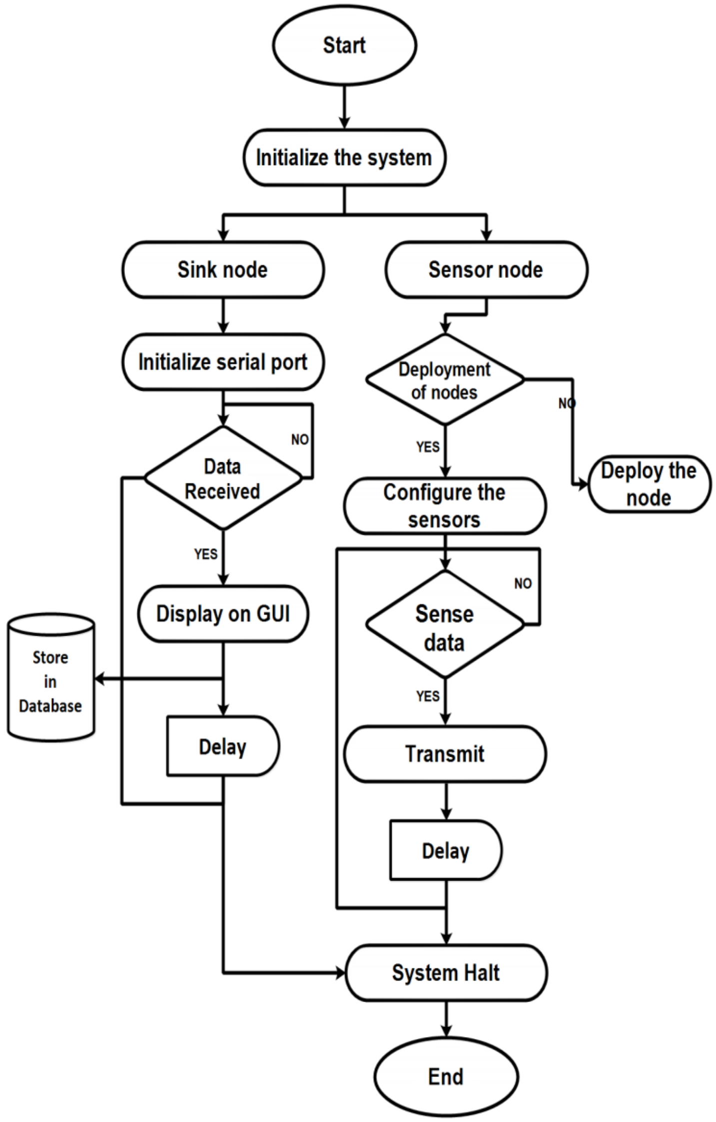
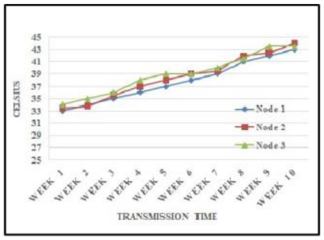
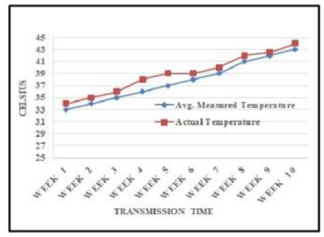


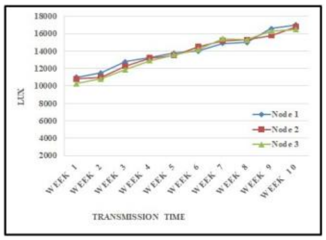

**Fig. 5.** Flowchart for sink and sensor nodes



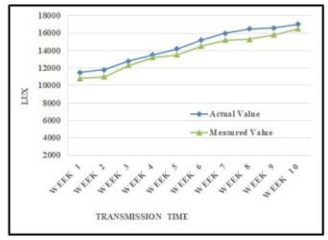

(a) Temperature measurement



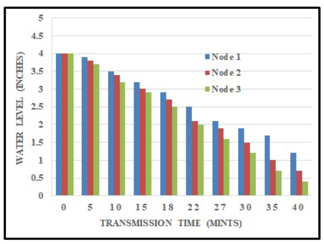

(b) Temp. measurement (actual v/s measured)



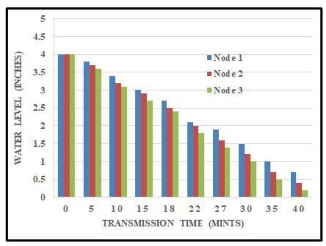

(c) LI measurement



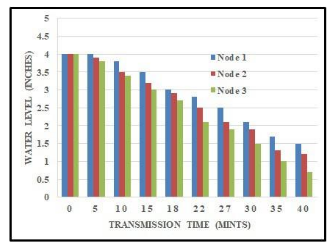

(d) LI measurement (actual v/s measured)




(e) 1st irrigation




(f) 2nd irrigation




(g) 3rd irrigation

**Fig. 6.** Deployment results

### 6.2 Luminous Intensity (LUX) Measurement

Fig. 6c describes sunlight detection in terms of Luminous Intensity (LI) for 10 consecutive weeks. We measured the LI during day time because there is no sunlight at night. It is measured in LUX. The LI is very necessary for crops to complete the process of photosynthesis [27]. Hence, the crops are dependent on sunlight for their growth. Different states of sunlight at different moments are mentioned in the form intensity unit LUX in Fig. 6c. Fig. 6d. shows the comparison between measured LI versus actual LI.

### 6.3 Water Level Measurement

Fig. 6e, Fig. 6f, and Fig. 6g plot the water levels of Wheat crops at different instant of times for 3 consecutive irrigations. We can observe that the readings for water level for all sensor nodes are different from each other with the passage of time; this is due to the rough surface layer of the Wheat crop field. Henceforth, by measuring water level, we can irrigate our crops accurately, precisely and timely. Moreover, it also helps in avoiding over and under irrigation. Thus, an ample amount of water can be saved. We provide a summary of the WSN-based smart agriculture application for Wheat crops, which is given in Table. 2.

TABLE II. SUMMARY OF THE WSN-BASED SMART AGRICULTURE APPLICATION FOR WHEAT CROPS

| # | Parameter | Module | Description | Remarks |
|---|---|---|---|---|
| 1 | Pre-deployment | Input | Budget = good<br>Outdoor<br>Crop = Wheat<br>Period = 10 weeks<br>Area = 100m$^2$<br>Node density = 4<br>Random deployment | Ample amount is available<br>Land/Field<br><br>One crop season<br><br>1 sink node and 3 sensor nodes |
| 2 | Sensors | Input | Temperature<br>Water level<br>LDR | |
| 3 | Hardware and software | Input | Sink/sensor node = IRIS motes<br>MDA 300<br>Water level sensor<br>LDR<br>OS=TinyOS | Different applications<br>Temperature sensor<br>For irrigation<br>For sunlight detection<br>Linux based OD |
| 4 | Communication | Communication | Topology = tree, direct<br>Data rate = 1 message/30sec<br>Node-node = Zigbee<br>Node-sink = Zigbee<br>BS-Server = WiFi<br>Server-User = Internet | Also depends on the deployment area<br>Application specific |
| 5 | Energy | Input | Rechargeable batteries | Can be replaced easily |
| 6 | Safety and maintenance | Output | Plastic casing<br>Security cameras | Few can be used across the field |

### 7.1 Requirement of Wheat Crop

The Wheat crop essentials and requirements are described in Fig. 8. According to it, the major requirements for the crops like Wheat are temperature, sunlight and water. So far, by keeping these parameters under consideration; the production rate and crop length can easily be estimated. If all or any of these significant parameters become less/more and are not maintained at their proper time, it will directly affect the production and quality rate of the crops, which is further clarified in Fig. 8. For further assistance and guidance, different Wheat researchers have provided the Wheat crop SOPs (Standard-Operating-Procedures), for example, season, crop length, irrigation period, harvesting time, and so on [62].

### 7.2 Users

There are two major users for the analysis of the results obtained during this work and that are farmers and researchers (or experts). The prime goal of this work is to provide essential up-to-date crop data and information to the farmers, landowners and researchers.

| Parameters | Devices | Requirments | Comments |
|---|---|---|---|
| Temperature |  | $80^\circ$ to $100^\circ$ F | temperature $\propto$ crop growth |
| LDR |  | Crop Length: 120 to 140 days | Light (sunny, dry, hot ) = crop length decreases<br>Dark (cloudy, fog, snow) = crop length increases |
| Water Level |  | 5 to 6 Irrigations | 3 inch/irrigation |

**Fig. 7.** Wheat crops essentials

**Farmers:** These are the principal and main users of this system. In Pakistan, the farmers are either uneducated or have little education. That's why in this system a simple GUI tool is used to display the results. By this, users can easily understand the results and take necessary actions accordingly.

**Experts:** The other users of this system might be highly professionals or agricultural experts. These people belong to research or agricultural training institutes. They can analyze the data in form of graphs, charts or statistical views. That's why, the results are also provided in form of charts and graphs, according to the need of the user.

### 7.3 Future Directions

Sensor network technology is growing day by day, and numerous platforms and tools are introduced to minimize and reduce human interactions and contributions. This research work can be enhanced by using some/any of the following fields.

#### 7.3.1 Wireless Platforms

We can implement different sensor network platforms, such as, Wasp mote Platform [19], Libelium Platform [19], Fleck Platform [5] and other platforms [5], [20], [21] for future correspondences.

#### 7.3.2 Sensors

There are various other sensing devices, like physical sensors (humidity sensor, soil moisture, watermark sensor, and leaf wetness sensor), mechanical sensors (flow sensors, injectors, and valves) and chemical sensors (biosensor, gas sensor) available in the market [38]. By incorporating these sensors, this work can easily be enhanced.

#### 7.3.3 Dynamic Controller

In this research, we haven't used any sort of dynamic (mechanical) controller. Dynamic controllers are helpful to perform mechanical tasks such as sprinklers, irrigation monitoring, water flow control using valves, injection of pesticides, vibrators and so on [7], [8], [15].

#### 7.3.4 Networking

In the wireless sensor network field, networking plays a pivotal role in all application domains in which a different number of sensor nodes are being employed. The researchers and application designers decide the networking according to the application and requirement scenarios. Classical networking (star, tree, P2P, mesh, bus topologies) [5], [7], [28], [44], opportunistic networking [3], [5], [45], and cognitive radios [46] are different classes of networking. Drones can also be used for networking purposes, such as measuring, intruder detection, and localization [43], [66], [67].

In this paper, we presented various sensing technologies that enable smart agriculture. In addition, we have performed a case study on Wheat crops. This work aims to minimize efforts and problems, which are being faced by the farmers while cultivating their crops. This work is based on the real-time deployment of sensor nodes in an irrigated Wheat field. We use IRIS motes as the WSN platform for incorporating as a sink and as sensor nodes. The IRIS motes were programmed in TinyOS and equipped with temperature, LDR and water level sensors to measure the field data. We measured these parameters; because, the cropped length has been estimated by analyzing the temperature and LDR parameters and by measuring water level, over-irrigation can be avoided.

Meanwhile, the measured data is transmitted as a base station accordingly. Where measured data is displayed, and analyzed on a GUI tool, which is developed in MS Visual Studio 2008. The results can easily be understood by every user without concerning any research expert. The database is used to maintain up-to-date information about collected data from the sensor nodes. This work offers much relaxation to the users for the measurements of required parameters of the crops, their presence at the field site all the time is not required, just once time performs the deployment of the nodes and gets the measured data at their workplace, base station and so forth.

## Acknowledgment

The authors appreciate the contributions of Quaid-e-Awam University of Engineering, Science, and Technology (QUEST), Nawabshah and Sir Syed University of Engineering and Technology (SSUET), Karachi to execute this work.

## REFERENCES


[1] Ahsan, A., and Ahmed, B., ”Identification of the type of agriculture suited for application of wireless sensor networks”. Russian Journal of Agricultural and Socio-Economic Sciences, 2012, 12(12).

[2] Fahmy, Hossam Mahmoud Ahmad. ”WSN applications.” Concepts, Applications, Experimentation and Analysis of Wireless Sensor Networks. Springer, Cham, 2021. 67-232.

[3] Reyes-Yanes, Abraham, et al. ”Wireless Sensing Module for IoT Aquaponics Database Construction.” International Journal of Electronics and Electrical Engineering 9.2, 2021: 43-47.

[4] Joshi Sarang, P., and R. U. Chavhan. ”Monitoring of Agriculture Parameter through Sensors for Productivity Enhancement.” New ARCH-International journal of Contemporary Architecture 8.2, 2021, pp. 1552-1560.

[5] Karim, S., and Shaikh, F. K., ”Wireless Sensor Network-Based Smart Agriculture.” In Opportunistic Networking, CRC Press 2017, pp. 251-276.

[6] Goel, Swati, Kalpna Guleria, and Surya Narayan Panda. ”Machine Learning Techniques for Precision Agriculture using Wireless Sensor Networks.” SPAST Abstracts 1.01, 2021.

[7] Anurag, D., Roy, S., and Bandyopadhyay, S., ”Agro-sense: Precision agriculture using sensor-based wireless mesh networks.” In Innovations in NGN: Future Network and Services, 2008. K-INGN 2008. First ITU-T Kaleidoscope Academic Conference (pp. 383-388).

[8] Garcia-Sanchez, A. J., Garcia-Sanchez, F., and Garcia-Haro, J., ”Wireless sensor network deployment for integrating video-surveillance and data-monitoring in precision agriculture over distributed crops.” Computers and Electronics in Agriculture, 75(2), 2011, pp. 288-303.

[9] Singh, I., and Bansal, M., ”Monitoring water level in agriculture using sensor networks.” International Journal of Soft Computing and Engineering, 1(5), 2011, pp. 202-204.

[10] Habtoor, M.S. and Rowaihy, H., ”Performance evaluation of IRIS and LOTUS motes when communicating in real environments”. International Journal of Distributed Sensor Networks, 12(2), 2016, pp.7347158.

[11] Datasheet MIB520. Programmable board. http://www.memsic.com/userfiles/files/Datasheets/ WSN/6020-0091-04 a mib520cb-t.pdf, (last accessed on Febuary 20, 2022).

[12] Balendonck, J., Hemming, J., Van Tuijl, B. A. J., Incrocci, L., Pardossi, A., and Marzialetti, P., ”Sensors and wireless sensor networks for irrigation management under deficit conditions (FLOW-AID)”. In: Conference Proceedings CD of the International Conference on Agricultural Engineering”. Agricultural & Biosystems Engineering for a Sustainable World. EurAgEng (European Society of Agricultural Engineers). 2008.

[13] Devadas, R., Jones, S. D., Fitzgerald, G. J., McCauley, I., Matthews, B. A., Perry, E. M., and Kouzani, A. Z., ”Wireless sensor networks for in-situ image validation for water and nutrient management.” In ISPRS 2010: Proceedings of ISPRS Technical Commission VII Symposium,. Institute of Photogrammetry and remote sensing, Vienna University of Technology, 2010, pp. 187-192.

[14] Valente, J., Sanz, D., Barrientos, A., Cerro, J. D., Ribeiro, and Rossi, C., ”An air-ground wireless sensor network for crop monitoring.” Sensors, 11(6), 2011, pp. 6088-6108.

[15] Ruiz-Garcia, L., Lunadei, L., Barreiro, P., and Robla, I., ”A review of wireless sensor technologies and applications in agriculture and food industry:

state of the art and current trends." Sensors, 9(6), 2009, pp. 4728-4750.

[16] Rahu, M. A., Kumar, P., Karim, S., and Mirani, A. A., "Agricultural Environmental Monitoring: A WSN Perspective." University of Sindh Journal of Information and Communication Technology, 2(1), 2018, pp. 17-24.

[17] Raut, R., Varma, H., Mulla, C., and Pawar, V. R., "Soil Monitoring, Fertigation, and Irrigation System Using IoT for Agricultural Application." In Intelligent Communication and Computational Technologies, 2018, pp. 67-73.

