UNIVERSITY OF RWANDA
COLLEGE OF SCIENCE AND TECHNOLOGY
AFRICAN CENTER OF EXCELLENCE IN INTERNET OF THINGS (ACEIoT)

MACHINE LEARNING BASED-PREDICTIVE MODELING WITH CONTROL MECHANISMS FOR ENHANCED BIOGAS YIELD THROUGH INTERNET OF THINGS-ENABLED SENSOR NETWORKS

Ph.D Thesis submitted in the fulfillment of requirements of the award of PhD. Degree in the Internet of Things – Wireless Sensor Network

ANGELIQUE MUKASINE

SEPTEMBER 2024

UNIVERSITY OF RWANDA
COLLEGE OF SCIENCE AND TECHNOLOGY
AFRICAN CENTER OF EXCELLENCE IN INTERNET OF THINGS (ACEIoT)

MACHINE LEARNING BASED-PREDICTIVE MODELING WITH CONTROL MECHANISMS FOR ENHANCED BIOGAS YIELD THROUGH INTERNET OF THINGS-ENABLED SENSOR NETWORKS

Ph.D Thesis submitted in the fulfillment of requirements of the award of PhD. Degree in the Internet of Things – Wireless Sensor Network

ANGELIQUE MUKASINE

Registration number: 219008518

Thesis Supervisor: Dr. LOUIS SIBOMANA, PhD

Thesis Co-Supervisors: Dr. KAYALVIZHI JAYAVEL, PhD

Thesis Co-Supervisor: Dr. KIZITO UKURIKIYEYEZU, PhD

SEPTEMBER 2024

Declaration

I hereby declare that the dissertation entitled **"Machine Learning-Based Predictive Modeling and Control Mechanisms for Enhanced Biogas Yield Through Internet of Things-Enabled Sensor Networks"** to be submitted for the Degree of Doctor of Philosophy in the Internet of Things (Wireless Sensor Network) at the University of Rwanda, College of Science and Technology is my original work and has not formed the basis for the award of any degree, diploma, associateship, or fellowship of similar other titles. I also declare that all sources of information used are acknowledged by a complete list of references and cited.

Signature:

Angelique Mukasine

Angelique Mukasine, a Ph.D student of UR-ACEIoT registration ID 219008518, successfully defended the thesis/dissertation entitled "**Machine Learning-Based Predictive Modeling and Control Mechanisms for Enhanced Biogas Yield Through Internet of Things-Enabled Sensor Networks**", which she prepared after fulfilling the requirements specified in the associated legislation, before the thesis examination members whose signatures are below.

Thesis Supervisor:	**Dr. Sibomana Louis**	
	National Council for Science and Technology	----------------------------
Co-Supervisor:	**Dr. Jayavel Kayalvizhi**	
	University of the Arts London	----------------------------
Resident Co-Supervisor:	**Dr. Kizito Nkurikiyeyezu**	
	University of Rwanda	----------------------------
Viva Voce Members:	**1**	

	2	

	3	

Date of Submission :

Date of Defense :

Acknowledgements

This thesis represents the culmination of doctoral research at the African Centre of Excellence in Internet of Things (ACEIoT), funded by the Government of Rwanda as part of the African Centers of Excellence (ACEs) initiative.

I extend my gratitude to the Government of Rwanda and the University of Rwanda for their efforts in enhancing and fortifying human capacity, fostering a knowledge-driven economy, and providing top-tier postgraduate education that aligns with market demands in Rwanda.

Allow me to extend my heartfelt appreciation to the administrative team at ACEIoT for their unwavering support and guidance throughout this journey. I am deeply grateful to Prof. Hanyurwimfura Damien and Dr. Gatera Omar for their mentorship, inspiration, administrative assistance, and facilitation in completing various research endeavors.

I extend my immense gratitude to my primary supervisor, Dr. Louis Sibomana, for his steadfast guidance, insightful discussions, provision of opportunities, and various forms of support that were instrumental in the success of this research endeavor. My heartfelt thanks to my co-supervisors, Dr. Kayalvizhi Jayavel and Dr. Kizito Nkurikiyeyezu, for their dedicated assistance and mentorship throughout this research journey. I deeply appreciate the time and motivation you shared with me, even in your busy schedule.

I am grateful to the Almighty for His continuous protection throughout my life's journey. My appreciation goes out to my parents for their unwavering moral support from the moment I joined the University of Rwanda until now. I also want to extend my thanks to my classmates and colleagues in the Department of Computer and Software Engineering for their encouragement. Special appreciation goes to Dr. Eric Hitimana and Omar Sinayobye Janvier for their inspiring words during exhaustion.

Finally, I want to express my gratitude to my husband, Shingiro Faustin, for his enduring patience and prayers throughout this lengthy journey. I'm grateful to my children, Jacey, Jonny, Jaiden, and Jollissa, as well as my siblings. Your unwavering encouragement has been crucial in keeping me motivated, and I acknowledge that your support

Table of Contents

List of Figures

List of Tables

Abbreviations

ACF	Annual Correction Factor
AD	Anaerobic Digestion
GB	Aboveground Biomass
AI	Artificial intelligence
ANFIS	Adaptive Neuro-fuzzy Inference System
ANN	Artificial Neural Network
BOD	Biological Oxygen Demand
BSC	Bioenergy supply chain
CatBoost	Categorical Boosting
CC	Charge Controller
CH_4	Methane
CO_2	Carbon Dioxide
COD	Chemical Oxygen Demand
CSV	Comma-Separated Values
DC	Direct current
DC	Daily Consumption
DoD	Depth of Discharge
DSS	Decision Support Systems
DT	Decision Tree
EAC	Eastern African Countries
GA	Genetic Algorithms
GA	Forest Biomass
GDP	Gross Domestic Product
GHI	Global Horizontal Irradiation
H_2	Hydrogen
H_2S	Hydrogen Sulfide
HHV	Higher Heat Value
HRT	Hydraulic Retention Time
ICT	Information Communication Technology

IEA	International Energy Agency
IoT	Internet of Things
ISR	Substrate Ratio
JSON	JavaScript Object Notification
LightGBM	Light Gradient-Boosting Machine
MAE	Mean absolute error
MINFRA	The Ministry of Infrastructure
ML	Machine Learning
MPPT	Maximum Power Point Tracking
MQTT	Message Queue Telemetry
MSE	Mean Squared Error
NDBP	National Domestic Biogas Program
OLR	Organic Loading Rate
OLS	The Ordinary Least
OMSW	Organic Municipal Solid Waste
pH	Potential of Hydrogen
PLSR	Partial least Squares Regression
PSO	Particle Swarm Optimization
PV	Photovoltaic
$\mathbf{R^2}$	R-Squared
REG	Rwanda Energy Group
S-MCDA	Spatial Multicriteria Decision Analysis
SVM	Support Vector Machine
USD	United States Dollar
VFAs	Volatile Fatty Acids
WSN	Wireless Sensor Node
XML	Extensive Markup Language

Executive Summary

In the recent decade, global energy demand has increased in importance driven by industrialization, urbanization, and population growth. The government of Rwanda has been placing significant emphasis on adopting renewable energy sources as essential elements for sustainable energy strategies, to reduce reliance on fossil fuels. Bioenergy is a renewable energy source globally available with agricultural residues being a primary supplier to bioenergy production. This makes it a promising alternative energy source, as its characteristics ensure accessibility and affordability for the local community.

Recently, Rwanda like other Eastern African Countries (EAC) has promoted biogas as an alternative source of cooking energy through various initiatives, specifically for rural communities. Despite various governments' support policies, the adoption and diffusion of biogas technology have been considerably low. This research reported a lack of computing technology controlling the operating parameters and predicting biogas yield within biogas production supply chains as challenges for efficient biogas production.

This research aims to achieve three objectives, first to develop an Internet of Things (IoT)-based system for controlling biogas production and analyse the correlation between environmental data and biogas output. Second, to identify the most suitable power harvesting method to sustain the designed sensor nodes, and third, to design a machine learning (ML) model for predicting the biogas yield, and compare its performance against traditional models using the collected dataset.

The research was executed in three sequential phases. For the first stage, an IoT prototype for data acquisition was developed, and tested to collect real-time data. During this phase, the thresholds for environmental parameters were determined experimentally and the actuation mechanisms were set to regulate the optimum condition. Additionally, an assessment of environmental data correlation was made with statistics and regression model, the variables are correlated in such a way that gives insights, and variability in biogas production is explained by R Squared (R^2) 73.4% for the environment parameters explored, indicating a relatively good fit.

In the second stage, an analysis of the power harvesting approach was made to ensure the sustainability of the deployed sensor node, and the power harvesting system specification was

derived from the sensor node energy consumption calculation. Further, a mathematical model predicting solar panel size was derived from two simultaneous functions, and global horizontal irradiation (GHI) as experimental input data. As a result, the designed sensor node can be powered by a solar panel size varied from 17.8 cm^2 to 21.7 cm^2. The model was tested on values data and it is generic to be adopted anywhere.

In the third phase, the research focuses on improving biogas yield prediction using a hybrid ML approach. This approach integrates two data-driven models such as a light gradient-boosting machine (LightGBM) and categorical boosting (CatBoost) as based models. The hyperparameter turning using random search optimization with 5-fold cross-validation is adopted to avoid overfitting in each model. Further, the evolutionary strategy optimization technique is adopted to optimize the metal model. The hybrid model is applied to environmental data from biogas facilities, the model achieved superior performance with a mean squared error (RMSE) of 0.004 and mean absolute error (MAE) of 0.0024, surpassing k-nearest neighbor (KNN), random forest (RF), and decision tree (DT) models. The findings underscore the potential of accurate biogas yield prediction for optimizing energy production. This research demonstrates the contribution of emerging technology solutions as a way to meet growing global energy needs and advance sustainable biogas operations.

In summary, developing and integrating IoT and predictive modeling for biogas production monitoring and prediction aligns with the International Energy Agency's (IEA) Net Zero by 2050 resolution. First, optimization of biogas production can lead to a reduction in greenhouse gas emissions, as the captured methane can be used for energy generation instead of being released into the atmosphere. Moreover, integrating these technologies can have a broader impact on the waste management sector, as optimizing biogas production from organic waste can incentivize the diversion of waste from landfills, decreasing the overall emissions and environmental impact.

Chapter 1

Introduction

This chapter provides the background of the study, including the problem area, aims, and objectives. Furthermore, it examines the research contributions and the thesis structure for the rest of the chapters.

1.1 Introduction

In early 2021, the IEA unveiled the NZE2050 plan, a comprehensive long-term strategy to achieve net-zero CO_2 emissions across the global energy sector by 2050[1]. This pathway is pivotal for enhancing air quality ensuring universal access to energy and addressing the pressing need for sustainable energy solutions, such as biogas, within communities.

Biogas energy is a renewable energy source generated from the degradation of organic waste under anaerobic conditions, and it is generated via several stages including hydrolysis, acidogenesis, acetogenesis, and methanogenesis, resulting in a mixture of gases primarily composed of methane (CH_4) and carbon dioxide (CO_2) [2]. Biogas is derived from prominent chemical materials including carbohydrates, proteins, and rich fat organic materials originating from various sources such as food waste, cow manure, and human extraction [3]. In addition to feedstock characteristics, several operating parameters such as temperature, organic loading rate, and moisture level, contribute to the performance of the anaerobic digestion process and the overall biogas production.

Researchers on factors affecting the production of biogas in the anaerobic digestion (AD) process have increased in importance over the past years [3,4]. The feedstock characteristics, organic loading rate, hydraulic retention time, digester structures, and environmental parameters present high importance [6]. To optimize biogas production, it is crucial to manage certain factors within practical and achievable ranges, although some of these factors may be beyond control. Today with emerging technologies like the IoT and Artificial Intelligence (AI), it is feasible to improve the adaptability of measurements across various environments. Therefore, advancing technologies to control biogas systems will be essential for energy-efficient systems, and provide potential operational support to biogas beneficial.

This thesis outlines the development of three essential aspects of the research project: (1) an IoT-based system for monitoring and controlling biogas production, (2) a modeling approach for solar energy harvesting, and (3) a hybrid prediction model. Each component is thoroughly discussed in the following subsections.

1.1.1 Internet of Things-based Biogas Monitoring and Control System

In the recent decade, bioenergy has been a promising renewable energy source globally available with agricultural residues, animal, and food waste being the primary supplier of bioenergy production. Moreover, challenges related to the management of various biomass types, and process control within supply chains are gradually increasing. Thus, process monitoring and control help to explain the functional behavior of anaerobic digesters and to achieve efficient production.

The increasing interest in biogas production has driven the exploration of advanced technologies, including the integration of IoT to enhance monitoring and control processes. As demonstrated in [6], IoT-based systems are utilized to monitor real-time parameters such as temperature, pH, and gas composition, enabling operators to make data-driven decisions and optimize biogas production. However, the system lacks an activation mechanism to control key influencing parameters. Another study [7] examined the use of IoT in sewage wastewater treatment for monitoring biogas yield. While the gas volume monitoring platform was successfully implemented and validated, the reliance on a cloud computing paradigm in these systems may introduce challenges related to high data latency and bandwidth consumption. Similarly in [7], an IoT-based system is proposed to remotely measure and monitor various gas compositions such as methane, carbon dioxide, oxygen, and nitrogen in the biogas production environment. However, the proposed solution relies on energy-constrained communication protocol which may require extra power consumption for the low-power IoT device.

The prior research demonstrates the contributions of IoT in the biogas generation process. However, lack of activation mechanisms and integration with other emerging data analytics technologies to optimize the decision-making capabilities of IoT-based biogas systems. There is a critical need to advance research towards utilizing real-time data for precise predictions. However, the existing literature relies on the cloud computing paradigm which may result in bandwidth, and data latency limitations since biogas facilities often operate in remote locations where continuous internet supply is not assured.

This research aims to develop an IoT-based intelligent system to control the biogas energy generation process. To address the existing limitation, The proposed system relies on wireless hybrid sensor networks, actuation mechanisms, energy-efficient communication protocol, and edge-based computing paradigms.

1.1.2 Solar Energy Harvesting Model Approach

Biogas energy is often proposed in environments with limited access to electrical power. To ensure the continuous operation of IoT-based intelligent systems deployed in biogas production settings, exploring energy harvesting as an alternative power source is crucial. This approach can enhance the efficiency and extend the lifespan of IoT applications in these constrained environments.

This research will analyze the energy harvesting techniques suitable for powering the designed Hybrid Wireless Sensor Node (HWSN). The proliferation of IoT devices has led to increasing demand for efficient and reliable power sources to support their continuous operation [9]. Energy harvesting has emerged as a promising approach to address energy challenges, enabling IoT devices to generate power from ambient sources.

This research investigated power harvesting techniques applied in IoT applications, including kinetic, thermal, radio frequency, and solar. Kinetic energy harvesting involves converting mechanical energy, such as vibrations or motion, into electrical energy [10]. This approach is particularly suitable for IoT devices that are exposed to ambient vibrations or movements, such as wearable sensors or industrial machinery [11][12]. Alternatively, thermal energy harvesting utilizes temperature gradients to generate electricity [13]. This technique is advantageous for IoT devices in environments with stable temperature differences, such as in industrial settings or near heat sources [14]. Radiofrequency energy harvesting focuses on capturing and converting ambient electromagnetic radiation into usable electrical energy [15]. This method is particularly useful for IoT devices in densely connected environments, where RF signals are abundant [16]. In contrast, Solar energy harvesting involves converting sunlight into electrical energy and is also a widely adopted power harvesting technique for IoT applications [17]. This approach suits IoT devices exposed to sufficient sunlight, such as outdoor applications [18]. In this research context, solar energy harvesting is a viable approach to address IoT devices' energy constraints in a biogas production environment which is an outdoor application.

The prior research on solar energy harvesting adoption in IoT ecosystems focuses on analyzing power harvesting circuit design and energy harvesting computation applied in precision agriculture [19][20]. Another research proposed a novel solar energy harvesting for WSN nodes [21]. The research focus was to increase the overall harvesting system efficiency. The state of the arts focuses on the analysis of power harvesting solutions that are generic for IoT devices. However, the uniqueness of the embedded systems' power requirements must be considered. There is a need for an analysis of power harvesting focused on the power requirement of individual systems to avoid insufficient supply.

This study examines the solar energy required to ensure the long-term viability and scalability of the designed IoT system. This research takes advantage of sensor node energy consumption, and solar irradiation differences in various environments to develop a mathematical model predicting the solar panel size required, eventually resulting in experimental analysis of the expected energy harvesting.

1.1.3 Hybrid Prediction Model

The integration of ML techniques in the field of biogas production has gained significant attention in recent years, as researchers and industry professionals seek to enhance the optimization of this renewable energy source [22][23]. The application of ML in biogas production has led to several key features, including improved process control, and enhanced stability [24]. Leveraging the power of data-driven models, biogas beneficial can make more informed decisions, and optimize process parameters.

ML algorithms have been successfully applied to various aspects of biogas production, including feedstock selection, process optimization, and real-time monitoring. One of the most commonly used ML models in biogas production is the artificial neural network (ANN), which has demonstrated its ability to accurately predict biogas yield, methane content, and other key performance indicators [25]. Another widely adopted model is the support vector machine (SVM), which has been used for classification tasks, such as identifying optimal operating conditions [26].

Additionally, genetic algorithms (GA) have gained popularity in the field of biogas optimization, as they can effectively explore the complex and nonlinear relationships between process variables and the overall system performance [27].