[18] Toth, G., Hermann, T., da Silva, M. R., and Montanarella, L., "Monitoring soil for sustainable development and land degradation neutrality." Environmental monitoring and assessment, 190(2), 2018.

[19] Guide, Technical, "Agriculture Board 2.0". document version: v4.002/2013, Libelium Communications Distribuidas S.L. http://www.libelium.com. 2013.

[20] Kusy, B. "Opal sensor node: Computation and communication in WSN platforms." 2012.

[21] F. K. Shaikh, S. Karim, S. Zeadally and J. Nebhen, "Recent Trends in Internet-of-Things-Enabled Sensor Technologies for Smart Agriculture," in IEEE Internet of Things Journal, vol. 9, no. 23, pp. 23583-23598, 2022.

[22] Mampentzidou, I., Karapistoli, E., and Economides, A. A., "Basic guidelines for deploying wireless sensor networks in agriculture." In Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2012 4th International Congress on, 2012, pp. 864-869.

[23] Kumar, K. A., Lingam, P. G., and Rao, K. M., "Wireless Sensor Based Remote Monitoring System for Fluoride Affected Areas Using GPRS and GIS. International Journal of Computer and Technology (IJCTT) volume, 7, 2014, pp. 178-182.

[24] Levis, P., Madden, S., Polastre, J., Szewczyk, R., Whitehouse, K., Woo, A., and Culler, D., "TinyOS: An operating system for sensor networks." In Ambient intelligence, Springer, Berlin, Heidelberg. 2005, pp. 115-148.

[25] Datasheet, MDA100CB, Data acquisition board. http://www.memsic.com/userfiles/files/Datasheets/ WSN/mts mda datasheet.pdf, (last accessed on Febuary 20, 2022)

[26] Datasheet, Temperature Sensor. http://www.ti.com/lit/ds/symlink/lm35.pdf, (last accessed on Febuary 20, 2022).

[27] Bareja, B.G., "What is Light Intensity, Effects on Plant Growth." April 2011.https://www.cropsreview.com/light-intensity.html

[28] Pijl, Anton, et al. "Remote sensing vs. field-based monitoring of agricultural terrace degradation." International Soil and Water Conservation Research 9.1, pp.1-10. 2021.

[29] Powers, L., and Snell M., "Microsoft visual studio 2008 Unleashed." Pearson Education, 2008.

[30] N. Ahmed, D. De, and I. Hussain, "Internet of things (iot) for smart precision agriculture and farming in rural areas," IEEE Internet of Things Journal, vol. 5, no. 6, pp. 48904899, 2018.

[31] M. S. Farooq, S. Riaz, A. Abid, K. Abid, and M. A. Naeem, "A survey on the role of iot in agriculture for the implementation of smart farming," IEEE Access, vol. 7, pp. 156 237156 271, 2019.

[32] I. Akyildiz, W. Su, Y. Sankarasubramanian, and E. Cayirci", A survey on sensor networks," IEEE Communications Magazine, vol. 42, no. 5, pp. 102114, August 2002.

[33] S. Karim, M.A. Rahu, A. Ahmed, A.A. Mirani, G.M. Jatoi, "Energy Harvesting for Water Quality Monitoring using Floating Sensor Networks: A Generic Framework." Sukkur IBA Journal of Emerging Technologies 1.2, 2018, pp. 19-32.

[34] F. K. Shaikh and S. Zeadally, "Energy harvesting in wireless sensor networks: A comprehensive review," Renewable and Sustainable Energy Reviews, vol. 55, pp. 10411054, 2016.

[35] S. Zeadally, F. K. Shaikh, A. Talpur, and Q. Z. Sheng, "Design architectures for energy harvesting in the internet of things," Renewable and Sustainable Energy Reviews, vol. 128, p. 109901, 2020.

[36] Sivagami, A., Pavai, K., Sridharan, D., and Murty, S.A.V., "Energy and Link Quality based Routing for Data gathering tree in wireless sensor networks under TINYOS-2. X." arXiv preprint arXiv:1005.1739, 2010.

[37] Vijayalakshmi, A., and Ranjan P.V., "Code Strategy Algorithm for Online Power Quality Monitoring of Electrical Equipment using WSN under Tiny-OS Environment." Journal of Theoretical & Applied Information Technology 57.3. 2013.

[38] White, R. M. "A sensor classification scheme". IEEE Transactions on ultrasonics, ferroelectrics, and frequency control, 34(2), 1987, pp.124-126.

[39] F. K. Shaikh, M. Memon, N. A. Mahoto, S. Zeadally, and J. Nebhen, "Artificial intelligence best practices in smart agriculture," IEEE Micro, vol. 42, no. 1, pp. 1724, 2022.

[40] F. K. Shaikh, M. Memon, and S. Zeadally, "Simultaneous wireless information and power transfer in internet of things," Energy Harvesting in Wireless Sensor Networks and Internet of Things, p. 223, 2022.

[41] B. B. Sinha and R. Dhanalakshmi, "Recent advancements and challenges of internet of things in smart agriculture: A survey," Future Generation Computer Systems, vol. 126, pp. 169184, 2022.

[42] W. Tao, L. Zhao, G. Wang, and R. Liang, "Review of the internet of things communication technologies in smart agriculture and challenges,"

Computers and Electronics in Agriculture, vol. 189, p. 106352, 2021.

[43] I. Bisio, H. Haleem, C. Garibotto, F. Lavagetto and A. Sciarrone, "Performance Evaluation and Analysis of Drone- Based Vehicle Detection Techniques From Deep Learning Perspective," IEEE Internet of Things Journal, vol. 9, no. 13, pp. 10920-10935, 2022.

[44] Bayrakdar, Muhammed Enes. "Employing sensor network based opportunistic spectrum utilization for agricultural monitoring." Sustainable Computing: Informatics and Systems 27 (2020): 100404.

[45] Mathapati, Mahantesh, et al. "A Study on Secure Data Aggregation and Routing for Wireless Sensor Networks." International Conference on Advanced Informatics for Computing Research. Springer, Singapore, 2020.

[46] Wang, L. F., and Wel, S. Q., "Cognitive engine technology".ZTE Communications, 2, 003. 2009.

[47] Habtoor, M. S., and Rowaihy, H., "Performance evaluation of IRIS and LOTUS motes when communicating in real environments". International Journal of Distributed Sensor Networks, 12(2), 7347158. 2016.

[48] Mythili, V., et al. "SEAT-DSR: Spatial and energy aware trusted dynamic distance source routing algorithm for secure data communications in wireless sensor networks." Cognitive Systems Research 58, 2019, pp. 143-155.

[49] Swami Durai, Senthil Kumar, Balaganesh Duraisamy, and J. T. Thirukrishna. "Certain investigation on healthcare monitoring for enhancing data transmission in WSN." International journal of wireless information networks, 2021, pp. 1-8.

[50] Jia, Xin-Chun. "Resource-efficient and secure distributed state estimation over wireless sensor networks: a survey." International Journal of Systems Science 52.16, 2021, pp. 3368-3389.

[51] Parrot Pot: smart pot, https://www.postscapes.com/wirelessplant-sensors/, (last accessed December 15, 2021).

[52] Open garden, https://www.cookinghacks.com/documentation/tutorials/open-gardenhydroponics-irrigation-system-sensors-plantmonitoring.html. (last accessed December 15, 2021).

[53] Apogee SQ-520, https://www.apogeeinstruments.com/sq-520-full-spectrum-smart-quantum-sensor-usb/. (last accessed December 15, 2021).

[54] Phoenix Nanotom 180NF, GE Sensing & Inspection Technologies GmbH, Wunstorf, Germany https://www.bakerhughesds.com/industrial-x-ray-ctscanners/phoenix-nanotom-m-3d-metrology-nano-ct. (last accessed December 15, 2021).

[55] Daly, Keith R., Saoirse R. Tracy, Neil MJ Crout, Stefan Mairhofer, Tony P. Pridmore, Sacha J. Mooney, and Tiina Roose. "Quantification of root water uptake in soil using Xray computed tomography and imagebased modelling." Plant, cell & environment 41, no. 1, 2018, 121-133.

[56] Tracy S.R., Daly K.R., Sturrock C.J., Crout N.M.J., Mooney S.J. and Roose T. "Three dimensional quantification of soil hydraulic properties using Xray computed tomography and image based modelling". Water Resources Research 51, 2015, pp. 10061022.

[57] Rogasik H., Crawford J.W., Wendroth O., Young I.M., Joschko M. and Ritz K. "Discrimination of soil phases by dual energy x-ray tomography." Soil Science Society of America Journal 63, 1999, pp. 741751.

[58] "Libelium Smart Agriculture Xtreme", https://www.libelium.com/iot-products/plugsense/smart-agriculture-xtreme, (last accessed on Febuary 20, 2022).

[59] "Arable mark irrigation tool", https://www.arable.com/solutions irrigation/, (last accessed on Febuary 20, 2022).

[60] "Grofit climate monitoring system," https://www.grofitag.com/product-page/grofit-iot-device-1/, (last accessed on Febuary 20, 2022).

[61] "Greeniq" https://easternpeak.com/works/iot/, (last accessed on Febuary 20, 2022).

[62] "Wheat Seed Production Techniques Manual", District Agriculture Development Office, Sindhupalchwok, https://www.jica.go.jp/nepal/english/index.html, (last accessed on September 02, 2022)

[63] "Meteohelix iot pro", https://allmeteo.com/agricultureiot-weather-station, (last accessed on Febuary 20, 2022).

[64] "Sky-lora weather station", https://get.pycno.co/collections/allproducts/products/sky lora-and-wifi-weather-station. (last accessed on Febuary 20, 2022).

[65] "Ec-1 controller", https://growlink.com/shop/ec-1/. (last accessed on Febuary 20, 2022).

[66] I. Bisio, C. Garibotto, H. Haleem, F. Lavagetto, and A. Sciarrone. "On the localization of wireless targets: A drone surveillance perspective." IEEE Network 35, no. 5, pp. 249-255, 2021.

[67] I. Bisio, C. Garibotto, H. Haleem, and Lavagetto, "Drone surveillance systemRF/WiFi-based drone detection, localization, and tracking: a surve," Aviation Cybersecurity: Foundations, Principles, and Applications, 49, 2021.

Vol. 5, No. 2 | July - December 2022

# Simulation based Analysis of Single Unit and Parallel Connected Three Phase AC Generators in QUEST Campus Larkana

Abdul Hameed Soomro[1]*, Syed Abid Ali Shah[1], AssadUllah Khauhwar[1], SanaUllah Talani[1], Aijaz Ahmed Solangi[1],Talha Soomro[1], Faraz Ali Khushk[1]

**Abstract:**
*Recently power crises are increasing day by day in Pakistan due to a shortage of energy resources. The financial position of Pakistan is too weak and is not in a position to install new power projects for the production of electrical power to meet the demand as per requirement. Every consumer is trying to install the power-producing devices to meet the demand in order to run their institutional laboratories, offices, commercial plants, and industrial machines. Electricity for resident people of villages of Pakistan is not sufficient to run their home appliances and mostly solar systems has installed to meet the requirement in day time but in night time energy storage system is required, so cost will be increased and less illumination is received at night time. The generator is also one of the power producer devices to meet the consumers' demand to supply the electrical power to the load in absence of failure of power supply from utility companies. A single large unit is very expensive to run the small loads and consumes large amounts of fuel, so the parallel connection of two small generators is very beneficial to meet the demand as per requirement, resulting in cost savings, and less fuel consumption. QUEST Campus Larkana has installed two 50KVA generators in parallel to supply the required power to the load in absence of a power supply to run their offices, workshops and laboratories, this minimizes the cost, and increases the reliability of the system. The MATLAB simulation model is developed to analyze the performance of the single unit generator and parallel connected generators.*

**Keywords:** *AC Generators, Power Supply*

## 1. Introduction

Recently power crises are increasing day by day in Pakistan due to a shortage of energy resources and poor financial position to install new power projects for the production of electrical power to meet the demand as per requirement [1]. Educational institutions, hospitals, laboratories, industries, residential, and commercial consumers are facing power crises and very difficult to manage the workstations in absence of electrical power [2]. As shown in figure 01 the power demand in Pakistan. Every consumer of electricity is trying to maintain his power system in boundary wall through addition of standby system to provide facilities to students and officers in educational institutes, patients and their staff in hospitals, and increase in revenue of production industries [3, 4]. Every country wants to run their industrial sector properly for development but due to the power quality problems and shortage of power, the required output is badly affected [5].

[1]Department of Electrical Engineering, Quaid-e-Awam University of Engineering Science and Technology Nawabshah Campus Larkana, Sindh Pakistan
Corresponding Author: abdul.hameed@quest.edu.pk

Power quality can be improved by installing the devices [6, 7] but the shortage of power is a serious problem for proper manufacturing in the industrial sector [1]. Electricity for resident people of villages of Pakistan is not sufficient to run their home appliances and mostly solar systems have been installed to meet the requirement during day time but at night time energy storage system is required, so the cost will be increased and less illumination is received at night time [4, 8]. The shortfall of electrical power is increasing day by day in Pakistan, cities are facing 6 to 8 hours of daily load shedding, villages are more than 12 hours in a day, and some of the villages are without electricity [2]. Hospitals, medical laboratories, institutions, industries, and large commercial offices have large loads and they have also required stand-by systems to run their loads at day and night time [9]. Electrical generator is one of the standby systems to meet the consumers' demand [10] which convert mechanical energy into electrical energy and available in small and large ratings to meet the consumers' requirement day and night.

Quaid-e-Awam University of Engineering and Technology campus Larkano was established in 2010 which is a constituent campus of the QUEST Nawabshah. There are four Engineering departments and initially, one batch was admitted in 2010 and installed 50KVA generator to meet the power demand in case of failure. After increasing the laboratories, staff offices, and induction of more student batches one more 50KVA generator was added in parallel to meet the demand. In case of low power requirement, only single unit is running and the other is remained off, this will give saving in consumption of fuel [11], so two small rating generators are beneficial instead of a single unit of large capacity which consumes large amount of fuel, and large in size. MATLAB simulation software is used to develop the model of a parallel connected generator to analyze the performance under small and large loads. Different power companies are intensively taking part in investments to enhance wind generation and overcome the shortage of electrical power in Pakistan [12], [13]. The utilization of oil and coal since 2015 has increased so we cannot more rely on oil and coal due to costly, which will directly hit the economy of the country [14]. Natural gas is available in Pakistan but the Oil and Gas Company of Pakistan (OGDCL) has declared that the natural gas will be reduced from 2025 to 2030 so we cannot rely on natural gas for the installation of more power plants in Pakistan in order to meet the power demand [15]. Figure 1 shows the supply-demand gap in Pakistan from 2002-2030.

Section 2 AC Generator

Section 3 Parallel Connection of generators

Section 4 Parallel connected generators in QUEST Campus Larkana

Section 5 Simulation Results and discussions

Section 6 Conclusions

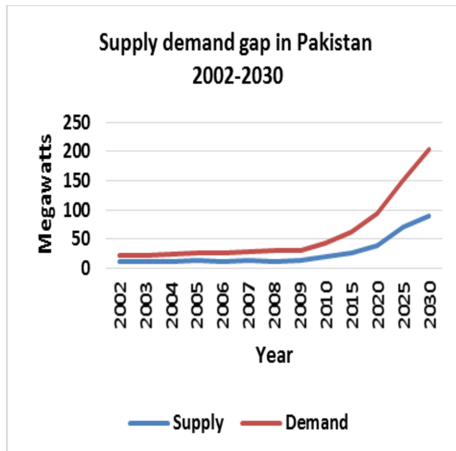



**Fig. 1.** Supply Demand Gap in Pakistan 2002-2030

## 2. AC Generator

The Generator is a device that converts mechanical energy into electrical energy and works on Faradays law of electromagnetic induction [16]. It consists of a coil of wire which rotates in a magnetic field. According to faradays law, the emf is induced when the current-carrying conductor is placed in a magnetic field [17], the simple AC generator is shown in figure 2 [18]. The direction of induced emf is calculated through Flemings right-hand rule, the process is repeated for each cycle [19] as shown in figure 3 [20]. It has been seen that at point A there is 00; no emf is generated because coil movement and direction of a magnetic field are in parallel with each other. At point B there is 900; maximum emf is generated because coil movement is at 900 to a magnetic field. At point C there is 1800; no emf is generated because coil movement and direction of a magnetic field are in parallel with each other. At point D there is 2700; maximum emf is generated because the coil movement is at 2700 to the magnetic field but opposite to point B. Finally, at point A there is 3600; no emf is generated because coil movement and direction of a magnetic field are in parallel with each other and the coil completed the rotation [3].

Assume that the coil shape is rectangular and has N no of turns with angular velocity ω and rotates in uniform magnetic field B and coil at any time is t then the angle between a magnetic field and coil time is given in equation (01) and magnetic field perpendicular to the plane of coil is given in equation (02) .

$$\theta = \omega t (01)$$

$$\varphi = B\cos\omega t (02)$$

When a magnetic field is linked with coil of $N$ turns then

$$\varphi = B\cos\omega t A (03)$$

Where A is area of coil, induced emf in the coil according to faradays law

$$\varepsilon = -d\emptyset/dt$$
$$\varepsilon = -d(NBA\ \cos\omega t/dt$$
$$\varepsilon = NBA\omega \mid \sin\omega t (04)$$

When coil is rotate with 900 then the value of $sin$ is 1 then equation (04) becomes

$$\varepsilon_0 = NB_m A\omega = NB_m A\ 2\pi (05)$$

Where $B_m$ denotes the maximum flux density and A is the cross sectional area and f is the frequency of coil rotation, now add equation (05) in equation (04) then we get

$$\varepsilon = \varepsilon_0 sin\omega (06)$$

The induce current is given in equation (07)

$$I = {}^{\varepsilon_0}/_{R} = \varepsilon_0 Sin\omega (07)$$

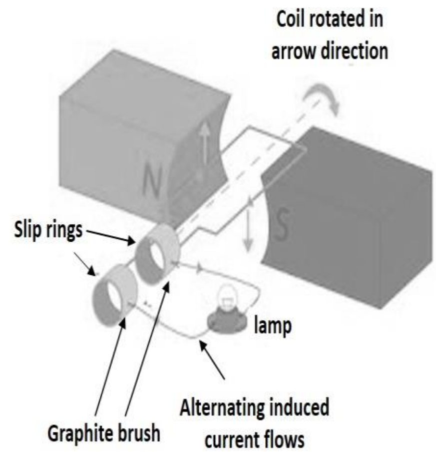



**Fig. 2.** Simple AC Generator

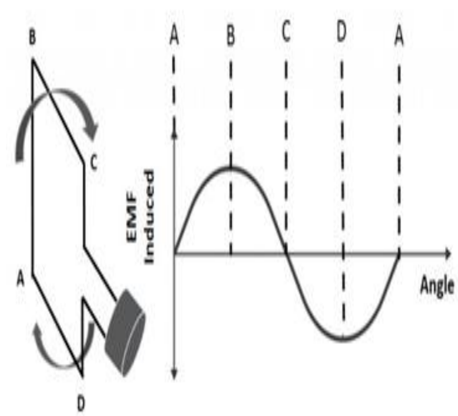



**Fig. 3.** Different Positions of a Coil

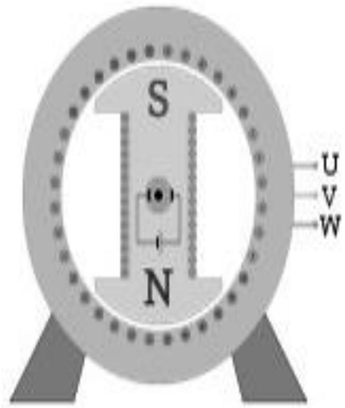



**Fig 4 (a)**

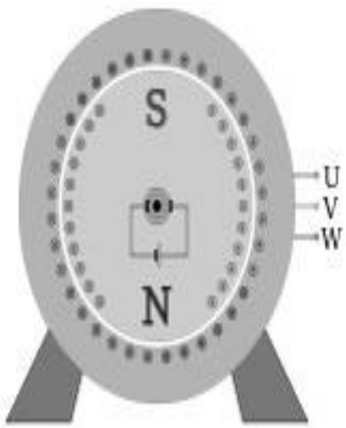



**Fig 4(b)**

**Fig. 4.** (a) Salient Pole Rotor, (b) Non-salient Pole Rotor

AC generator is a very essential device that meets the requirement of electrical power in the shape of generation. In AC generators the slip rings are employed to produce the alternating current and are available in small and large capacities and mostly applied to the power plant for the generation of electrical power and in commercial and industrial sectors to meet the energy gap [21]. The main parts of the generator are the stator and rotor. The stator is a stationary part fixed an armature winding that carries the current for the load, it consists of a frame, stator core, and armature winding [22]. The rotor is the rotating part of the generator, it consists of magnetic field winding and a dc supply is used to magnetize the magnetic poles. Salient and non-salient type rotors as shown in figures 4(a) and 4(b) are used depending on the point of application [23].

## 3. Parallel Connections of AC Generators

When a load is increased beyond the capacity of a single unit of AC generator, another generator is connected in parallel to meet the required demand [11]. The single unit of large rating is heavy, large in size, and more costly to replace a single unit of small capacity with a large one and in case of fault the whole unit will be replaced with another one for maintenance, resulting in revenue loss, large labor charges and also disturb the reliability of power to run the offices, laboratories, and industries [24]. It is a dire need to connect a small rating of two AC generators in parallel to meet the required demand. If one unit of generator is faulty then we can run our major machines, laboratories, and office equipment till the availability of another unit safely [25]. Two small units of parallel connected generators give result in small size, less weight, easy maintenance, and reduced revenue loss [26]. The parallel connection is possible when the terminal voltage of the two generators must be the same otherwise high current will flow between the two machines, resulting in damaging the equipment or generator [27]. The frequency of both generators must be the same otherwise the low-frequency generator takes more loads and results in the overloading of the generator [28]. The output voltage must be the same in phase otherwise large output voltage is developed in phase [29].

## 4. Parallel Connected AC Generators in QUEST Campus Larkana

QUEST Campus Larkana was established in 2010. There are four Engineering departments and initially, one batch was admitted in 2010 and installed a 50KVA generator to meet the power demand in case of failure. After increasing the laboratories, staff offices, and induction of more student batches one more 50KVA generator was added in parallel to meet the demand and run the load of 400Amperes.


**Fig. 5.**Parallel Connected Generators at QUEST Campus Larkana

In case of low power requirement, only a single unit is running, and the other is remained off, this will give saving in consumption of fuel [11], so two small rating generators are beneficial instead of a single unit of large capacity which consumes a large amount of fuel, and large in size. Figures 5 and 6 show the AC generators available at campus Larkana.

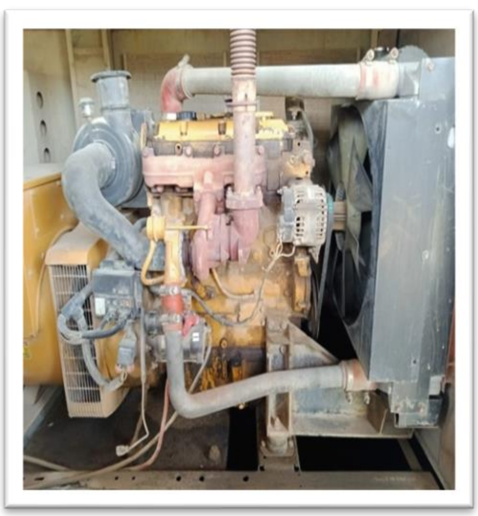


**Fig. 6.**Inner look of 50KVA AC Generator

## 5. Simulation Results and Discussion

The simulation model is developed in MATLAB software and analyzes the performance of the AC generator when operating alone and in parallel under normal and overload conditions. Figure 7 shows the Simulation model of parallel connected generators and Table 1 shows the parameters of generators.

TABLE I. PARAMETRS OF AC GENERATORS

| S.No | Generator Parameters | Value |
|---|---|---|
| 1 | Capacity of AC Generator | 50 KVA |
| 2 | Output voltage per phase | 220 volts |
| 3 | Frequency | 50 Hz |
| 4 | Inertia | $3.7e^3$ |
| 5 | Pair of poles | 2 Nos |
| 6 | Internal Resistance of AC generator | 0.0001 Ohm |
| 7 | Internal Inductance of AC Generator | 0.05 Henry |
| 8 | DC Machine Field type | Wound |
| 9 | Armature Resistance of DC machine | 24 Ohm |
| 10 | Field Resistance of DC machine | 618 Ohm |
| 11 | Field Inductance of DC machine | 0.05 Henry |
| 12 | Field Armature mutual inductance of DC machine | 1.8 Henry |
| 13 | Initial Field Current of DC machine | 0.3 Ampere |

| | | |
|---|---|---|
| 14 | DC voltage source | 169 Volts |
| 14 | Mechanical input | Speed |
| 15 | Sampling Time | -1 |

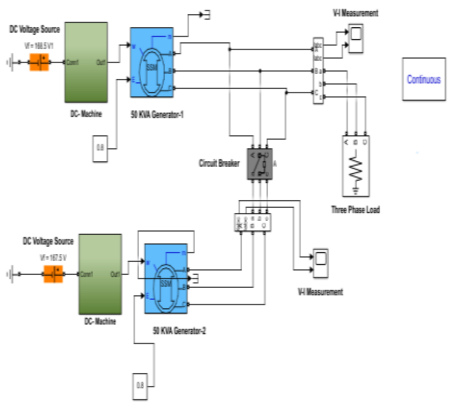



**Fig. 7.** Simulation Model of Parallel Connected Two 50KVA AC Generator

### 5.1. Single Unit Generator Opearting alone (At 150 Ampere Load)

In the initial stage, a load of QUEST campus Larkana was 195 Amperes and one single unit was sufficient to meet the supply gap. It has been seen in this simulation result from the single unit of generator produces the 220Volts single phase without any oscillations and meets the requirement satisfactorily. Figure 8 shows the simulation result of the generator under normal load conditions.

### 5.2. Single Unit Generator Opearting alone under variation of input and reference parameters (At 210 Ampere Load)

When the load of the campus was increased beyond the capacity of a single generator, so it has seen in this simulation result that the output voltage of a single unit of the generator is reduced from 220Volts to 170Volts and not meets the requirement satisfactorily and after some time heat will be generated and shut down. Figure 9 shows the simulation result of the generator under variation of input and reference parameters.

### 5.3. Single Unit Generator Opearting alone under variation of input and reference parameters (At 230 Ampere Load)

When the load of the campus was increased beyond the capacity of a single generator i.e. up to 230Amperes, so it has seen in this simulation result that the output voltage of a single unit of the generator is reduced from 220Volts to 135Volts and not meet the requirement satisfactorily and after some time heat will be generated and shut down. Figure 10 shows the simulation result of the generator under variation of input and reference parameters.

### 5.4. Two Generator Opearting Parallel (At 320 Ampere Load)

When the load was increased and to run the offices, labs, classes, and other loads satisfactorily, another generator of 50KVA was added to the first generator in parallel. It has been seen in this simulation result that the parallel operation of two generators share the load and produces the 220Volts single phase and meets the requirement satisfactorily. Figure 11 shows the simulation result of the parallel operation of two generators.

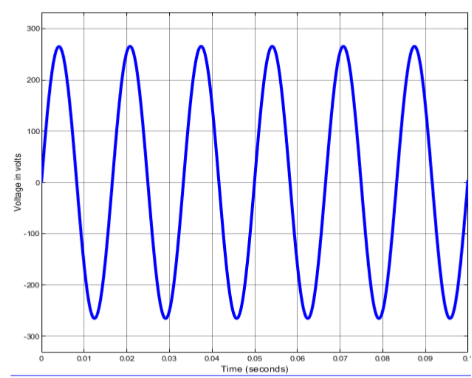



**Fig. 8.** Single Unit Operation of Generator Under Normal Input and Reference Parameters

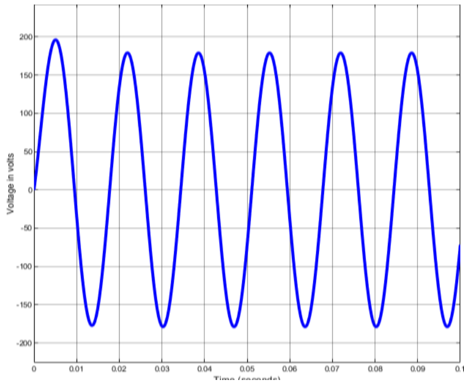



**Fig. 9.** Single Unit Operation of Generator Under Variation of Input and Reference Parameters

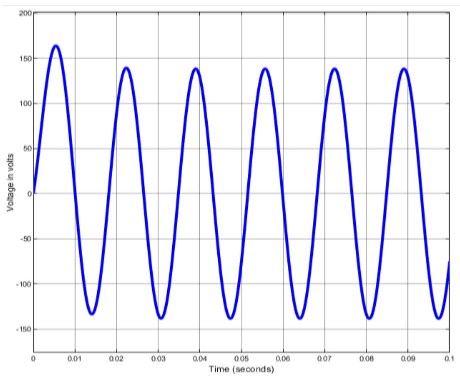



**Fig. 10.** Single Unit Operation of Generator Under Normal Input and Reference Parameters

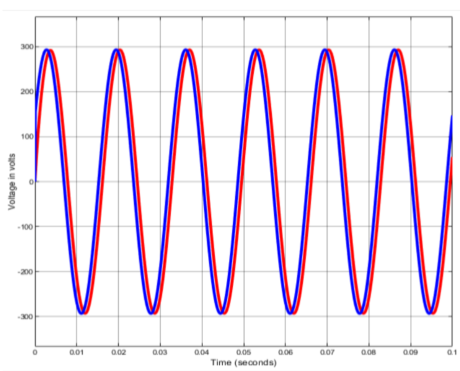



**Fig. 11.** Parallel Operation of Two AC Generators

## 5.5. Discussion

It has been seen from the literature that the power crises are increasing day by day in Pakistan due to that our industrial sectors, public sector, and commercial sectors are facing very difficulty in keeping the reliable operation of their equipment. The Solar system is more costly and difficult to get energy at the night time. So an AC generator is the best option to meet the energy gap in day and night time. It has been seen from simulation results that the performance of a single-unit generator is not satisfactory under overload conditions and consumers received less voltage for utilization and the generator overheated and shut down. To eliminate this problem another generator of the same rating is connected in parallel to get the satisfactory performance of generators under overload conditions. Table 2 shows the simulation results of the operation of AC generators operating alone and in parallel.

TABLE II. SIMULATION RESULTS OF AC GENERATORS OPERATING IN PARALLEL AND ALONE

| **Load** | **Single Unit Generator** | **Parallel Connected Generators** |
|---|---|---|
| 150 Ampere | 220 Volts per phase | Operation not started |
| 210 Ampere | 170 Volts per phase | Operation not started |
| 230 Ampere | 135 Volts per phase | Operation not started |
| 320 Ampere | ---- | 220 Volts per phase |

## *6.* Conclusion

This research study is carried out in the Quaid-e-Awam University of Engineering and Technology Nawabshah Campus Larkana to analyze the performance of the single unit generator and parallel connection of two generators to meet the energy demand in case of failure of the power supply. In Pakistan, the energy demand is increasing day by day, and currently facing a shortage of energy due to that 6-8 hours load shedding is carried out in cities and 12-14 hours in villages. Hospitals, Laboratories, Industries, Commercial loads, and Universities are very responsive places so the availability of electrical power is essential to carry out the work properly. For the continuity of power supply stand-by system is necessary in shape of AC generators because the solar system will be affected in the rainy season, cloudy conditions, and nighttime. The two generators of 50KVA are available in QUEST Campus Larkana to meet the required energy gap. It has been seen from the simulation results that when a load is increased beyond the capacity of a single unit of the generator then the output voltage is decreased and the generator is overheated after some time generator is shut down, so it is necessary to add another generator in parallel with a single one to share the load. The output voltage remains within the limit and the problem of overheating and shutdown is eliminated. Singe large unit of generator is costly and heavy and also a problem in maintenance, so it is concluded that the two generators of small capacity are installed, resulting in easy maintenance, reduce cost, and small in size to meet the energy demand in Industries, Hospitals, Universities, and laboratories.