The explored ML techniques have proven effective in learning from historical datasets to forecast outcomes, their application in biogas generation primarily relies on laboratory data for prediction and relies on a single model [28][29]. However, the applicability of these models may be limited to specific feedstock types or process conditions, and they may not fully capture the dynamic and nonlinear nature of the biogas production system [25] [10]. In addition, the lack of integration of ML models with other modeling approaches can limit the accuracy and robustness of biogas predictions required in modeling the complex biological processes involved in biogas production [30][31][32][33]. There is a need to integrate multiple modeling techniques to potentially enhance predictive accuracy and robustness.

This research explored the stacking ensemble modeling approach. The proposed model combines data-driven ML techniques with an optimization algorithm to predict the biogas yield. These hybrid models have the potential to provide more accurate and robust predictions of biogas production, as they can leverage the strengths of both modeling approaches. The simulations of the hybrid model were benchmarked against each model such as a decision tree, random forest, and k-nearest neighbors to find the highest accuracy system based on MAE, and RMSE. Furthermore, the fitness of the model was evaluated using R^2 metrics.

1.2 Motivation

The existing research studies proposed an IoT-based adaptive system for biogas management highlighting various IoT infrastructures as solution for biogas operatoring in the communities. The potential system should provide biodigester’s environmental parameters control mechanisms and alert the users to ensure the stability of the AD process and enable informed decision-making. Another consideration of the system will be energy harvesting, it will require a power harvesting mechanism to ensure the continuous operations of the system. As biogas energy plays a significant role in the renewable energy sector, an accurate predicting system for a biogas yield should be in place. However, it should not require high-cost equipment or inflexible infrastructure, as these would lead to significant expenses and hinder adaptability. In conclusion, this research should integrate IoT with AI technologies, such as ML techniques for promoting data analytics and forecasting biogas production to enhance the optimization and decision-making capabilities of IoT-based biogas management systems. The key selection of the Rwandan context is motivated by the national initiative to ensure that the network coverage is almost the same in all districts and sectors of the country, to allow system usability by the community.

1.3 Problem Statement

As the global community faces the challenges of climate change and energy security, finding sustainable ways to meet energy needs is critically important. Research indicates that worldwide fossil fuel reserves may only last for another 25 years, underscoring the urgent need to shift to alternative sources such as renewable energy [35]. Currently, global awareness of renewable energy adoption is a key pillar for economic growth, regulating environmental pollution, and decreasing dependence on fossil fuels [36]. In rural areas with restricted access to conventional energy sources, alternative energy solutions like biogas production show great potential. Utilizing bio-waste resources sustainably for biogas production is an effective way to address rural energy needs. Agricultural residues, food waste, and animal manure are substantial sources of organic material that can be transformed into biogas through anaerobic digestion [37].

Biogas is a renewable energy source derived from the anaerobic digestion of organic matter and holds immense potential in addressing the global energy crisis and environmental concerns. Despite this, biogas production is still inefficient. However inaccurate prediction of the input materials as well as inefficient control of environmental parameters and the fermentation process is reported as challenges. This research aims to investigate the best approach for designing and developing an IoT-Driven Predictive Control System for Biogas Management. The research outcome enables precise control of the fermentation process improving overall biogas production efficiency.

1.4 Research Questions

The main objectives of this research work are to contribute to renewable energy growth and access to clean cooking. This research intends to provide an IoT-Driven Predictive Control System for Biogas Management for controlling environmental parameters with biogas yield prediction capability. The research outcomes intend to address the following research questions:

1. How can IoT technology be integrated into the biogas production environment to control operational parameters and address the current needs and challenges biogas operators face?
2. How can an experimental analysis of power harvesting be conducted to identify the most suitable approach for powering IoT sensor nodes applied in a biogas generation context?
3. How can the existing ML models be enhanced to outperform the classical models applied in the biogas generation context?

The primary hypothesis of this research is to investigate whether the integration of IoT-based sensors with ML algorithms can optimize the anaerobic digestion process, leading to a significant increase in biogas yield and operational efficiency compared to traditional monitoring methods. This hypothesis will be examined through the design and development of an IoT-based monitoring system and ML algorithms designed to enhance biogas production and system performance.

1.5 Research Objectives

This research aims to design and develop an IoT-based intelligent system with an energy harvesting mechanism for controlling the effect of environmental parameters on biogas production. The development is feasible with the help of various technologies such as ML, cloud services, web technology, and IoT.

The specific objectives of the research are:

1. To develop an IoT-based system applied in a biogas context, and perform correlation between environmental data and biogas production.
2. To analyze the most appropriate power harvesting approach for sustaining the designed sensor node.
3. To develop an ML model for predicting biogas production and perform a comparative analysis with the classical models on the gathered dataset.

1.6 Research Methodology

This research employs a comprehensive, multi-faceted methodology to integrate IoT, energy-harvesting mathematical modeling, and ML approaches for enhancing biogas production. The proposed solution is designed to be relevant and cost-effective for biogas operators, addressing needs at both local and industrial levels.

The methodology encompasses qualitative, quantitative, and experimental approaches. The qualitative phase involved interviews and focus groups with stakeholders, including biogas operators, technicians, and industry experts, to understand their challenges and expectations. The data collected has been analyzed to identify themes and guide the IoT and ML solution design. The quantitative phase collected numerical data through surveys on production rates, raw material usage, and efficiency, which was analyzed using statistical methods to validate qualitative findings and refine the models. The experimental phase tested the IoT, ML, and energy harvesting models in selected sites. Sensors monitor production parameters, while ML algorithms analyze and predict biogas production. Energy harvesting mathematical models are used to optimize energy capture and utilization in the biogas production process. This comprehensive approach aims to develop an innovative tool for more efficient and sustainable biogas production. The detailed workflow is presented in Figure 1.1.

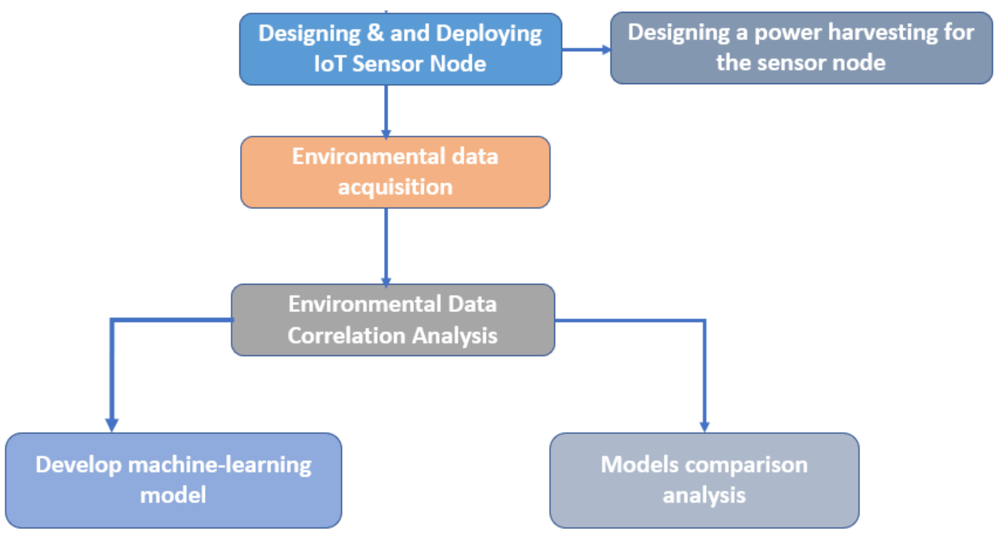


Figure 1.1: Research implementation flow

During the need assessment phase, the interview was conducted with key stakeholders gaining insight into the needs, challenges, and expectations of the integration of IoT and ML in biogas generation. The interview was conducted in four selected districts, namely Rwamagana, Musanze, Bugesera, and Gasabo as samples. The interviewers are classified into three categories (Rwanda Energy Group, technicians, and biodigester owner) based on their role and knowledge in the biogas industry. The interview questions were designed with the following considerations: the current biogas generation practices, perceived needs, and challenges, expectations on potential integration of IoT, and ML monitoring tools. Structured questions are chosen to ensure uniformity in responses and facilitate participants in the process. A sample size of a thousand interviewers was selected in the mentioned districts.

Figure 1.2 indicated that 47.7% of respondents cited feedstock supply and quality as a significant issue, making it the clear top challenge for biogas producers. According to 30.6% of respondents, monitoring and control systems are also a big concern, highlighting the need for efficient process management. Although less noticeable, equipment dependability and maintenance 23.0% and budget and financial constraints 16.7% are noteworthy.