## Acknowledgment

The authors would like to thank QUEST, Nawabshah Pakistan and MUET, Jamshoro Pakistan for the research facility provided.

## Funding

This research is without any external funding

### References


[1] M. A. Raza, K. L. Khatri, A. Israr, M. I. U. Haque, M. Ahmed, K. Rafique, et al., "Energy demand and production forecasting in Pakistan," Energy Strategy Reviews, vol. 39, p. 100788, 2022.

[2] F. Arshad and M. Shamshad, "Energy Crisis in Pakistan: Socio-Economic Implications and the Way Forward," Annals of Social Sciences and Perspective, vol. 3, pp. 105-115, 2022.

[3] H. Radmanesh, H. Jashnani, A. Khaledian, and H. Sobhani, "Optimal and stable electric power system for more electric aircraft: Parallel operation of generators and weight reduction," Journal of Energy Management and Technology, vol. 5, pp. 23-31, 2021.

[4] G. D. Valasai, M. A. Uqaili, H. R. Memon, S. R. Samoo, N. H. Mirjat, and K. Harijan, "Overcoming electricity crisis in Pakistan: A review of sustainable electricity options," Renewable and Sustainable Energy Reviews, vol. 72, pp. 734-745, 2017.

[5] B. Singh, A. Chandra, and K. Al-Haddad, Power quality: problems and mitigation techniques: John Wiley & Sons, 2014.

[6] A. Farooqi, M. M. Othman, M. A. M. Radzi, I. Musirin, S. Z. M. Noor, and I. Z. Abidin, "Dynamic voltage restorer (DVR) enhancement in power quality mitigation with an adverse impact of unsymmetrical faults," Energy Reports, vol. 8, pp. 871-882, 2022.

[7] R. Adware and V. Chandrakar, "Comprehensive Analysis of STATCOM with SVC for Power Quality Improvement in Multi Machine Power System," in 2022 2nd International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), 2022, pp. 1-5.

[8] V. Khare, S. Nema, and P. Baredar, "Solar–wind hybrid renewable energy system: A review," Renewable and Sustainable Energy Reviews, vol. 58, pp. 23-33, 2016.

[9] A. Chauhan and P. Thakur, Power Quality Issues and Their Impact on the Performance of Industrial Machines: Anchor Academic Publishing, 2016.

[10] I. Boldea, "Electric generators and motors: An overview," CES Transactions on Electrical Machines and Systems, vol. 1, pp. 3-14, 2017.

[11] M. Stone, "Parallel operation of AC generators—Action of governors and Damper windings," Electrical Engineering, vol. 52, pp. 44-44, 1933.

[12] G. J. Herbert, S. Iniyan, E. Sreevalsan, and S. Rajapandian, "A review of wind energy technologies," Renewable and sustainable energy Reviews, vol. 11, pp. 1117-1145, 2007.

[13] Z. Yao, Z. Yuxing, K. Yaqian, and B. Sobhani, "Research on an integrated power and freshwater generation system from natural gas energy and geothermal sources," Desalination, vol. 525, p. 115494, 2022.

[14] M. Asim, S. Jamil, R. Shad, N. Hayat, A. Moaz, M. T. Akram, et al., "Comparison of Reanalysis, Analysis and Forecast datasets with measured wind data for a Wind Power Project in Jhimpir, Pakistan," in Journal of Physics: Conference Series, 2018, p. 012004.

[15] M. S. Sarfraz, M. Asim, M. S. Kamran, S. Imran, and N. Hayat, "Evaluation of ERA-interim and NCEP-CFSR reanalysis datasets against in-situ measured wind speed data for keti bandar port, Pakistan," in Journal of Physics: Conference Series, 2018, p. 012001.

[16] J. Darke, "The primary generator and the design process," Design studies, vol. 1, pp. 36-44, 1979.

[17] T. Chan, L. Lai, and L.-T. Yan, "Performance of a three-phase AC generator with inset NdFeB permanent-magnet rotor," IEEE transactions on energy conversion, vol. 19, pp. 88-94, 2004.

[18] M. D. Lee, P. San Lee, R. C. H. Chiong, P. Karunakaran, and H. J. Ngu, "Design and Development of a Small-Scale Mechanical Energy Conversion Device," in 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-5.

[19] S. Singirikonda, G. Sathishgoud, and M. Harikareddy, "Transient stability of AC generator controlled by using fuzzy logic controller," Int J Eng Res Appl, vol. 4, 2014.

[20] I. Boldea, The Electric Generators Handbook-2 Volume Set: CRC press, 2005.

[21] H. Polinder, J. A. Ferreira, B. B. Jensen, A. B. Abrahamsen, K. Atallah, and R. A. McMahon, "Trends in wind turbine generator systems," IEEE Journal of emerging and selected topics in power electronics, vol. 1, pp. 174-185, 2013.

[22] C. Williams and R. B. Yates, "Analysis of a micro-electric generator for microsystems," sensors and actuators A: Physical, vol. 52, pp. 8-11, 1996.

[23] W. Freitas, A. Morelato, W. Xu, and F. Sato, "Impacts of AC generators and DSTATCOM devices on the dynamic performance of distribution systems," IEEE transactions on power delivery, vol. 20, pp. 1493-1501, 2005.

[24] L. Larson, "Parallel operation of aircraft AC generators," Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry, vol. 72, pp. 403-407, 1954.

[25] L. Dreller, "THE FUNDAMENTALS OF PARALLEL OPERATION OF DIRECT AND ALTERNATING CURRENT GENERATORS," Journal of the American Society of Naval Engineers, vol. 49, pp. 273-306, 1937.

[26] V. Popov, "AC Generator Paralleling Method," 2020.

[27] O. Honorati, F. Caricchi, and E. Santini, "High speed AC generator (7200 RPM, 60 Hz) for autonomous power systems," IEEE transactions on energy conversion, vol. 4, pp. 544-550, 1989.

[28] A. Arkadan, Y. Abou-Samra, and N. Al-Aawar, "Characterization of stand alone AC generators during no-break power transfer using radial basis networks," IEEE transactions on magnetics, vol. 43, pp. 1821-1824, 2007.

[29] X. Liu, P. C. Loh, F. Blaabjerg, and P. Wang, "Load sharing using droop control for parallel operation of matrix converters as distributed generator interfaces in isolated mode," in 2012 IEEE Energy Conversion Congress and Exposition (ECCE), 2012, pp. 962-968.

Vol. 5, No. 2 | July - December 2022

# An Optimized Proportional Integral Derivative (PID) based Power System Stabilizer (PSS) for Damping of Active Power Oscillations

Veer Bhan[1]*, Ashfaque Ahmed Hashmani[1], Abdul Hakeem Memon[1]

**Abstract:**
*Rising load demand and possibilities of short interruptions either due to temporary faults or equipment switching may lead to transient instability in terms of growing oscillations and eventually result in cascaded outages if not being damped properly and timely. To dampen such oscillations, the power system stabilizers (PSS) are invariably installed to regulate the excitation of synchronous generators. In this research, a comparative analysis among the two design variants of PSS i.e., Lead Lag and PID, is performed by comparing the overshoot and settling time of generator active power after disturbance, to observe the effect of each on the system characteristics and nominal conditions. The settling time with PID based PSS is reduced by 46% in case of three phase short circuit fault and 80% in case of the single line to ground fault as compared to Lead Lag PSS. Thus, the results obtained show a better performance of PID based PSS having better overshoot response and reduced settling time for symmetrical and unsymmetrical faults.*

**Keywords:** *Active power oscillations, Lead-Lag PSS, PID-based PSS*

## 1. Introduction

Notwithstanding the improved security and tolerance of modern power systems towards temporary and permanent contingencies, the issues of stability and protection have been prevailing yet [1][2]. The stability of a power system refers to the condition where it maintains the state of equilibrium for normal operation and retains its acceptable characteristics after being subjected to local or global disturbances. The common disturbances include faults and load switching that introduce low-frequency oscillations into the power system and may lead to oscillatory instability if not being damped out adequately. One of the promising ways to compensate for such oscillations is to install Power System Stabilizer (PSS) at the excitation of the synchronous generator, as shown in Fig. 1.

The PSS aims at expanding the system stability limit with the help of modulation of generator excitation So that it can dampen the power swing modes with additional positive damping torque [3].

The modeling and design of different types of PSS have been well-studied in the literature. Using the static output feedback (SOF) control approach, the researchers in [4] have proposed the direct implementation of PSS having PID properties and compared it with various existing passive architectures. The idea is, however, simple and feasible to implement, but it does not ensure global stability. A novel PSS structure to ensure the global transient stability of a Microgrid system is discussed in [5]. The method has been evaluated against the

[1]Department of Electrical Engineering Mehran University of Engineering and Technology Jamshoro, Sindh, Pakistan
Corresponding Author: veerhablani@gmail.com

loading and faulty conditions considering the small-signal and transient stability criterion.

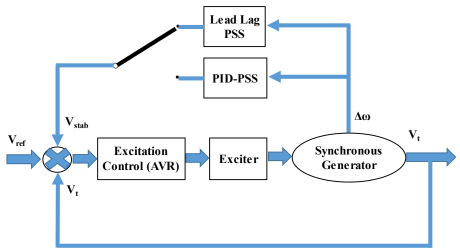



**Fig. 1.** Conceptual representation of installation of PSS into the power system

In [6], the researcher has exploited the concept of model predictive control (MPC) to design a robust architecture of PSS for improving the dynamic stability of a single-machine infinite bus (SMIB) system. This system, owing to its predictive nature, can adapt to unknown conditions but is computationally ill condition and has a marginally slow response time.

Considering above discussed approaches, the overall objectives of this paper are:

1. To formulate and model the two designs of PSS, i.e., Lead- Lag and PID.
2. To investigate the effects of installing these types of PSS onto the characteristics of the power system in the simulated environment of MATLAB Simulink.

To evaluate and compare the performance of these PSS designs in terms of overshoot and settling time.

## 2. Literature Review

Various researchers have analyzed the performance of PSS with different computational methodologies [7]. Authors in [8] worked on the analysis of steady-state stability and improved it using eigenvalues and conventional PSS for a thermal power plant in Jamshoro, Pakistan, taken as a case study. Most of the power plants are equipped with a conventional power system stabilizer (CPSS), which is of the Lead-Lag type and has fixed parameters [9]–[12]. However, Lead-Lag based conventional PSS lacks reliability against different loading conditions. [4]. Industrial utilities have used proportional-integral-derivative (PID) controllers for their essential functionality and simplicity in structure. To function properly for a wide range of operating conditions, PID-based PSS is proposed in [13]. Also, PID controller-based PSS structure has been presented for system stability enhancement by applying the controller on a generator connected with the infinite bus, and its performance has been compared without PSS, with proportional-integral (PI-PSS), and with proportional-integral-derivative (PID-PSS).

In [14], the researchers investigated the performance of a large hydro station using PID-based PSS against the damping of active power oscillations. The gains of the installed PSS model were chosen considering the frequency characteristics of the system under different loading and faulty conditions. Likewise, the case study in [15] examined the PID-based PSS model for a multi-machine system, but the gains were determined using particle swarm optimization considering the dynamic stability characteristics of the power system. The results of three different loading conditions from eigenvalue analysis show that the PID-based PSS has improved steady state and dynamic performance compared with the Lead-Lag PSS [16]. While tuning for stabilizer parameters has been presented in [17]–[20].

## 3. Research Methodology

To design a robust PSS model that is not only resilient to external disturbances, thereby providing positive damping to local mode oscillations but also ensures optimal behavior by injecting torque variations in phase with speed deviations by considering the loading characteristics of the system. To evaluate the performance and effectiveness of the designed PSS model, its dynamic damping test is carried out to validate that the resultant model can dampen active power oscillations effectively. For this purpose, two different models of PSS, i.e., Lead-Lag and PID, are designed and analyzed in this paper.

## 3.1. Lead-Lag PSS Model

The block diagram representation of the Lead-Lag PSS model is shown in Fig. 2. Where the stabilizing gain $K_{stab}$ determines the amount of damping provided and plays an essential role in the damping of active power oscillations. By increasing the gain value, damping increases up to a certain point, beyond which further increase in gain results in a decrease in damping. The wash-out filter is of a high-pass type used to remove the low-frequency components or to prevent steady changes in speed from modifying the field voltage and enables the stabilizer to act upon the speed deviations only. The phase compensation is used to eliminate the phase lag between the exciter input (i.e., PSS output) and the resulting electrical torque of the synchronous generator [3]. The input to this model can be either the speed deviation ($\Delta\omega$) or the acceleration power ($P_a$) i.e., the difference of mechanical power ($P_m$) and electrical power ($P_e$), as expressed by the equation shown in (1).

$$P_a = P_m - P_e \quad (1)$$

Whereas the output is the stabilizing voltage ($V_{stab}$) with upper and lower bounds to restrict the terminal voltage of the synchronous generator within predefined limits. The overall transfer function of this PSS model is expressed by the equation shown in (2).

$$\frac{V_{stab}}{\Delta\omega} = K_{stab} \frac{sT_W + s^2 T_W T_1}{1 + s(T_2 + T_W) + s^2 T_W T_2} \quad (2)$$

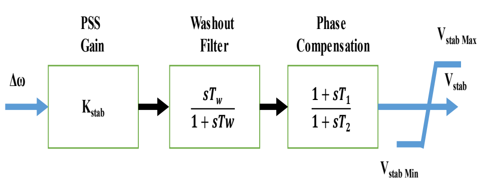



**Fig. 2.** Block diagram of Lead-Lag PSS Model

The choice of Tw, T1, and T2 depends upon the characteristics of the power system and the rating and configuration of the synchronous machine. The sufficient condition to determine their values states that the poles of the equation shown in (2) should be real and lie on the negative complex plane. The parameters given in equation (2) are enumerated in Appendix A.

## 3.2. PID-based PSS Model

The Lead-Lag architecture, however, is simple and easy to realize but does not guarantee global stability and robustness against varying operating conditions under all disturbances. The standard PID controller with unity feedback is exploited to achieve state stabilization of rotor swings, as shown in Fig. 3.

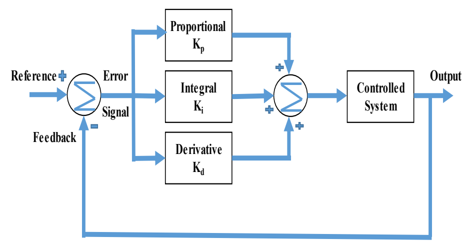



**Fig. 3.** Schematic representation of PID-based PSS Model

This design works on an error signal originating either in speed ($\omega$) or acceleration power ($P_a$), as mathematically described by the equation shown in (3).

$$V_{stab} = K_P(\omega_d - \omega_c) + K_I \int(\omega_d - \omega_c)dt + K_D(\dot{\omega}_d - \dot{\omega}_c) \quad (3)$$

Where $K_P$, $K_I$, and $K_D$ are the controller gains, $\omega_d$ and $\omega_c$ represent desired and current rotor speed, respectively, and their associated acceleration are denoted by $\dot{\omega}_d$ and $\dot{\omega}_c$. The output of the controller is a voltage signal ($V_{stab}$), like the Lead-Lag PSS, given to the excitation system, which as a result, produces the additional torque responsible for the damping of active power oscillations.