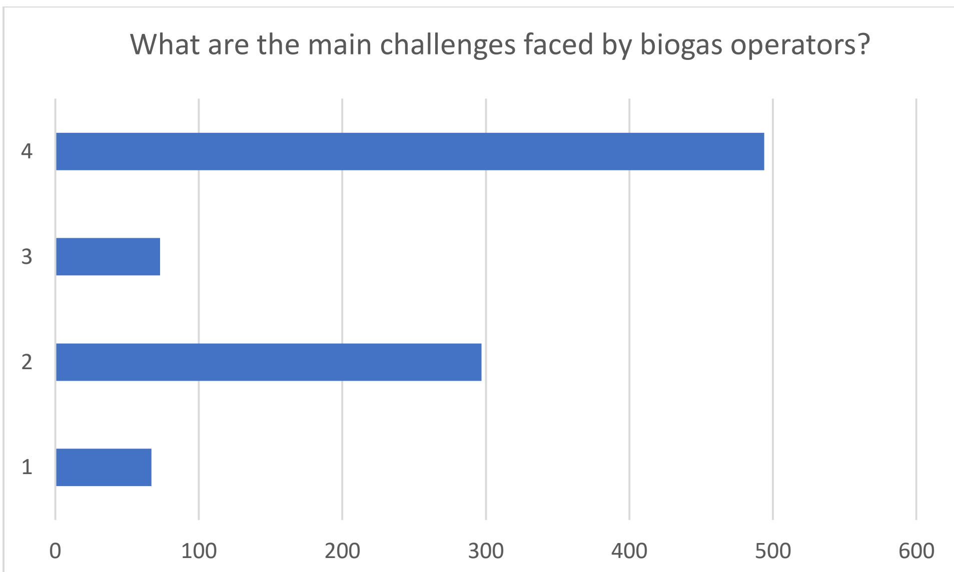


Figure 1.2: Results on main challenges for biogas operators.

Figure 1.3 presents the result indicating the most operational parameters affecting biogas production. The most important operational characteristic was found to be temperature, as indicated by 50.6% of respondents. A significant consensus regarding its impact was indicated by the volume of feeding, which came in at 35.7%. Also, 14.2% of respondents mentioned the composition of the feedstock, whilst 5.7% and 7.2% of respondents focused more on moisture level and pH, respectively, and 18.2% were unaware of the precise factors influencing the formation of biogas.

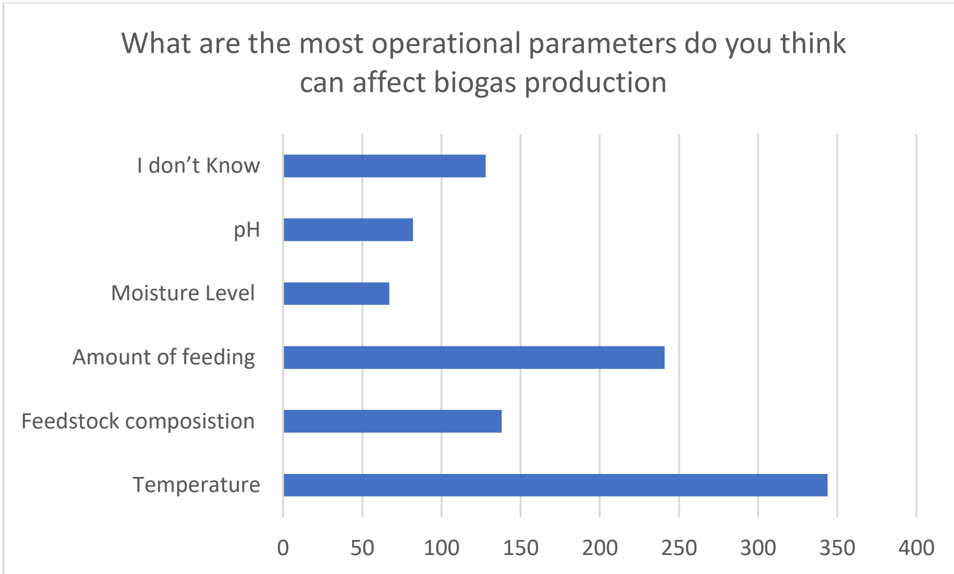


Figure 1. 3: Results for operating parameters affecting biogas production.

The category of responses for identifying technological tools used to monitor biogas production is depicted in Figure 1.4. It demonstrates that there is a monitoring tools access gap in the biogas generation industry. The 99% of respondents shown that they did not know any monitoring tools that are used to track the biogas production process. While 1% of respondents knew anything about monitoring technologies. This implies that more information about the various monitoring options has to be shared.

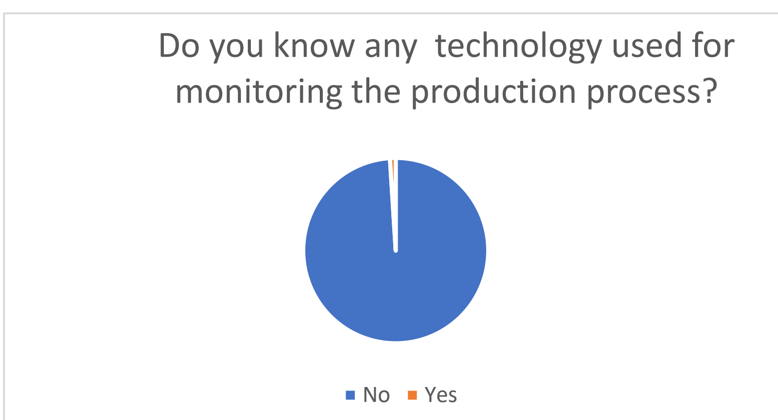


Figure 1.4: Results for technological tools used in monitoring.

Figure 1.5 indicates that the majority of respondents 81.8% strongly endorse the idea that real-time monitoring will likely increase the efficiency of biogas production. Nonetheless, a lower percentage of respondents 10.5% do not recognize the benefits of real-time monitoring, and there are still some uncertainties equivalent to 14.3%. Overall, while there is broad agreement on the potential benefits.

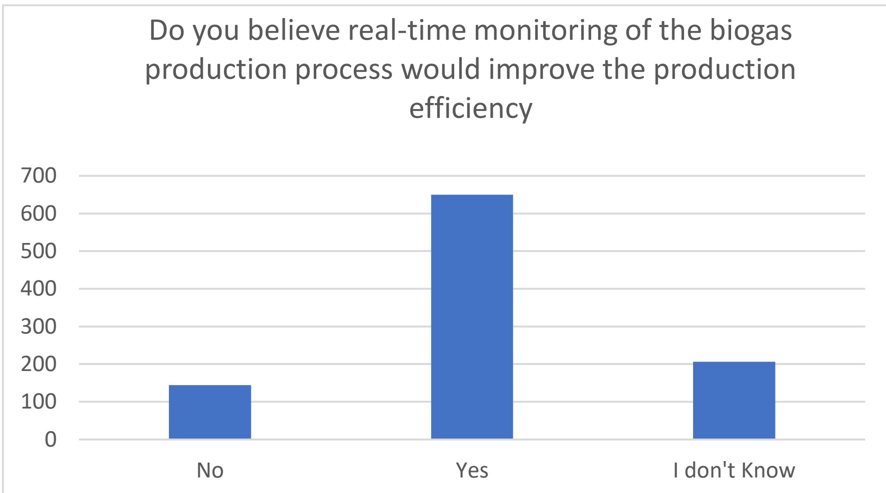


Figure 1.5: Results for the belief of improved production efficiency by real-time monitoring of the biogas production process.

The main outcome of the first objective is to develop a smart system capable of monitoring environmental parameters in a biogas generation context and analyzing their impact on biogas yield. The smart system's development involves integrating IoT technology to monitor and control environmental parameters. This integration addresses the needs and challenges identified through qualitative research with stakeholders, ensuring that the system is both relevant and practical. By deploying this system in real-world settings and continuously refining it based on feedback, the research demonstrates how IoT technology can be effectively utilized to optimize biogas production processes, thus directly answering the first research question. Data are acquired at a home digester for three months. Additionally, the correlation of environmental data on biogas yield was made using Pearson distribution and a multiple linear regression model.

To ensure the sustainability of the deployed sensor node, the second objective of this research is to analyze appropriate power harvesting to power the device. To achieve this objective sensor node energy consumption was computed, power harvesting system specification was analysed, and a polynomial base mathematical model was developed for predicting the expected solar panel size. By generating data on the efficiency and reliability of different power harvesting methods, the research provides concrete answers to the second research question, offering new knowledge on how to sustain power IoT sensors in biogas facilities.

The third objective aims to develop AI-based algorithms integrated with IoT-based architecture. The stack assemble-based model was developed and compared with an existing model from the collected dataset. This comparison is grounded in quantitative and experimental data, validating the superior performance of the new ML models. Consequently, the research answers the third question by demonstrating the practical improvement and benefits of using hybrid ML models in biogas production.

1.7 Major Contributions

During the development of this study, the following contributions have been achieved:

- This research contributes by developing an intelligent IoT system that monitors and controls environmental parameters in real-time within biogas production. By integrating IoT technology and addressing stakeholder needs through quantitative insights, the system optimizes biogas production processes effectively.
- The study provides new insights into energy harvesting methods for powering IoT sensor nodes in biogas facilities. Through rigorous experimental analysis, it identifies sustainable approaches, enhancing operational efficiency and reliability in biogas production environments.
- This research advances ML models tailored for biogas yield prediction. By leveraging advanced ML algorithms and comparing them with classical models using quantitative data, the study demonstrates improved prediction accuracy and efficiency, thus optimizing biogas production processes.

1.7 Thesis Outline

The thesis structure is illustrated in Figure 1.6. Chapter 2 presents the literature review on bioenergy, key influencing factors, and the existing technology-based mechanisms to monitor production. Chapter 3 details the effect of environmental parameters on biogas yield using an IoT system. Chapter 4 presents the experimental analysis of power harvesting required to sustain the sensor node. Chapter 5 discusses the biogas yield prediction modeling using a stacking assemble-based approach. Chapter 6 presents the thesis conclusion and recommendation.

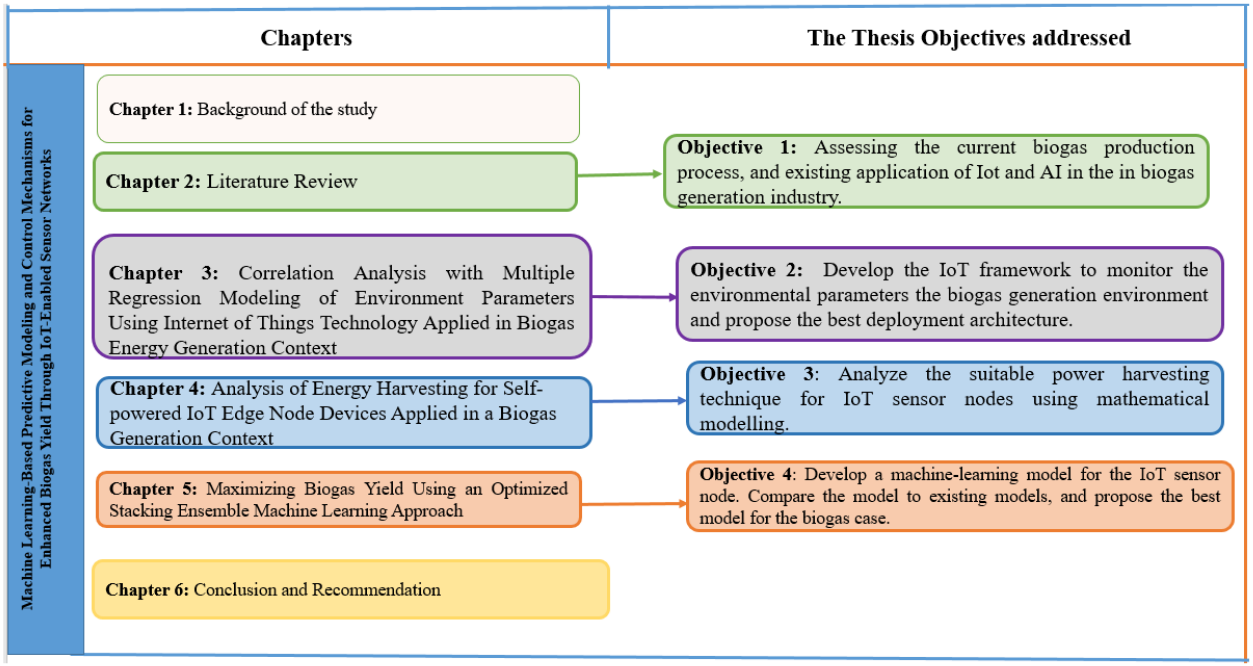


Figure 1.6: The Structure and organization of the thesis.

Chapter 2

Literature Review

2.1 Introduction

This section details the literature review of the current research field and proposes a way forward for this research. Firstly the current status and challenges of biogas energy adoption globally, in East Africa, and particularly in Rwanda. It also presents the details of the AD process and performance-influencing parameters. Additionally, the IoT-based architecture and its current application and limitations biogas generation context are presented. Lastly, the ML techniques and their application in biogas prediction are detailed.

2.2 The Current Status and Challenges for Biogas Production in Rwanda

In recent decades, the importance of global energy consumption has escalated, forced by industrialization, urbanization, and population growth. Consequently, there has been a significant emphasis on the adoption of renewable energy sources as integral to sustainable global energy strategies. It has been recognized that renewable energy sources have the potential to provide reliable and affordable energy while reducing greenhouse gas emissions and improving access to energy [34][35]. According to a global report in 2023, the renewable energy technology workforce has doubled from 7.3 million in 2012 to 13.7 million in 2022 [36].

Biogas energy holds a significant appeal among renewable energy sources due to its versatility and potential to address multiple challenges simultaneously. Biomass resources, such as agricultural residues, organic waste, and dedicated energy crops, can be readily converted into biofuels, or used directly for electricity and heat generation [37]. Moreover, bioenergy can help mitigate waste management, produce rich fertilizer for rich agriculture activities, and improve sanitation for the local community.

In the global context, the relevance of biogas has gained considerable attention in line with the United Nations' Sustainable Development Goals. Many countries have adopted renewable energy support policies, acknowledging the importance of providing clean and green energy [38]. Figure 2.1 presents, the global biogas energy production that has substantial increase, growing from 0.29 exajoules in 2000 to 1.46 exajoules by 2020, this growth has been particularly notable in the European Union, which accounts for nearly half of the world's biogas production.

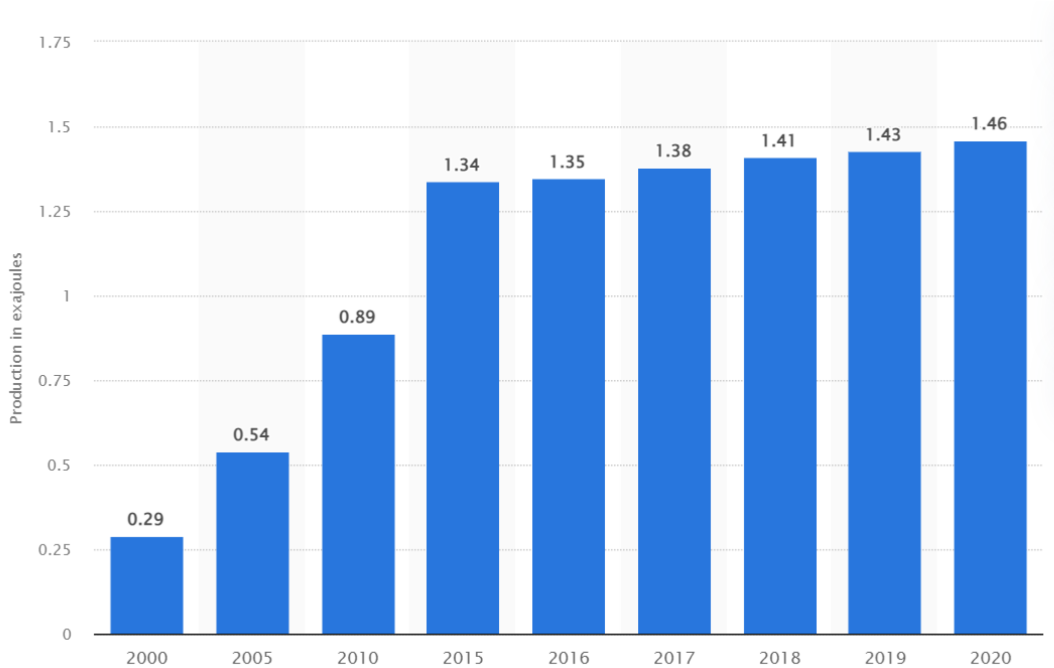


Figure 2.1: Global biogas production [39].

The East Africa Community (EAC) has made progress in promoting renewable energy, by setting ambitious targets and implementing favorable policies for the sector [40]. The renewable energy sector is core to East Africa's goal of expanding access to electricity and clean cooking, while also supporting entrepreneurship and the region's productive transformation. However, at the end of 2020, only 49% of the population had access to electricity, and a mere 14% had access to clean cooking [41]. In the East African region, the potential for biogas production and utilization has been recognized as a viable option to meet the energy needs of rural and urban communities. An Africa-focused global investment platform for the energy sector demonstrates that Kenya, a leader in biogas policies, has made the greatest progress in creating viable biodigester markets followed by Tanzania.

In recent years, Rwanda's energy efficiency roadmap has emphasized the importance of energy efficiency in meeting the country's energy goals [42]. There is much potential for Rwanda to rely on bioenergy using native renewable energy resources since its characteristics are naturally available in Rwanda's climate presented as subtropical with adorable temperatures to generate the energy. The efforts to promote clean energy have been initiated through various projects and initiatives [43]. Rwanda Domestic Biogas Programme (RDBP) has successful increased the number of biogas installations, as seen in Figure 2.2, along with several larger-scale biogas plants serving institutions and communities. The government has also introduced policies and regulations to support the biogas sector, such as the provision of subsidies and the establishment of quality standards for biogas systems.