For this PSS model, asymptotic stability is achieved by the proper selection of gain matrices. The optimal choice of these gain matrices is made in such a way that equation shown in (4) is satisfied [21].

$$\ddot{e} + K_D\dot{e} + K_I \int e dt + K_P e = 0 \quad (4)$$

## 4. Results and Discussion

The presented models of PSS have been evaluated onto the Simulink model of a thermal power plant connected with an infinite bus, i.e., single machine infinite bus (SMIB) under symmetrical and unsymmetrical fault. The two-evaluation metrics, i.e., overshoot, i.e., % of the rise in the value of respective quantity over to its nominal value and settling time, i.e., the duration within which the system has retained its normal state. The generator active power for the presented PSS models against a three-phase short circuit (symmetrical) fault at constant loading conditions is shown in Fig. 4 to Fig. 7.

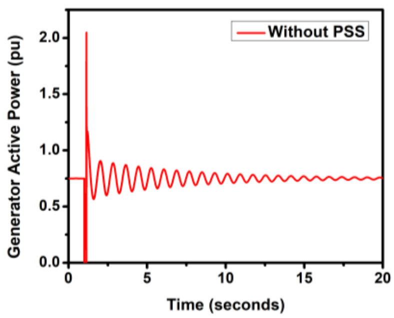



**Fig. 4.** Generator Active Power against three-phase short circuit fault without PSS

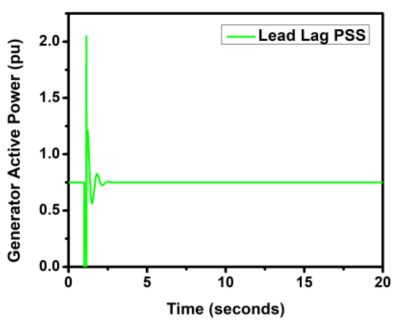



**Fig. 5.**Generator Active Power against three-phase short circuit fault with Lead-Lag PSS

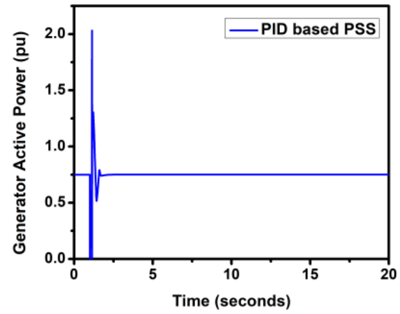



**Fig. 6.** Generator Active Power against three-phase short circuit fault with PID-based PSS

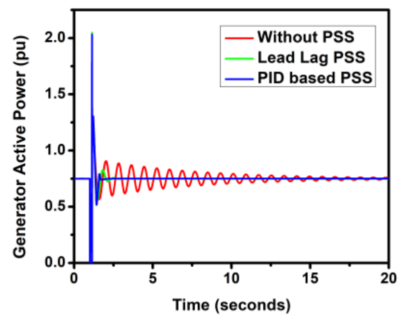



**Fig. 7.** Generator Active Power against three-phase short circuit fault with different PSS

TABLE I. SUMMARY OF OVERSHOOT AND SETTLING TIME AGAINST THREE-PHASE SHORT CIRCUIT FAULT

| **Model** | **Three Phase Short Circuit Fault** | |
|---|---|---|
| | Generator Active Power | |
| | Overshoot (pu) | Settling Time (s) |
| **Without PSS** | 0.41 | >10 |
| **Lead-Lag PSS** | 0.45 | 1.3 |
| **Optimized-PID PSS** | 0.35 | 0.7 |

The fault occurs at 1 sec, as can be seen from Tables I and II, that owing to the severity of three-phase short circuit fault, large and high frequency deviations and transients are observed into the output power of the generator as compared to a single line to ground fault. Further, PID-based PSS is robust and effective in mitigating the local area oscillations with reduced overshoot and settling time.

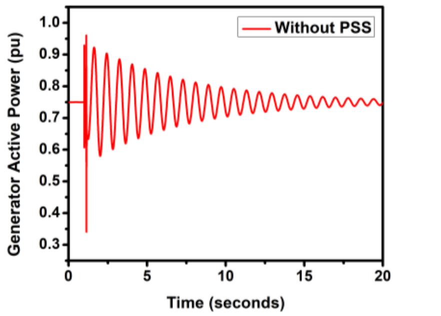



**Fig. 8.** Generator Active Power against the single line to ground fault without PSS

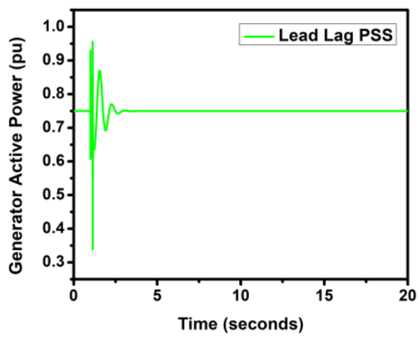



**Fig. 9.** Generator Active Power against the single line to ground fault with Lead-Lag PSS

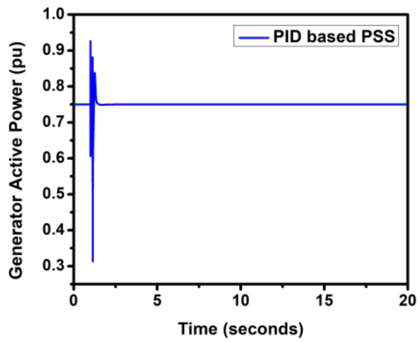



**Fig. 10.** Generator Active Power against the single line to ground fault with PID-based PSS

In comparison, the Lead-Lag PSS performs better than without PSS in both scenarios, with reduced overshoot and settling time. The generator active power for the presented PSS models against the single line to ground (unsymmetrical) fault at constant loading conditions is shown in Fig. 8 to 11.

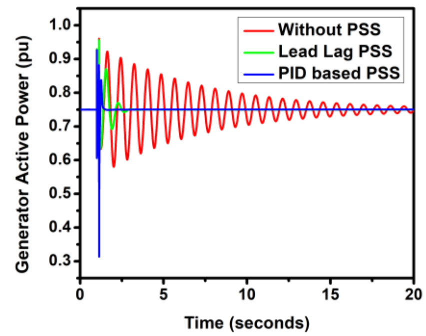



**Fig. 11.** Generator Active Power against the single line to ground fault with different PSS

TABLE II. SUMMARY OF OVERSHOOT AND SETTLING TIME AGAINST SINGLE LINE TO GROUND FAULT

| Model | **Single Line to Ground Fault** Generator Active Power | |
|---|---|---|
| | Overshoot (pu) | Settling Time (s) |
| **Without PSS** | 0.17 | >10 |
| **Lead-Lag PSS** | 0.12 | 1.3 |
| **Optimized-PID PSS** | 0.08 | 0.26 |

It is evident from Table I and II that the PID-based PSS shows lesser overshoot as well as settling time compared to Lead-Lag and without PSS. On the other hand, Lead-Lag PSS performs better than without PSS, but unfortunately, it under-performs in achieving less overshoot and settling time than PID-based PSS.

## *5.* Conclusion

This research analyzed and discussed the characteristics and effects of integrating the different designs of PSS into the performance of the overall power system. The power system under investigation was the thermal power plant in the MATLAB Simulink environment, and the models of respective PSS were designed by considering the rating, loading, structural properties of the power system, and as well as severity of the simulated contingencies. It has been observed that the PID-based PSS adapts well to varying conditions irrespective of the nature and severity of the conditions and has reduced overshoot and reduced settling time, thus providing better damping efficiency but requiring explicit tuning of its hyper-parameters. Whereas the Lead-Lag design, however simple and uses intrinsic properties of the system, has lower damping efficiency than PID-based PSS and reasonable settling time. The settling time with PID-based PSS is reduced by 0.6 seconds in case of a three-phase short circuit fault and 1.04 seconds in case of a single line to ground fault as compared to Lead-Lag PSS. Therefore, it can be concluded that with respect to the oscillations in generator active power, PID-based PSS performs better.

Considering this, in future work, the model of the fuzzy system will be exploited to achieve such results. Other than this, an iterative learning approach will be used that will update its internal state based on the varying dynamic conditions of investigated power system.

## Acknowledgment

This is to declare that there is no conflict of interest.

## Funding

The authors are pleased to express their gratitude for the technical support provided by the Department of Electrical Engineering, Mehran University of Engineering and Technology authorities for their support.

## APPENDIX A

TABLE III. PARAMETERS FOR THE LEAD-LAG PSS

| Parameter | Value |
|---|---|
| $K_{stab}$ | 20 |
| $T_w$ | 2 |
| $V_{stab\ Max}$ | 0.15 |
| $V_{stab\ Min}$ | -0.15 |

### REFERENCES

[1] Katyara, S., Hashmani, A., Chowdhary, B.S., Musavi, H.A., Aleem, A., Chachar, F.A. and Shah, M.A., 2021. Wireless networks for voltage stability analysis and anti-islanding protection of smart grid system. Wireless Personal Communications, 116(2), pp.1361-1378.

[2] Shaikh, M.F., Katyara, S., Khand, Z.H., Ali Shah, M., Staszewski, L., Bhan, V., Majeed, A., Shaikh, S. and Zbigniew, L., 2021. Novel Protection Coordination Scheme for Active Distribution Networks. Electronics, 10(18), p.2312.

[3] Kundur, P.S. and Malik, O.P., 2022. Power system stability and control. McGraw-Hill Education.

[4] Chaubey, P., Lather, J.S., Yelisetti, S., Manda, S. and Yadav, N.K., 2019. Robust power system stabilizer based on static output feedback approach to enhance power system stability. Energy Procedia, 158, pp.2960-2965.

[5] Kim, J.J. and Park, J.H., 2021. A novel structure of a power system stabilizer for microgrids. Energies, 14(4), p.905.

[6] Ismail, M. and Bendary, A., 2021. Improvement of Dynamic Stability of a Single Machine Infinite-Bus Power System using MPC based Power System Stabilizer. ERJ. Engineering Research Journal, 44(4), pp.385-390.

[7] Devarapalli, R., Sinha, N.K. and García Márquez, F.P., 2022. A review on the computational methods of power system stabilizer for damping power network oscillations. Archives of Computational Methods in Engineering, pp.1-27.

[8] Shahani, Z.A., Hashmani, A.A. and Shaikh, M.M., 2020. Steady state stability analysis and improvement using eigenvalues and PSS. Engineering, Technology & Applied Science Research, 10(1), pp.5301-5306.

[9] Kyriakides, E. and Farmer, R.G., 2004. Modeling of damping for power system stability analysis. Electric Power Components and Systems, 32(8), pp.827-837.

[10] Larsen, E.V. and Swann, D.A., 1981. Applying power system stabilizers part I: general concepts. IEEE Transactions on

Power Apparatus and systems, (6), pp.3017-3024.

[11] Furuya, S. and Irisawa, J., 2006. A robust H∞ power system stabilizer design using reduced-order models. International Journal of Electrical Power & Energy Systems, 28(1), pp.21-28.

[12] Tse, C.T. and Tso, S.K., 1993. Refinement of conventional PSS design in multimachine system by modal analysis. IEEE Transactions on power systems, 8(2), pp.598-605.

[13] Soliman, M., Elshafei, A.L., Bendary, F. and Mansour, W., 2010. Robust decentralized PID-based power system stabilizer design using an ILMI approach. Electric Power Systems Research, 80(12), pp.1488-1497.

[14] Prabhakaran, K.K., Tiwari, R., Nayak, S.R. and Chelliah, T.R., 2020, December. Performance investigation on damping of active power oscillation in the large hydro-power plant with power system stabilizer. In 2020 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES) (pp. 1-6). IEEE.

[15] Dasu, B., Kumar, M.S. and Rao, R.S., 2019. Design of robust modified power system stabilizer for dynamic stability improvement using Particle Swarm Optimization technique. Ain Shams Engineering Journal, 10(4), pp.769-783.

[16] Mahdavian, M., Amini, A.A. and Hashemi, M., 2022. Design of a PID-PSS Power System Stabilizer for Single-Machine Infinite-Bus Power System. Signal Processing and Renewable Energy, 6(2), pp.67-83.

[17] Farhad, Z., Ibrahim, E.K.E., Tezcan, S.S. and Safi, S.J., 2018. A robust PID power system stabilizer design of single machine infinite bus system using firefly algorithm. Gazi University Journal of Science, 31(1), pp.155-172.

[18] ASV, V.L., Manyala, R.R. and Mangipudi, S.K., 2020. Design of a robust PID-PSS for an uncertain power system with simplified stability conditions. Protection and Control of Modern Power Systems, 5(1), pp.1-16.

[19] Ghany, M.A. and Shamseldin, M.A., 2020. Model reference self-tuning fractional order PID control based on for a power system stabilizer. International Journal of Power Electronics and Drive Systems, 11(3), p.1333.

[20] El-Dabah, M.A., Kamel, S., Abido, M.A.Y. and Khan, B., 2022. Optimal tuning of fractional-order proportional, integral, derivative and tilt-integral-derivative based power system stabilizers using Runge Kutta optimizer. Engineering Reports, 4(6), p.e12492.

[21] Genesio, R., Tartaglia, M. and Vicino, A., 1985. On the estimation of asymptotic stability regions: State of the art and new proposals. IEEE Transactions on automatic control, 30(8), pp.747-755.

Vol. 5, No. 2 | July - December 2022

# Efficient Detection and Recognition of Traffic Lights for Autonomous Vehicles Using CNN

Tayyaba Sahar[1]*, Hayl Khadhami[2], Muhammad Rauf[1]

**Abstract:**

*Smart city infrastructure and Intelligent Transportation Systems (ITS) need modern traffic monitoring and driver assistance systems such as autonomous traffic signal detection. ITS is a dominant research area among several fields in the domain of artificial intelligence. Traffic signal detection is a key module of autonomous vehicles where accuracy and inference time are amongst the most significant parameters. In this regard, this study aims to detect traffic signals focusing to enhance accuracy and real-time performance. The results and discussion enclose a comparative performance of a CNN-based algorithm YOLO V3 and a handcrafted technique that gives insight for enhanced detection and inference in day and night light. It is important to consider that real-world objects are associated with complex backgrounds, occlusion, climate conditions, and light exposure that deteriorate the performance of sensitive intelligent applications. This study provides a direction to propose a hybrid technique for Traffic Light Detection (TLD) in the daytime and at night. The experimental results successfully improve the night-time detection accuracy of traffic lights from 71.84% to 79.51%.*

**Keywords:** *Object Detection; Convolutional Neural Networks; You Only Look Once; Intelligent Transportation Systems; Hough Transform; Traffic Signal Lights*

## 1. Introduction

Humans have eyes and brains that signify their natural ability to detect and classify objects around them. Their detection and recognition capabilities are uncompilable to artificially intelligent systems. Recent progressions prompting upgrades in proficiency and execution in this field would ease human existence by facilitating them through intelligent systems. Similarly, the provision of convenient applications for drivers and road safety is crucial for Intelligent Transportation Systems (ITS). It aims to modernize the operation of vehicles and promotes driver-assisted systems and driverless cars. Traffic light detection in real-time is challenging since it is associated with real-world problems. This includes complex backgrounds, occlusion, climate conditions, and light exposure that depreciates the performance of sensitive intelligent applications. The efficiency of ITS is crucial, whereas less efficiency is reported due to the variation of light exposures specifically day and night light.

The motivation for this study is inspired by some real-life problems associated with TLD. Problems such as biasness between traffic lights and signs, and variations in day and night light recognition efficiency. Therefore, different experiments and techniques are needed to be tested to enhance the detection accuracy of traffic lights in day and night light.

[1]Department of Electronic Engineering, Dawood University of Engineering & Technology, Karachi, Pakistan
[2]Department of Industrial Engineering and Management. Dawood University of Engineering & Technology, Karachi, Pakistan
Corresponding Author: sahartayyaba@gmail.com

The development of ITS has evolved in two stages: data acquisition and processing followed by the development of technologies for vehicle safety such as collision detection and avoidance [1]. These systems are important for urban planning and future smart cities, because of transportation and transit efficiency [2,3]. TLS is a salient module in Driver Assistance Systems (DAS) and autonomous vehicles [4]. The placement of cameras, the distance of objects, light exposure, and the processing ability of vehicular chips affect the traditional computer vision-based systems used for TLD [4-6]. Deep learning-based CNN models have been deployed in numerous applications after the achievement of incredible results of Alex Net [7]. You Only Look Once (YOLO) version 3 is a 106-layer [8] fully convolutional underlying architecture that is a variant of the original Darknet which comprises 53 layers of a network trained on ImageNet.