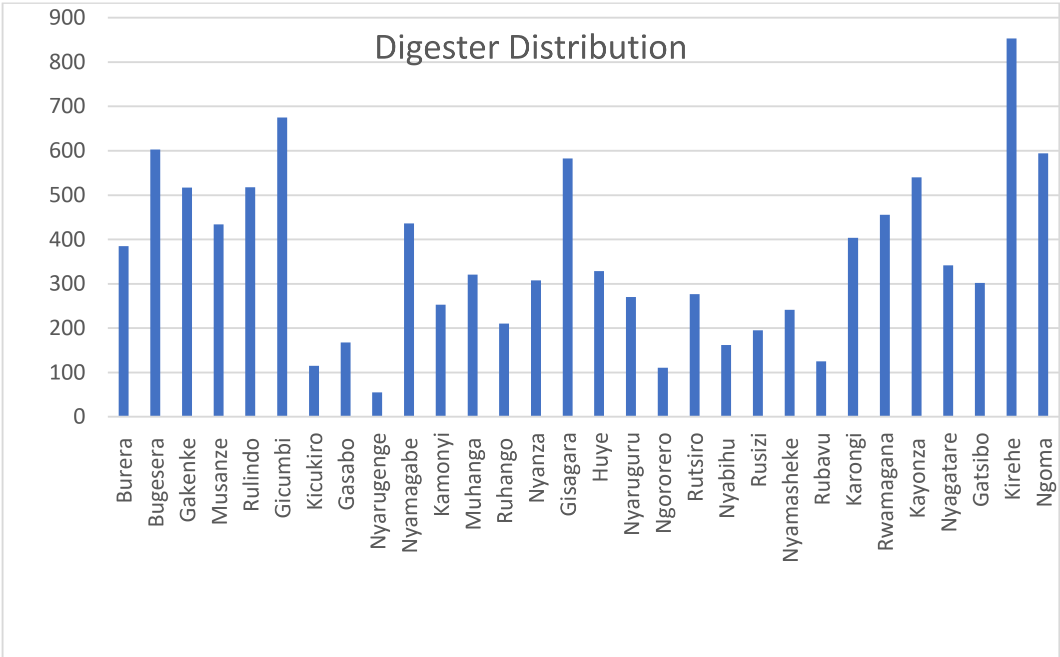


Figure 2.2: Digester distribution in Rwanda

Despite the abundant availability of organic waste resources, the utilization of biogas technology in East Africa remains limited. This is primarily due to a lack of technical knowledge, high installation and operation costs, and limited end-user applications.

Regarding Rwanda, during the needs assessment, this research discovered a high percentage of non-operational systems. The primary reasons for this include a lack of technical skills, inadequate feeding practices, limited feedstock availability, and the lack of end-user technology to control the production environment.

This section discussed the motivation behind the adoption of renewable energy, particularly bioenergy, the current development status, and challenges. The next section presents the bioenergy production process, details of its process parameters, and their roles in efficient production.

2.3 Biogas Energy Generation Process

The effectiveness of biogas production depends on various factors, including biogas plant structure, performance of the AD, and the process parameters. This section discusses the overview of the biogas production process from literature papers and site visits gathered information.

Biogas is the energy produced from an anaerobic digestion process, within the process biomass is converted to biogas through various biochemical sequential reactions. During biogas production, biomass is placed in a container called a digester. Literature has shown various digesters such as fixed domes, floating drums, and low-cost polyethylene tubes [44]. Figure 2.3 illustrates the structure and components of a single-stage low-cost polyethylene tube digester experiment. It was chosen due to its availability, and flexibility enough to adapt to extra inlets and outlets for small-scale production. In a single stage, wastes are loaded in the reactor and four processes happen in the reactor sequentially. The biodigester is both a storage and mixing room, with an upper layer reserved for accumulating biogas produced, the inlet, and outlets. Figure 2.4 presents the deployed digester in the experimental site.

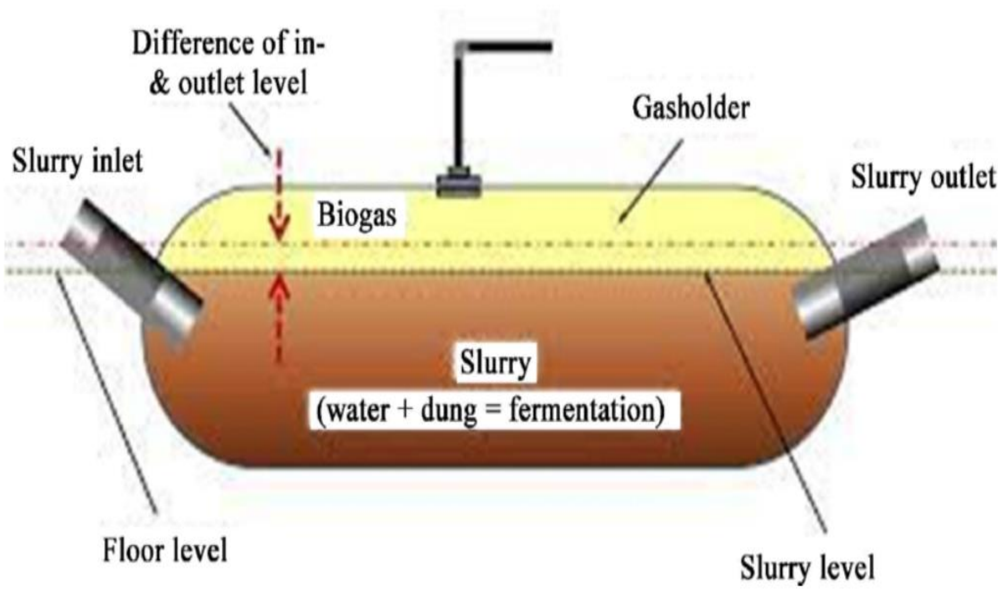


Figure 2.3: Biogas digester structure and components [45].

Figure 2.4: Experimented home digester

Anaerobic digestion involves a sequence of biochemical processes in which bacteria decompose the organic components of various substrates into a mix of gases including methane (CH_4), carbon dioxide (CO_2), hydrogen (H_2), hydrogen sulfide (H_2S), under anaerobic conditions [46]. Methane is the primary component and is valuable as a renewable energy source, often used for electricity generation and heating. The organic materials are mainly a composite of carbohydrates, proteins, and lipids which can be degraded to simpler compounds by microorganisms with the following process stages: hydrolysis, acidogenesis, acetogenesis, and methanogenesis stage [47] as presented in Figure 2.5.