Experimental studies reveal that compared to other deep learning models YOLOv3 is a faster, stronger, and more reliable real-time object detector [9]. For automated vehicles, timely detection of traffic lights and their changing states is very important whereas YOLO does generalize object representation without precision losses than other models [10]. A single neural network predicts the boundary boxes and class possibilities from an image of the video frame in a single evaluation. Usually, traffic signs and lights are placed together, these images appear relatively smaller in road view images. Thus, true recognition and detection become challenging as it covers only 1%-2% of the total image area [11]. The combination of CNN and hand-crafted techniques is applied to determine the best input features [12]. Well-acknowledged pattern analysis tools such as Hough Transform (HT) result in realistic outcomes against distortion and diffraction. On the other hand, high material needs and computation costs are two disadvantages that come with it [13]. The latest research and advancements in the field of ML enable to combine the different algorithms from similar or different domains to complement each other [14].

Throughout the years, an increment in the number of cars on roads has increased the frequency of causalities. This endangers human life and safety therefore computer vision techniques are needed to be utilized for observing the immediate data in real-time [15]. Existing studies reveal various intrusive and non-intrusive such as in situ techniques and in-vehicle technologies that are used for traffic monitoring. Computer vision-based techniques have shown better performance than traditional ones [16]. Vitas et al. proposed a hybrid model for the traffic light recognition system using adaptive thresholding and deep learning for region proposal and localization of traffic lights [17]. The researchers used an open-source LISA dataset and custom augmentation to increase the number of data samples. On the other hand, the classification part of the algorithm gave off an 89.60% true detection rate, while the regression correctly localized 92.67% of the traffic lights [17].

TLD is challenging due to the small size and colors that may be similar to the backgrounds. Faster Region-based Convolutional Neural Networks (R-CNN) and Grassmann Manifold Learning have a high degree of accuracy and are robust [18]. Comparative studies of several state-of-the-art methods show that variations and cascaded detection techniques help to deal with real-life issues that affect the detection process [18, 19]. Moreover, other kinds of traffic lights including traffic signs and pedestrian lights have been identified as the main cause of false positives. Deep Convolutional TLD shows an overall detection performance of 0.92 average precision of traffic lights [19].

The brief analysis of the state-of-the-art reveals that TLD using color space and shape detection is dominant for finding the exact parameters of traffic lights [20].In this paper, we propose a framework for efficient TLD using deep learning and hand-crafted technique to monitor the day and night time efficiency. Moreover, the core contributions of the paper include proposing a framework for the hybrid technique for TLD. The foremost target is to improve the night light efficiency of the algorithm. Whereas, it also emphasizes the availability of the dataset having an equal proportion of day and night light images for accurate training. In addition, a subset of this type of dataset has been prepared for conducting this experiment.

This technique is practically implementable using real-sense cameras having day and night vision capabilities mounted behind the windshield of autonomous vehicles. Enhanced real-time detection and recognition of traffic lights will help to cope with the increasing rate of road accidents and causalities.

The research is organized in sections as follows: The introduction outlines the background, motivation, brief review of the state-of-the-art, contribution, and application of this technique. The subsequent sections of this paper include methods and then the experimental results for the applied methodology. The last section concludes the research outcomes.

## 2. Method

This paper outlines an efficient TLD technique to strike a balance between recognition accuracy and overall model complexity. This experimental study aims to propose a hybrid technique that has been summarized in Fig. 1 & 2. With the proposed method TLD accuracy in different lightening conditions can be detected through three major stages: image preprocessing, feature extraction, and image identification.

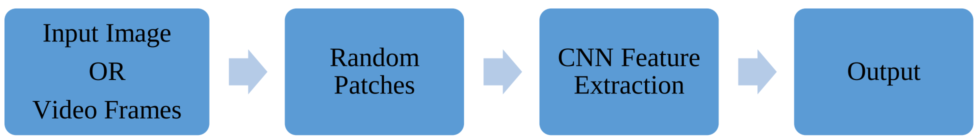



**Fig. 1.**Traffic Light Detection Using YOLO v3


**Fig. 2.**Traffic Light Detection Using Hybrid Method

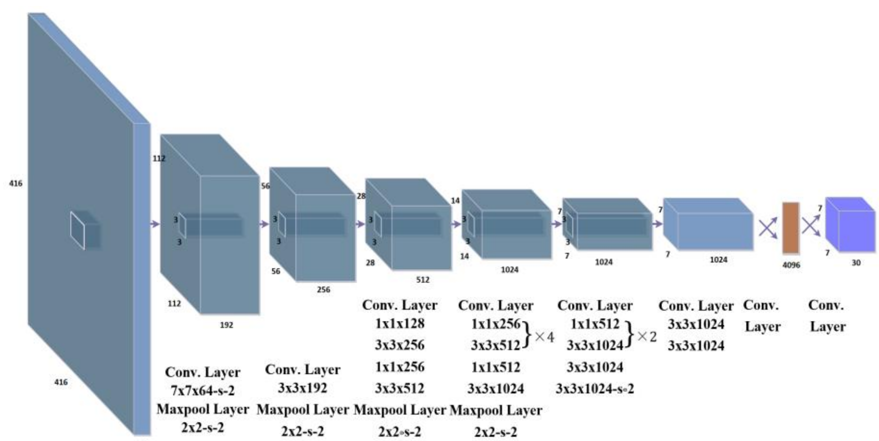



**Fig. 3.**YOLO Network Architecture [23]

### 2.1. YOLO v3

In our approach, we have used YOLO, often called a clever CNN, because it is super-fast and supposed to be run in real-time [21-22]. It incorporates features learned by a Deep Convolutional Neural Network for detecting the objects it was first defined by Joseph Redmon and Ali Farhadi in the seminal 2015 paper. This model depends on the "Darknet" architecture shown in Fig.3 [24]. It processes all image features; 2 fully connected layers are used for bounding box prediction for objects.

Mathematically, in terms of regression, an input image is divided into an S × S grid. Furthermore, the boundary boxes comprise five elements including the object's x and y coordinates that need to be detected in the input image. The other two elements are w and h. that is the width and height of the same image. The last and fifth element is the most important that reveals the confidence score. The confidence score predicts the presence of the object in the box, along with the accuracy of the boundary box that is indicated in Fig. 4

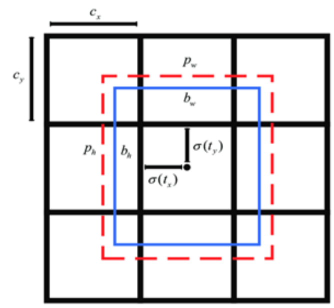



**Fig. 4.** Elements of boundary boxes [24

This algorithm also has an appealing feature that enables the end-user to change the model size to make the trade-off between detection accuracy and speed [27]. Consequently, we have made changes within the config file and a number of filters are being changed as per the classes selected for traffic light detection.


**Fig. 5.**Steps of YOLO algorithm for TLD

### 2.2. Hough Transform

To improve the results, a pattern recognition technique called Hough Transform (HT) is deployed. Usually, it is used to extract lines, circles, and similar ellipses or conic sections. All around the world traffic lights are usually circular, in this case, YOLOv3 gives FP (False Positives) due to biasness between the traffic light and traffic signs. In addition to this, YOLOv3 also struggles while detecting small objects in larger images. Moreover, ITS need more enhanced detection and recognition capabilities in night light as well. In this regard, CHT (Circular Hough Transform) is used for feature extraction [24] that first converts an image from cartesian to polar coordinates.

Mathematically, a circle is represented as follows:

$$(x - x_{center})^2 + (y - y_{center})^2 = r^2 \quad (1)$$

The center of the circle is (x_center ,y_center ) and the radius is r, similarly for using Circle Hough Transform (CHT) for TLD we have used a 2D accumulator with known r as shown in Fig. 6.

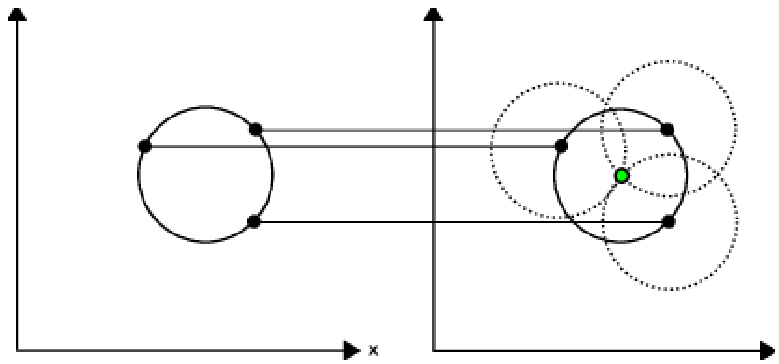



**Fig. 6.**Circle detection using HT [28]

In other words, the number of unknown parameters is equal to the dimension of the accumulator of a given Hough Transform. HT requires pre-processing to find edges in the original traffic light image, at first traffic light images are converted into binary operation with threshold operation that is followed by a morphological operation to remove background noise, these are depicted in Fig. 8.

As compared to CNN-based models image processing approach is quite uncomplicated, however, it undergoes critical phases such as thresholding, and filtering [26]. In this case, OpenCV is preferred because it enables to Hough Gradient method for gradient information of traffic light edges.

## 3. Experimental Results

The proposed model has been trained and tested on a total of 500 images at a resolution of 608 x 608 pixels of the Bosch Small Traffic Lights Dataset, 85 % of images are used for training; while 15 % are used for the test. The annotations include bounding boxes of traffic lights as well as the color of each traffic light. It is known as an accurate dataset for vision-based traffic light detection [21]. The Bosch dataset is selected because it allows easy testing of objection detection approaches, especially for small objects in larger images in different lighting environments.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FP} \quad (2)$$

At first, traffic light detection has performed using YOLO v3. The Bosch Lite dataset is trained on images with 3 classes (Red, Yellow, and Green) while available unseen test images are used for testing purposes. The following figures show detected traffic lights on the trained model, all three lights including red, yellow, and green are indicated within bounding boxes along with confidence scores.

(a) (b) (c) (d)

**Fig. 7.** Output Images YOLO V3 in different lightening conditions. (a) Green light detection (b) Yellow light detection (c) Red Light detection (d) Misc. lights detection

TABLE I. SUMMARY OF RESULTS

| Model | Frames | TP | FP | Day Time Efficiency | Night Time Efficiency |
|---|---|---|---|---|---|
| YOLOV3 | Day: 1509 | 1504 | 5 | 99.68 % | 71.84 % |
| | Night: 1450 | 1042 | 408 | | |
| | Total: 2959 | | | | |

TABLE II. SUMMARY OF CHT RESULTS

| Model | Frames | TP | FP | Day Time Eff. | Night Time Eff. |
|---|---|---|---|---|---|
| CHT | Day: 1509 | 1308 | 202 | 86.62% | 79.51% |
| | Night: 1450 | 1153 | 297 | | |
| | Total: 2959 | | | | |

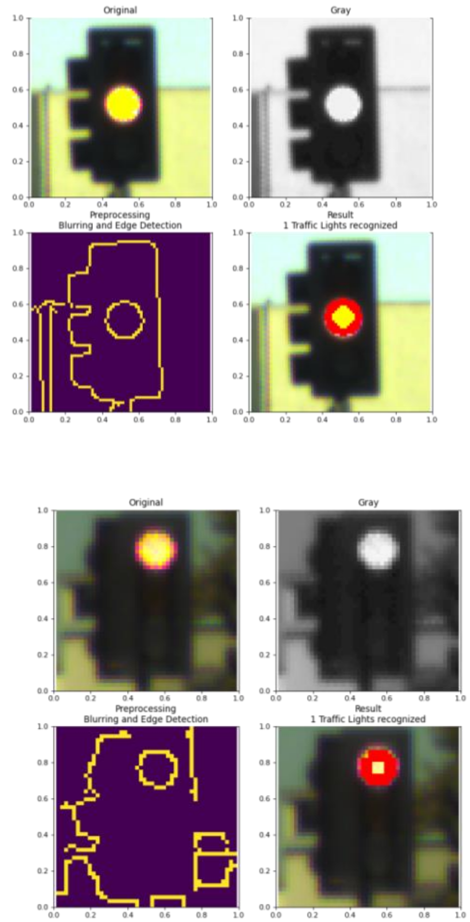



**Fig. 8.** Output Images of CHT

## 4. Conclusion

Intelligent Transportation Systems can provide convenient, sustainable, safe, and secure transportability. However, autonomous vehicles undergo several challenges including infallible detection and recognition of traffic lights. These problems can be eliminated by using enhanced computer vision techniques. Over the decades, the expeditious evolution of deep learning models introduced an object detection model YOLO that can efficiently detect traffic lights in real time. Contrary to this, real-life problems limit the accuracy and inference time. In essence, the development of a hybrid approach is needed to encounter the biasness between traffic lights and signs along with improved efficiency at night time. This study supports the development of a hybrid technique combining deep learning and handcrafted technique to improve traffic light detection and recognition in the best possible way. The experimental results successfully improve the night-time detection accuracy of traffic lights from 71.84% to 79.51%.

## REFERENCES

[1] Rauf, Muhammad, et al. "Response Surface Methodology In-Cooperating Embedded System for Bus's Route Optimization." Research Journal of Applied Sciences, Engineering and Technology 5.22 (2013): 5170-5181

[2] Sharma, S. and S.K. Awasthi, Introduction to intelligent transportation system: overview, classification based on physical architecture, and challenges. International Journal of Sensor Networks, 2022. 38(4): p. 215-240.

[3] Yuan, T., et al., Machine learning for next-generation intelligent transportation systems: A survey. Transactions on Emerging Telecommunications Technologies, 2022. 33(4): p. e4427.

[4] Ouyang, Z., et al., Deep CNN-based real-time traffic light detector for self-driving vehicles. IEEE transactions on Mobile Computing, 2019. 19(2): p. 300-313.

[5] Li, X., et al., Traffic light recognition for complex scene with fusion detections. IEEE Transactions on Intelligent Transportation Systems, 2017. 19(1): p. 199-208.

[6] Jensen, M.B., et al., Vision for looking at traffic lights: Issues, survey, and perspectives. IEEE Transactions on Intelligent Transportation Systems, 2016. 17(7): p. 1800-1815.

[7] Krizhevsky, A., I. Sutskever, and G.E. Hinton, Imagenet classification with deep convolutional neural networks. Communications of the ACM, 2017. 60(6): p. 84-90.

[8] Redmon, J. and A. Farhadi, Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.

[9] Mujahid, A., et al., Real-time hand gesture recognition based on deep learning YOLOv3 model. Applied Sciences, 2021. 11(9): p. 4164.

[10] Chandana, R. and A. Ramachandra, Real Time Object Detection System with YOLO and CNN Models: A Review. 2022.

[11] Rehman, Y., et al., Small Traffic Sign Detection in Big Images: Searching Needle in a Hay. IEEE Access, 2022. 10: p. 18667-18680.

[12] Sani, U.S., O.A. Malik, and D.T.C. Lai, Improving Path Loss Prediction Using Environmental Feature Extraction from Satellite Images: Hand-Crafted vs. Convolutional Neural Network. Applied Sciences, 2022. 12(15): p. 7685.

[13] Kumar, S., et al. Lane and Vehicle Detection Using Hough Transform and YOLOv3. in 2022 2nd International Conference on Intelligent Technologies (CONIT). 2022. IEEE.

[14] Sahar, T., Rauf, M., Murtaza, A., Khan, L. A., Ayub, H., Jameel, S. M., & Ahad, I. U. (2022). Anomaly detection in laser powder bed fusion using machine learning: A review. Results in Engineering, 100803.