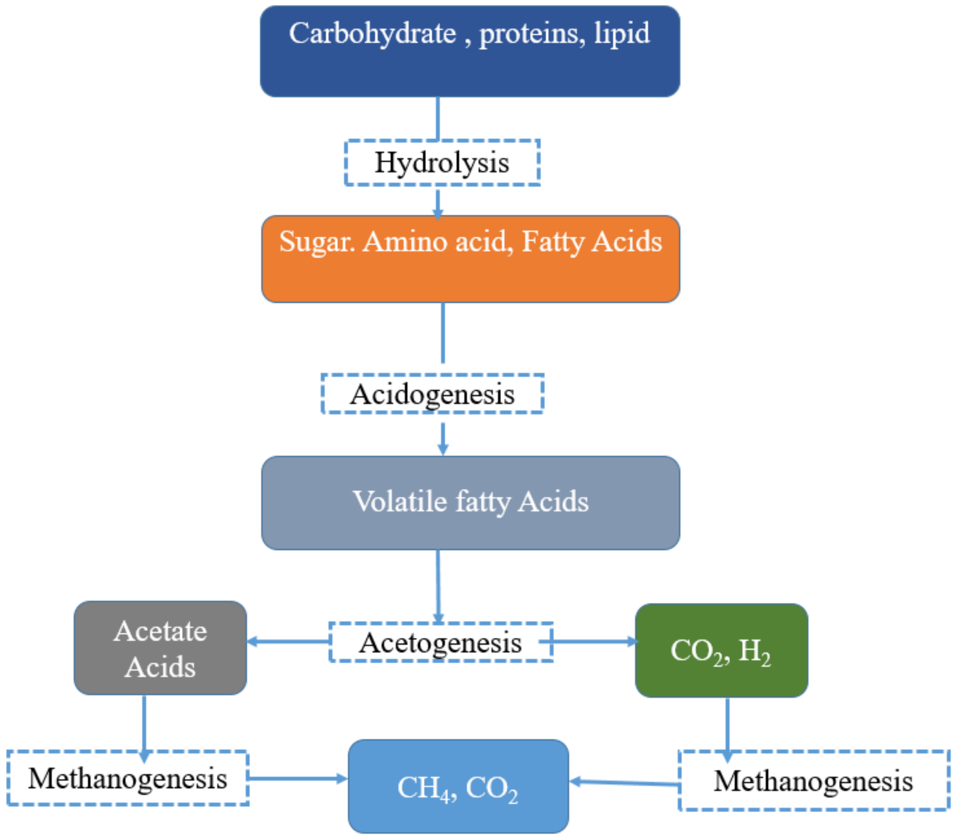


Figure 2.5: Biomass decomposition stages in the AD process

During the initial phase, microorganisms release extracellular enzymes that break down complex organic substances like long-chain polymers, proteins, lipids, and carbohydrates into simpler forms [48]. During the acidogenesis phase, many of the products from hydrolysis remain as large molecules that need to be broken down further into smaller components like acetic acid. In the subsequent stage, acidogenic bacteria metabolize simple sugars, amino acids, and fatty acids, producing acetate, and CO_2. Additionally, volatile fatty acids (VFAs) such as acetic acid, butyric acid, propionic acid, and other organic acids are generated during this phase [49]. In a perfect operating anaerobic system, approximately 70–80% of the hydrolysis products are transformed into hydrogen, CO_2, and acetate, that are readily utilized by methanogenic microorganisms. The 20–30% remaining is converted into various other intermediate products [49]. During the acetogenesis phase, agents can either produce or consume hydrogen to generate acetate. Those producing hydrogen oxidize acids into acetate [50]. In the last stage, methanogenic microorganisms utilize acetate, H_2, and CO_2 to synthesize methane [51].

2.4 Anaerobic Digestion Process Parameters

The optimum biogas production mainly depends on the movement and speed of AD bacteria. The bacteria are sensitive to several process parameters, impacting the effectiveness of anaerobic digestion. The efficient management of process parameters is essential to prevent reactor failure. The key factors include the substrate's chemical composition, mixing methods, inoculum-substrate ratio (ISR), biological oxygen demand (BOD), chemical oxygen demand (COD), organic loading rate (OLR), substrate mixing proportions, nutrient balance, hydraulic retention time (HRT), the pH level, and digester operating temperature [52].

These elements significantly influence anaerobic digestion performance and are extensively examined and discussed in subsequent sections, aiming to provide clarity and direction. Table 2.1 summarises the optimal operating conditions for a better output.

Table 2.1: Optimum range of parameters for a better digestion process

Parameters	Optumun values	References
pH	6-8.50	[53][54]
Temperature	Mesophilic (35-40^{o}C)	[55][56]
C/N ration	20-3-:1	[57]
ISR	>2:1	[58]
TA	1000-5000mg caco3/l	[48]
VFA/TA ration	0.20-0.40	[59]
OLR	1-6 Gvs OR	[60]
Moisture contents	>75%	[61]

To guarantee AD's performance, it is essential to maintain the operating factors within the desired range. This is practical and possible for some parameters while others are uncontrollable. In this research context, we focus only on controlling environmental parameters to optimize the digestion process.

2.4.1 Temperature

The anaerobic digestion decomposition rate is influenced by environmental factors. The operating temperature is the most important factor determining anaerobic digestion's performance since it boosts microorganisms' growth [62]. Thus impacting methane content, production rate, and overall system efficiency. Anaerobic digestion processes typically operate under three temperature ranges: cryophilic (15–25°C), mesophilic (35–40°C), and thermophilic (50–60°C) [63].

Studies have shown that biogas yield tends to be slightly higher under thermophilic anaerobic digestion conditions compared to mesophilic conditions [64]. Example in [65] Researchers have investigated the impact of temperature on the anaerobic digestion of lignocellulosic biomass, revealing that biogas production rates increase with rising temperatures. However, challenges such as operational instability may arise during thermophilic operation. Another research [55] explored the effect of temperature variation on biogas production from the digestion of rice straw and animal waste in bench-scale bioreactors. They found that all bioreactors adapted to both mesophilic and thermophilic conditions, however, mesophilic conditions offered higher stability thus presenting the most promising option for biogas generation. Further investigation of the co-digestion of poultry manure and kitchen waste to assess the impact of temperature and mixing ratio on biogas production under room temperature (28°C) and mesophilic conditions (37°C) [57]. The literature concludes that maximum biogas yield was achieved under mesophilic conditions.

2.4.2 Potential of Hydrogen

The potential of hydrogen (pH) value is key in assessing the stability and efficiency of the digestion process. The pH affects microbial activity and methane production yield. The pH value varies depending on the types of anaerobic microbes involved in the system. In [66], the analysis of the experiment conducted showed the effect of pH levels on CH_4 and H_2 production in a two-stage digester on sorghum-based substrates. The result shows the highest H2 yield of 0.92 mol H2/mol carbohydrates consumed at a pH equal to 5. According to the research in [67], the preferred pH range for anaerobic digestion is between 6.8 and 7.2. However, they highlight that the most suitable pH for methanogenesis falls around 7.0, while for hydrolysis and acidogenesis, it ranges between 5.5 and 6.5. This is the reason researchers advise separating the hydrolysis/acidification and acetogenesis/methanogenesis processes in two-stage systems.

2.4.3 Moisture Content

Moisture content stands out as a crucial factor influencing anaerobic digestion. Research has indicated that moisture has the benefit of regulating cell movement facilitating the transport of nutrients, and participating in the hydrolysis of complex organic matter [68]. Higher moisture levels support anaerobic digestion because water content can significantly influence the process.

Anaerobic digestion processes are categorized based on the total solids (TS) content of the slurry in the digester reactor. These categories include low solids or wet digestion (less than 10% TS), medium solids or semi-dry (10–20% TS), and high solids (more than 20% TS). Most studies on degrading the organic fraction of municipal solid waste have focused on dry anaerobic digestion processes due to the high solid content of organic municipal solid waste (OMSW) [69][70]. However, increasing the moisture content of OMSW by adding water or co-digesting it with low-solid wastes such as sewage sludge and manure can render it suitable for semi-dry anaerobic digestion processes. Research has shown that elevating the initial moisture content in mesophilic anaerobic digesters from 90% to 96% resulted in increased methanogenic activity in high-solids sludge digestion [71]. A separate investigation [69], indicated that digesters with higher initial moisture content exhibited higher methane production rates. However, a study stated that raising the moisture content of organic OMSW decreased the methane production rate in anaerobic digesters employing periodic cycles of leachate drainage and water addition [72].

2.5 Traditional Technologies for Monitoring Anaerobic Digestion

Bioenergy is a promising renewable energy source globally available with agricultural residues being a primary supplier to bioenergy production. Moreover, challenges related to the management of various biomass types, and process control within supply chains are gradually increasing. Thus process monitoring and control can help to explain the functional behaviour of anaerobic digesters and to achieve efficient biogas production. This systematic review examines the traditional methods and technologies employed for monitoring and optimizing the key parameters of the anaerobic digestion process. The performance of the anaerobic digestion process is highly dependent on the careful monitoring and optimization of various process parameters, such as pH, temperature, organic loading rate, hydraulic retention time, and nutrient balance.

Numerous traditional methods have been employed to monitor and optimize these parameters. For example, temperature is a critical parameter that can significantly impact the activity of the anaerobic microorganisms and the overall efficiency of the process. Various temperature measurement techniques, such as thermocouples and resistance temperature detectors, have been used to monitor and control the temperature within the reactor [81, 82].

Similarly, pH is another crucial parameter that must be carefully maintained within an optimal range to support the growth and activity of the anaerobic microorganisms. Traditional pH measurement methods, such as the use of pH electrodes, have been widely utilized to monitor and adjust the pH of the digestion process [75].

In addition to these physical parameters, the monitoring and optimization of the biochemical parameters, such as volatile fatty acids (VFAs), and nutrient levels, are also essential for the successful operation of an anaerobic digestion system. Traditional monitoring tools, such as titration, spectrophotometry, and gas chromatography, were adopted to measure these biochemical parameters and inform the decision-making process for process optimization [84[77].

The traditional biogas monitoring approaches provide valuable insights into the fermentation process by monitoring key parameters and making informed decisions to optimize the system. However, those approaches often rely on manual data collection and intermittent measurements, which can lead to incomplete and inaccurate data. Moreover, traditional monitoring approaches can be labor-intensive and time-consuming, requiring dedicated personnel for data collection and analysis. This can be particularly challenging for remote or decentralized biogas plants, where on-site monitoring may be logistically difficult or cost-prohibitive.

To address traditional monitoring limitations, this research has explored the potential of emerging technologies, such as the IoT and cloud platform to enhance biogas monitoring capabilities. The combination of those technologies can provide real-time, continuous data on the biogas production process, enabling more precise control and optimization. The next section highlights the recent potential of AI and the IoT to address bioenergy challenges and foster a sustainable future.