[15] Bao, C., et al. Safe driving at traffic lights: An image recognition based approach. in 2019 20th IEEE International Conference on Mobile Data Management (MDM). 2019. IEEE.

[16] Jain, N.K., R. Saini, and P. Mittal, A review on traffic monitoring system techniques. Soft Computing: Theories and Applications, 2019: p. 569-577.

[17] Vitas, D., M. Tomic, and M. Burul, Traffic light detection in autonomous driving systems. IEEE Consumer Electronics Magazine, 2020. 9(4): p. 90-96.

[18] Gupta, A. and A. Choudhary. A Framework for Traffic Light Detection and Recognition using Deep Learning and Grassmann Manifolds. in 2019 IEEE Intelligent Vehicles Symposium (IV). 2019. IEEE.

[19] Bach, M., D. Stumper, and K. Dietmayer. Deep convolutional traffic light recognition for automated driving. in 2018 21st International Conference on Intelligent Transportation Systems (ITSC). 2018. IEEE.

[20] Iftikhar, M., et al., Traffic Light Detection: A cost effective approach. 2022.

[21] Behrendt, K., L. Novak, and R. Botros. A deep learning approach to traffic lights: Detection, tracking, and classification. in 2017 IEEE International Conference on Robotics and Automation (ICRA). 2017. IEEE.

[22] Yadav, P.V., et al., AquaVision: Real-Time Identification of Microbes in Freshwater Using YOLOv3, in Soft Computing for Security Applications. 2022, Springer. p. 437-448.

[23] Redmon, J., et al. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[24] Li, X., et al., A modified YOLOv3 detection method for vision-based water surface garbage capture robot. International Journal of Advanced Robotic Systems, 2020. 17(3): p. 1729881420932715.

[25] Chun, L. Z., Dian, L., Zhi, J. Y., Jing, W., & Zhang, C. Yolov3: Face detection in complex environments. International Journal of Computational Intelligence Systems, (2020). 13(1), 1153-1160.

[26] Gothankar, N., C. Kambhamettu, and P. Moser. Circular hough transform assisted cnn based vehicle axle detection and classification. in 2019 4th International Conference on Intelligent Transportation Engineering (ICITE). 2019. IEEE.

[27] Kulkarni, R., S. Dhavalikar, and S. Bangar. Traffic light detection and recognition for self driving cars using deep learning. in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). 2018. IEEE.

[28] Ribeiro, B.M., et al. Arbitrary ball detection using the circular Hough transform. in Proc. of the 15th Portuguese Conf. on Pattern Recognition, RECPAD. 2009.

Vol. 5, No. 2 | July – December 2022

# Evaluation of Machine Learning based Network Attack Detection

Muhammad Awais Rajput[1], Muhammad Umar[2], Adnan Ahmed[2], Ali Raza Bhangwar[3], Khadija Suhail Memon[2], Misbah[2]

**Abstract:**

The growth in the internet and communication technologies has driven tremendous developments in various application areas such as smart cities, cloud computing, internet-of-things, e-banking, e-commerce, and e-government. However, the advancements in networking infrastructure, hacking tools, and methodologies have enabled hackers to attempt newer and more complicated cyber-attacks. Consequently, cyber-security has now emerged as a vital research area to address security concerns. Traditional security mechanisms such as firewalls and anti-viruses are not enough to accurately detect intrusions. Therefore, an Intrusion Detection System (IDS) provides an additional layer of security to prevent intrusions through continuous surveillance of the network traffic. Machine Learning (ML) and Deep Learning (DL) techniques have been exploited to overcome the inherent deficiencies of IDS such as accurately detecting intrusions, countering zero-day cyber-attacks, and reducing false positive rates. Existing research has demonstrated that ML and DL-based techniques can efficiently detect patterns (features) from the network traffic and predict the behavior (normal or abnormal activity) based on these patterns. This research work first presents the concepts of IDS, followed by a comprehensive review of the recent ML and DL-based schemes. Later, a performance analysis of various ML algorithms such as Decision Trees (DT), Random Forest (RF), Gradient Booster (GB), and Deep Neural Networks (DNNs) is presented on a publicly available dataset. The performance is reported in terms of accuracy, F1-Score, cross-entropy loss, and training and testing times.

**Keywords:** *machine learning; intrusion detection systems; cyber security; anomaly detection*

## 1. Introduction

The recent advancement in emerging technologies such as healthcare, telecommunication, education, intelligent transportation systems, smart grids, e-commerce and e-government, manufacturing and infotainment provide ease and quality of services to the people [1], [2]. Despite the significant advantages, these technologies are exposed to numerous cyber threats and unauthorized access [3]. To mitigate and eliminate the impacts of these attacks, cybersecurity provides various methods and technologies, such as antivirus, encryption/decryption, firewalls, access control and IDS, to protect data and network infrastructures [4]–[6]. Although these methods can prevent various attacks, however, in-depth traffic analysis cannot be performed using these security techniques.

[1]Dept. of Artificial Intelligence, Quaid-e-Awam University of Engineering, Science and Technology, Nawabsah, Pakistan
[2]Dept. of Telecommunication, Quaid-e-Awam University of Engineering, Science and Technology, Nawabsah, Pakistan.
[3]Dept. of Computer Systems, Quaid-e-Awam University of Engineering, Science and Technology, Nawabsah, Pakistan.
Corresponding Author: adnan.ahmed03@quest.edu.pk

For continuous network traffic surveillance, IDS have been developed to perform deeper traffic analysis. However, the IDS still poses challenges in detecting zero-day attacks and minimizing false alarm rates [7]. To overcome the limitations of existing IDS schemes and to provide accurate, cost-effective and efficient IDS, machine learning and deep learning techniques have been integrated to learn network traffic's features and then predict and distinguish between the benign and normal traffic patterns. In this paper, we evaluate the performance of machine learning algorithms such as Decision Trees (DT), Random Forest (RF), Gradient Booster (GB) and Deep Neural Networks (DNNs) on various cyber security attacks: Denial of Service (DoS), Remote to Local (R2L), User to Root (U2R) and probe. The work was tested with the publicly available dataset KDD Cup99 and performance was evaluated in terms of accuracy and training time.

The contributions of this paper are highlighted in the following:

1) A machine learning-enabled attack detection flow is implemented.
2) The performance of various machine learning algorithms including deep neural networks is evaluated on a publicly available dataset.

The rest of the paper is organized as follows: Section 2 presents the relevant literature and discusses the ML and DL-based IDS schemes. Section 3 presents the methodology and overall flow of the work. Section 4 presents the results and discussion. Finally, section 5 concludes the paper.

## 2. Literature Review

The intrusion detection scheme based on linear discriminant analysis (LDA), classification and Regression trees CART has been used in [8]. For testing purpose, Random Forest and KDD Cup 99 dataset has been used in manner that it is divided on 80-20 rule. The performance of IDS schemes (LDA, RF and CART) is evaluated in terms of accuracy and Cappa. According to result analysis, the RF scheme performs better in accuracy (99.65%) than LDA (98.1%) and CART (98%).

In paper [9], a performance comparison of machine learning-based algorithms, KNN, decision tree and AdaBooost, is performed using TON-IoT data set. The 99.8% accuracy was achieved for AdaBoost scheme which was better than KNN and decision tress schemes.

The Intrusion Detection Tree (IntruDTree) machine-learning-based security model was proposed in [10], which predicts the unseen test cases accurately and also reduces the computational complexity by optimizing feature dimensions. The efficacy of proposed IntruDTree was examined with cybersecurity datasets and performance was measured in terms of ROC, accuracy, recall and precision. The performance of IntruDTree was compared with Naive Bayes (NB), Logistic Regression (LR), K-Nearest Neighbor (KNN), and Support Vector Machines (SVM). The IntruDTress achieved better results than their counterpart.

The paper [11] presented the variant of the FNN known as Self-normalizing Neural Networks (SSN) and compare its performance with the FNN. The evaluation was implemented on the BoT-IoT dataset. The SSN scheme is better than FNN based on multiple metrics such as accuracy, precision, and recall as well as multi-classification metrics such as Cohen Cappa's score.

The authors in [12] used Radial Based Function (RBF) for support vector machine for classifying DoS, Probe, R2L and U2R types of attack. Two datasets namely "Mixed" and "10% KDD Cup99" datasets have been used for evaluating performance of intrusion detection scheme. According to result analysis, validation accuracy was estimated to 89.85% and 99.9% for mixed and KDD respectively.

Table 1: Comparative analysis

| PAPERS | DATASET | ML METHODS | DL METHODS | PERFORMANCE METRICS | ATTACK DETECTION |
|---|---|---|---|---|---|
| [7] | KDD cup 99 | LDA, RF, CART | -- | Accuracy and Kappa | Normal, probe, U2R and R2L |
| [8] | TON IoT | KNN, Ada Boost | -- | Precision, Recall, F1 score, Accuracy, and ROC | Normal, scanning, DDoS, DoS, password attacks |
| [9] | IntruDTree | KNN, SVM, LR, NB | -- | precision, recDOI, F1-score, accuracy, and ROC | DoS Malware |
| [10] | BoT IoT | -- | FNN, SNN | precision, recall and F1-score, Kappa Score and MC | Normal, DDoS, DoS, and Reconnaissance |
| [11] | KDD cup 99 | SVM | -- | Accuracy | Denial of service (Dos), Probe, User to Root (U2R), Remote to User (R2L) |
| [12] | DARPA 1999 | KNN | -- | Accuracy and F1-score | Snort based attack detection and minimizing false alarms |
| [13] | MalShare | Decision Tree | -- | Accuracy | APT Attacks |
| [14] | KDD cup 99 | -- | DNN | Accuracy, Training and testing time | Denial of service (Dos), Probe, User to Root (U2R), Remote to User (R2L) |
| [15] | KDD Cup 99 | LR | -- | Accuracy and CPU time(s) | Normal. DOS, probe, U2L, R2L |
| [16] | NSL-KDD | -- | RNN-IDS | Accuracy and Training time | Normal, DoS, probe, U2R and R2L |

Meng et al. [13] developed an intelligent knowledge-based alarm filter based on KNN-classifier with objective function to minimize the false alarm rate. The DARPA 1999 dataset was used to train the filter. A network environment was setup using snort and Wireshark where real-world web traffic was monitored. Snort detects various types of attacks and generated alerts. The generated alters were forwarded to KNN-based alarm filter for further analysis thereby filtering and minimizing the false alarm rate. The result analysis showed that accuracy of the design system was 85.2% and F-score was 0.82.

Moon et al. [14] proposed a DTB-IDS (Decision Tree-based IDS) for detection and preventing Advance Persistent Threat (APT). The proposed system executes the malicious code on the virtual environment and then analyze the behavior to detect APT. the result analysis depicts the accuracy of proposed system was 84.7%.

The author in [15] proposed an IDS that classify the non-labeled data using ladder network and then classify the non-labeled data using Deep belief Network (DBN). Moreover, the proposed scheme also integrates the semi-supervise learning with neural network with aim to achieve high accuracy with small number of labeled samples using KDD Cup99 dataset. The detection accuracy was 99.18%.

An IDS based on Deep Belief Network (DBN) using logistic regression was proposed in [16]. In order to improve the overall performance of IDS system, the multi-class logistic regression was trained on 10 epochs with pre-trained data of the 10% KDD Cup99 dataset. A low false rate of 2.47% was measured and detection rate of 97.9% was achieved for the proposed scheme.

A Recurrent Neural Network (RNN) based IDS was proposed by Ying et al. [17] using NSL-KDD data set. The performance of proposed model was evaluated in binary classification, influence of number of neurons, learning rates and multi-class classification. The result analysis showed that multi-class classification achieved the training accuracy 99.53% and test accuracy 81.29%. The binary classification model achieves the training accuracy 99.81% and test accuracy 83.28%.

A categorical comparison of recent ML and DL based works is provided in Table -1 in an attempt to provide further insight on ML/DL methods, various attack types and datasets that have been recently adopted in this domain.

## 3. Methodology

This section explains our ML-based intrusion detection flow. First, we provide an overview of the main steps involved in the methodology. Later, the key steps are explained in the subsequent subsections.

Figure 1 visualizes the overall flow of the intrusion detection using ML. The flow starts with the dataset available as the input. Initially, a preprocessing step performs tasks to prepare the dataset for the next steps. The preprocessing is an essential step which when overlooked, could affect the accuracy of the later steps of the flow. An in-depth understanding of the dataset is made via visualization and correlation mapping of the features. Figure 2 shows the correlation map of the dataset with only those features for which the correlation is higher than 0.6.

The cleaned dataset is forwarded to the training step where essentially the dataset is divided into train and test sets. The train data is provided to the selected ML model for learning the patterns. The test part of the dataset is kept for prediction purposes for the later step. In the testing step, the trained model from the last step is evaluated by applying unseen data (that was initially kept in the test part of the dataset).

### 3.1. Preprocessing

The capability of quickly learning patterns in the data is what makes ML models preferrable over the traditional approaches [18]. However, for many applications areas, the datasets are not in a standard form. Moreover, the data obtained from repetitive readings requires excessive steps to bring it in a form suited for further processing in the ML pipeline.

The tasks to be performed in preprocessing step vary depending on the application and the type of data that is being dealt with. Nevertheless, there is a set of techniques that have been commonly employed in most of the ML applications. In case of intrusion detection, we apply a sequence of common preprocessing tasks along with some specific tasks that favor the IDS.

Next important task involves understanding of the feature set. For this purpose, correlation map could provide useful insight to check which features could be eliminated due to redundancy. This is really important in many cases including IDS since the dataset (*KDDCUP99*) contains 41 features out of which many could be dropped. Other visualizations could also be used to undertake dataset formatting such as class balancing. In certain supervised ML methods such as SVM, feature mapping is required to represent features of data in suitable space. Finally, suitable and relevant features are kept only for the next step.

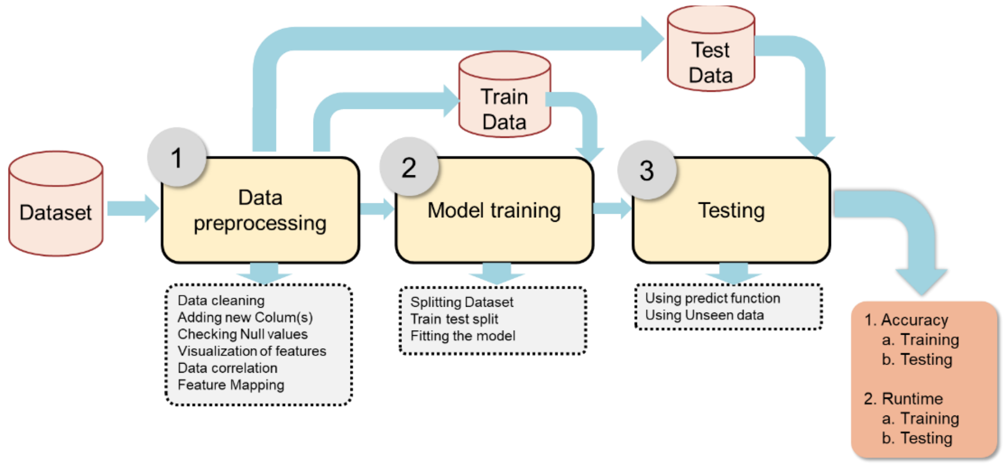



Figure 1: Overall flow

### 3.2. Training and Testing

The training step starts by first splitting train/test data. This is done to avoid over/under fitting of the model. The overfitting effect refers a scenario where the model learns the patterns of a dataset during the train and when tested on the same data, provides extremely accurate results since it has already seen the data. In underfitting, the model suffers to finds patterns in the training data and when tested on the new data, the model accuracy is very low.

The overfitting effect can be minimized by breaking the dataset with appropriate proportions of the training data to let the model learn the relationship of the data and enough samples in the test data to evaluate the model. The underfitting effect can be tackled by using well suited ML model and careful selection of the features to capture essential input/output pattern of the data.

We divide the processed dataset with a proportion of *70%* and *30%* for training and testing of the model. This is done by invoking *train_test_split* function from *sklearn* library. Then for the available training data, we utilize the most common seven ML methods from different python's ML libraries e.g., *sklearn* and *keras*, to train the models. The testing data is prepared by stripping the class label column. The accuracy score of the model is computed by applying the test data to the trained model.

## 4. Results

In this section, we provide details on our experimental setup and obtained results for various ML techniques on the intrusion detection.

### 4.1. Setup of experiment

As discussed in the previous section, we use a standard dataset i.e., *KDDCUP99*, to evaluate ML techniques for accurate prediction of anomalies in the network traffic data. The dataset manipulation is performed mainly via python library *pandas* and *numpy*. For visualization purposes, we use *matplotlib*. For ML models except for DNN, python's *sklearn* library is utilized, whereas for DNN model we use *keras*. The model parameters for all the ML models used in our evaluation are shown in Table-2.

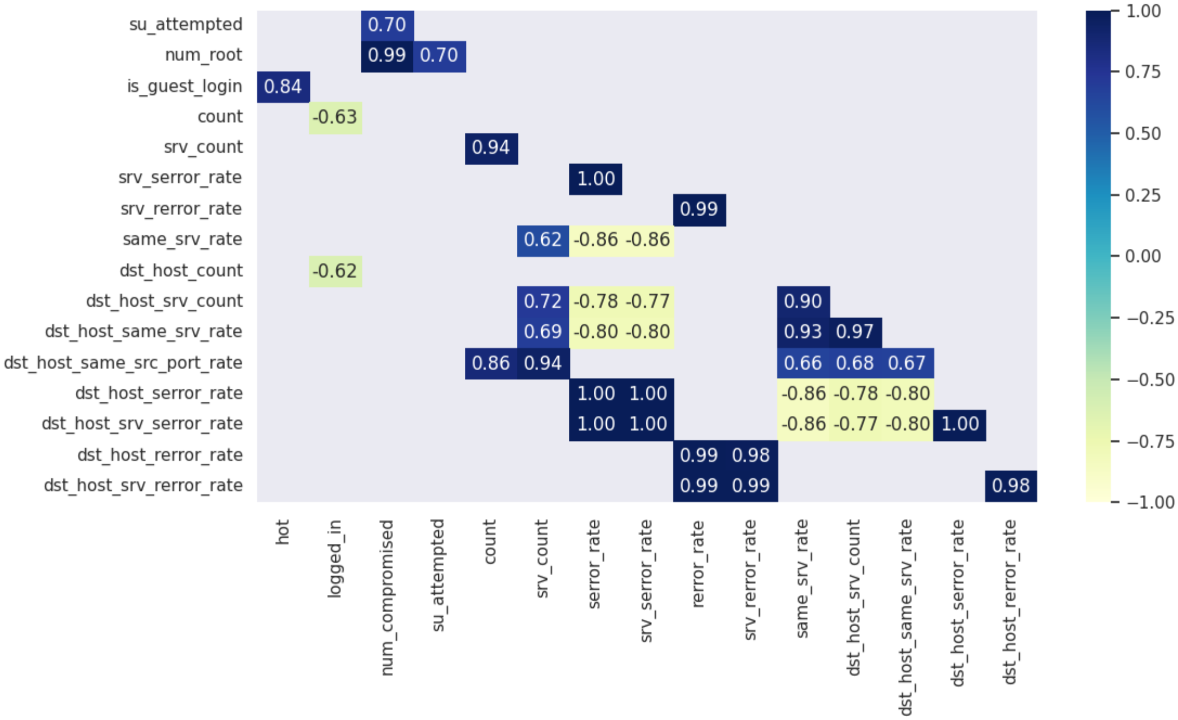



Figure 2: Correlation matrix for attributes of the dataset

## 4.2. Results and Discussion

For each ML classification method from Table-2, we evaluate the performance along two key parameters i.e., time and accuracy. For time, we report training and testing times separately.

Figure 3 and Figure 4 show the training and testing times, respectively, for all ML methods on the dataset. As can be seen from Figure 1, simpler ML models tend to have smaller training times whereas models that involves more complex learning structure could spend more time in learning the relationship of data. In this case, the DNN took the largest training time on the dataset followed by GB, SVC and LR. NB took the smallest training time but as we will see that costs us performance penalty on the accuracy front.

**Table-2: ML algorithm parameters**

| S. # | Model | Parameter |
|---|---|---|
| 1 | Naïve Bayesian (NB) | *default* |
| 2 | Decisions tree (DT) | *criterion="entropy max_depth = 4* |
| 3 | Random Forest (RF) | *n_estimators=30* |
| 4 | Support Vector Machine (SVM) | *gamma = 'scale'* |
| 5 | Linear Regression (LR) | *max_iter =1200000* |
| 6 | Gradient Boosting (GB) | *random_state=0* |
| 7 | Deep Neural Network (DNN) | *epochs=100 batch_size=64 activation=relu optimizer=adam* |

For testing, all algorithms perform equally well except SVC which took significantly larger time on the prediction for the testing part of the dataset. We believe that this increase might be the result of an increase in the support vectors during the prediction step.

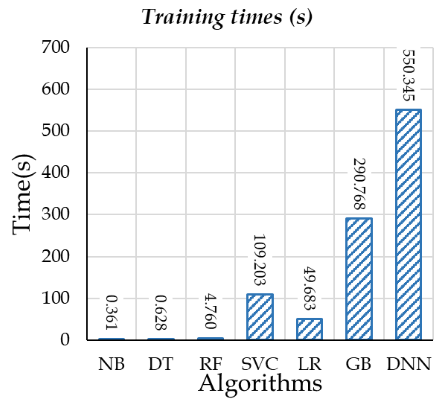



Figure 3: Training times for all ML methods

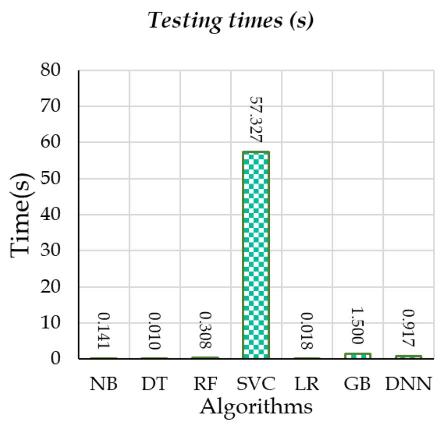



Figure 4: Testing times for all ML methods

In Figure 5 and 6, we report the accuracy scores for all algorithms. In case of accuracy results, we see a similar trend for training and testing accuracy. All algorithms except NB could achieve accuracy higher than 90%. The highest accuracy is achieved by the RF algorithm (1.0). DT, SVC, LR, GB and DNN achieve 0.990, 0.999, 0.993, 0.997, and 0.998 accuracy respectively. For testing accuracy, a similar trend can be observed. Overall RF algorithm turns out to be the most accurate classifier with a very low training and testing time.

To further evaluate the performance of the ML models, we employ a commonly used loss function i.e., cross-entropy loss and report the values for all algorithms in Table-3. It is important to mention that other most common metrics for loss such as means squared error or mean arithmetic error are not applicable in our case since the problem at hand is classification rather than regression and thus these metrics cannot capture the loss function efficiently. Cross-entropy loss function can better represent the training and testing performance in multi-class classification problems [19]. Results of cross-entropy represent a general trend of slight increase in loss for testing except in case of NB classifier where testing loss is a bit lower.

**Table-3: Cross-entropy values for all algorithms**

| **Alg.** | ***Cross-Entropy*** | |
|---|---|---|
| | **Training** | **Testing** |
| NB | 2.777 | 2.776 |
| DT | 0.025 | 0.026 |
| RF | 0.000 | 0.002 |
| SVC | 0.005 | 0.005 |
| LR | 0.017 | 0.017 |
| GB | 0.048 | 0.052 |
| DNN | 0.041 | 0.042 |

In addition to accuracy which might not be sufficient to indicate a model's performance, we attempt get further insight into the performance of the ML algorithms by computing another important quality metric i.e., F1-score that combines precision and recall into a single composite quality indicator. For F1-scores, we observe that the RF archives the highest F1-score (see Figure-7) followed by SVC and LR. The NB classifier could only obtain 0.45 making it the lowest in terms of F1-score.

## 5. Conclusion

This paper presents a work on classification of network traffic for intrusion detection using machine learning techniques.

Intrusion detection has become a challenging task due to data explosion in the recent years driving novel attacks and penetration techniques from the hackers. Manual techniques based on signatures of malicious activity could miss certain attack types or even could not respond to unknown threats. ML techniques are capable of learning complex patterns in the data and can classify unseen data with high accuracy. The experimental results provided in this paper show that ML techniques could separate normal and bad connections with great accuracy (up to 99.9%) on a standard dataset with computationally light-weight inference models.

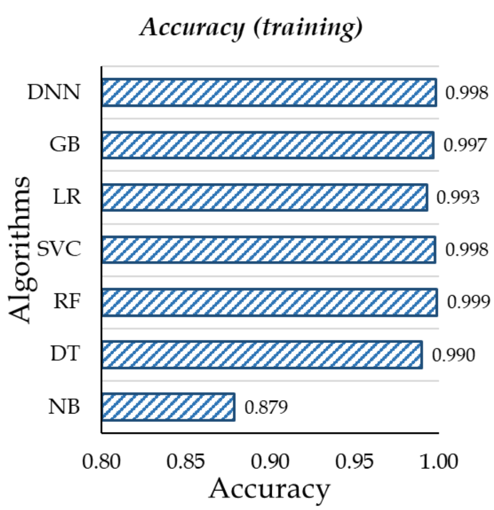



Figure 5: Training accuracy for all ML methods

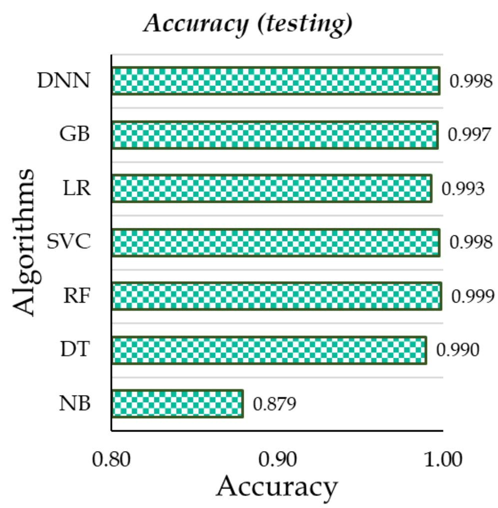



Figure 6: Testing accuracy for all ML methods

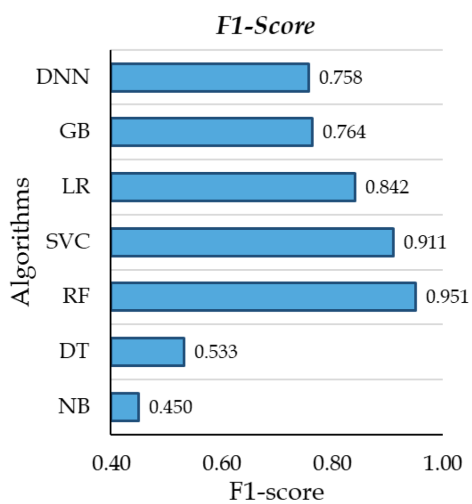



Figure 7: F1-score for all algorithms

## REFERENCES

[1] Z. Zhang *et al.*, "Artificial intelligence in cyber security: research advances, challenges, and opportunities," *Artif Intell Rev*, vol. 55, no. 2, pp. 1029–1053, 2022, doi: 10.1007/s10462-021-09976-0.

[2] Z. Tang *et al.*, "Data Augmentation for Graph Convolutional Network on Semi-supervised Classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12859 LNCS, pp. 33–48, 2021, doi: 10.1007/978-3-030-85899-5_3.

[3] T. T. Nguyen and V. J. Reddi, "Deep Reinforcement Learning for Cyber Security," *IEEE Trans Neural Netw*

*Learn Syst*, 2021, doi: 10.1109/TNNLS.2021.3121870.

[4] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, "Network Intrusion Detection for IoT Security Based on Learning Techniques," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 3, pp. 2671–2701, Jul. 2019, doi: 10.1109/COMST.2019.2896380.

[5] P. Parkar and A. Bilimoria, "A survey on cyber security IDS using ML methods," in *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, May 2021, pp. 352–360. doi: 10.1109/ICICCS51141.2021.9432210.

[6] F. Tao, M. Akhtar, and Z. Jiayuan, "The future of Artificial Intelligence in Cybersecurity: A Comprehensive Survey," *EAI Endorsed Transactions on Creative Technologies*, vol. 8, no. 28, p. 170285, 2021, doi: 10.4108/eai.7-7-2021.170285.

[7] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 1, Jan. 2021, doi: 10.1002/ett.4150.

[8] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. K. A. A. Khan, "Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review," in *Procedia Computer Science*, 2020, vol. 171, pp. 1251–1260. doi: 10.1016/j.procs.2020.04.133.

[9] I. ; Hidayat, M. Ali, ; Zulfiqar, and A. Arshad, "Machine learning based intrusion detection system: an experimental comparison." [Online]. Available: https://edshare.gcu.ac.uk/id/eprint/5179

[10] I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. Khan, "IntruDTree: A machine learning based cyber security intrusion detection model," *Symmetry (Basel)*, vol. 12, no. 5, May 2020, doi: 10.3390/SYM12050754.

[11] O. Ibitoye, O. Shafiq, and A. Matrawy, "Analyzing adversarial attacks against deep learning for intrusion detection in IoT networks," in *IEEE global communications conference (GLOBECOM)*, 2019, pp. 1–6.

[12] M. v. Kotpalliwar and R. Wajgi, "Classification of attacks using support vector machine (SVM) on KDDCUP'99 IDS database," in *Proceedings - 2015 5th International Conference on Communication Systems and Network Technologies, CSNT 2015*, Sep. 2015, pp. 987–990. doi: 10.1109/CSNT.2015.185.

[13] W. Meng, W. Li, and L. F. Kwok, "Design of intelligent KNN-based alarm filter using knowledge-based alert verification in intrusion detection," *Security and Communication Networks*, vol. 8, no. 18, pp. 3883–3895, Dec. 2015, doi: 10.1002/sec.1307.

[14] D. Moon, H. Im, I. Kim, and J. H. Park, "DTB-IDS: an intrusion detection system based on decision tree using behavior analysis for preventing APT attacks," *Journal of Supercomputing*, vol. 73, no. 7, pp. 2881–2895, Jul. 2017, doi: 10.1007/s11227-015-1604-8.

[15] M. Nadeem, O. Marshall, S. Singh, X. Fang, and X. Yuan, "Semi-Supervised Deep Neural Network for Network Intrusion Detection," 2016. [Online]. Available: https://digitalcommons.kennesaw.edu/ccerphttps://digitalcommons.kennesaw.edu/ccerp/2016/Practice/2

[16] K. Alrawashdeh and C. Purdy, "Toward an online anomaly intrusion detection system based on deep learning," in *Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, Jan. 2017, pp. 195–200. doi: 10.1109/ICMLA.2016.167.

[17] C. Yin, Y. Zhu, J. Fei, and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," *IEEE Access*, vol. 5, pp. 21954–21961, Oct. 2017, doi: 10.1109/ACCESS.2017.2762418.

[18] F. Deeba *et al.*, "A novel image dehazing framework for robust vision-based intelligent systems," *International Journal of Intelligent Systems*, Dec. 2021, doi: 10.1002/INT.22627.

[19] Z. Zhang, M. S.-A. in neural information, and undefined 2018, "Generalized cross entropy loss for training deep neural networks with noisy labels," *proceedings.neurips.cc*, Accessed: Jan. 20, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2018/hash/f2925f97bc13ad2852a7a551802feea0-Abstract.html