2.6 IoT Concept and Architecture

The concept of the IoT refers to an interconnected system of Internet-enabled devices [78]. These devices are designed to carry out three primary functions: receiving data, processing data, and transmitting data [79]. The function of IoT is to enhance daily practices by presenting significant real-time monitoring capabilities. The IoT ecosystem involves various complementary technologies to bridge the gap between virtual and physical spaces.

IoT technology has been increasingly applied in biogas production systems to enable data collection, remote control of the production environment, and energy management. The key IoT Architectures in biogas systems comprises three, four, and five layers. However, the three-layer architecture is the most widely adopted [80][81]. Figure 2.6 presents a generic IoT architecture adopted in biogas generation systems comprised of three layers.

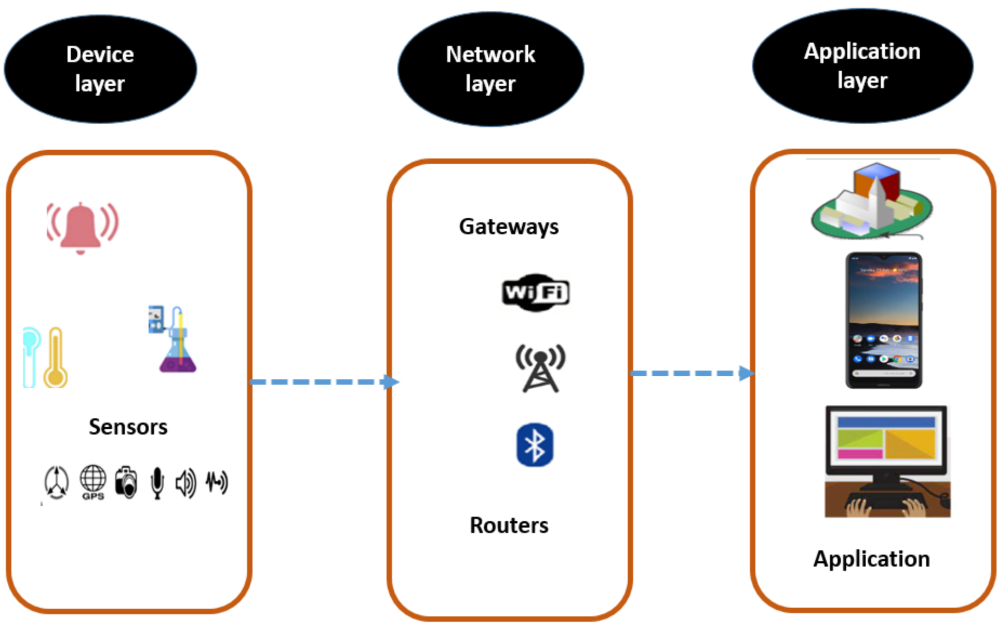


Figure 2.6: Generic IoT architecture

Device layer: Involves sensors and actuators, and the microcontroller data processing unit that collects data from the environment, and makes the necessary processing.

Network layer: Middleware comprises routers, gateways, and switches that facilitate the interaction between the lower and upper layers.

Application layers: comprises of web or mobile app hosted on a cloud platform responsible for data storage, processing, and analysis tasks of the acquired device data. To provide specific services to end users.

In addition to IoT architecture, various communication protocols are employed to establish connections to the upper layers such as message queuing telemetry transport (MQTT), constrained application protocol (CoAP), HyperText Transfer Protocol Secure (HTTPS), and Long Range Wide Area Network (LoRaWAN)[82]. Those Communication protocols play a critical role in ensuring efficient data transmission and reliable system performance. This review examines the various IoT communication protocols applied in biogas systems, highlighting their characteristics, advantages, and limitations. The choice depends on specific application requirements such as range, power consumption, data rate, and implementation complexity.

MQTT: It is one of the key communication protocols powering the IoT ecosystem. The MQTT is designed for low-power and low-bandwidth devices. It has gained significant traction in IoT applications due to its lightweight nature and suitability for resource-constrained devices [83]. In the biogas industry, an MQTT-based system for real-time monitoring and control was integrated with existing supervisory control and data acquisition systems to improve operational efficiency in urban biogas facilities in Europe.

CoAP: In low-power communication systems, CoAP is particularly recommended for applications where energy efficiency, low bandwidth, and limited computational resources are crucial, making it an attractive choice for various IoT applications [84].

HTTP: In the context of the IoT, HTTP plays a crucial role in facilitating communication and data exchange between IoT devices, cloud platforms, and various web-based applications and services [85]. IoT systems often rely on HTTP to establish a common language for data exchange. IoT devices can utilize HTTP to send sensor data to cloud platforms, where it can be aggregated, analyzed, and acted upon [86].

LoRaWAN: It has emerged as a promising solution for IoT applications in the agricultural sector due to its long-range communication capabilities, and low power consumption. LoRa-based IoT systems have been utilized in a variety of agricultural applications, such as smart irrigation, soil moisture monitoring, rice field management, and the provision of intelligent agricultural services [87].

In this context, the IoT architecture adopts the MQTT communication protocol due to its lightweight nature, which ensures efficient data transmission with minimal bandwidth and power consumption, which makes it suitable for real-time monitoring and control in resource-constrained environments.

2.7 Wireless Sensor Networks (WSN) for Biogas System

WSNs have emerged as vital tools in providing real-time data collection, analysis, and control over various parameters. These networks are typically classified based on their deployment environment into underground, territorial (aboveground), and hybrid systems [88]. Underground WSNs are deployed beneath the surface to monitor critical parameters like soil moisture, temperature, and gas emissions. Recent studies have highlighted the importance of underground WSNs in applications where precise subsurface monitoring is critical to optimizing biogas production [89]. In contrast territorial WSNs, are deployed aboveground and are typically used to monitor parameters such as air quality, temperature, and methane levels in open spaces. These networks benefit from line-of-sight communication, which reduces signal interference and enhances data transmission reliability [90]. Further Hybrid WSNs combine the strengths of both underground and territorial networks, offering a more comprehensive monitoring solution for both subsurface and surface parameters simultaneously [91]. This research adopts a hybrid WSN due to its high capability to support biogas generation systems where indoor and outdoor conditions are important.

2.8 IoT Applications in Biogas Generation Context

The rapid advancements in the IoT have paved the way for a wide range of applications in various domains, including energy management [93]. In recent years, the energy sector has witnessed a remarkable transformation with the advent of IoT technology [94]. Smart grids and green IoT have been highlighted as key areas where IoT drives significant improvements, and smart grids, in particular, offer immense potential for optimizing energy consumption through demand-management applications [95]. Concerning biogas energy, the growing interest in biogas production has led to the exploration of various technological advancements, including the integration of the IoT to enhance the monitoring and control of biogas production processes [96].

Recently, numerous biogas automated applications aimed at inducing users by sending direct responses to remote locations. For example in [97], IoT-based systems have been utilized to monitor real-time parameters, such as temperature, pH, and gas composition, enabling operators to make informed decisions and optimize the biogas production process. Another example is the research that investigated the application of IoT for sewage wastewater treatment processes to monitor the biogas yield in real-time [98].

The application of IoT in biogas production can be observed in various aspects. IoT-based sensors can be utilized to collect data on key parameters such as temperature, pH, and organic matter content, allowing for precise monitoring and control of the anaerobic digestion process [99][100]. Additionally, IoT-enabled devices can regulate the input of feedstock, water, and other necessary components, optimizing the overall efficiency of the biogas plant [101]. Furthermore, in [102], research introduced a biogas facility integrated with IoT technology, aiming to showcase innovative designs. The proposed prototype incorporated pH, temperature, and oxidation-reduction potential electrodes linked to a programmable logic controller. The system enabled real-time remote monitoring which is essential for maintaining reactor stability and ensuring quality biogas production. Furthermore, in [102], research introduced a biogas facility integrated with IoT technology, aiming to showcase innovative designs. The proposed prototype incorporated pH, temperature, and oxidation-reduction potential electrodes linked to a programmable logic controller. The system enabled real-time remote monitoring which is essential for maintaining reactor stability and ensuring quality biogas production.

The integration of IoT technology in the biogas production industry has gained significant attention in recent years, as it offers promising solutions to enhance productivity, monitoring, and control of biogas plants. However, the successful implementation of an IoT system requires addressing the following limitations: (1) the reliable sensors and consistency of the data transmitted by the IoT sensors to the cloud platform; (2) inaccurate data can lead to suboptimal decision-making and ultimately undermine the effectiveness of the monitoring system; and (3) the integration of IoT technology with the existing biogas production infrastructures is technically complex and costly, posing a barrier to scalability. Biogas is mostly adopted as an alternative energy source in an electricity-contained environment, the consistent power supply of the sensor node, requires time-consuming battery charging.

In the existing IoT-based biogas monitoring systems, the real-time data are transmitted wirelessly to a central control system, allowing for remote monitoring and providing valuable insights into the factors that influence biogas production. However, the following research gaps have been identified: (1) the existing IoT-based biogas monitoring systems often lack advanced analytics and decision-support capabilities. (2) While these systems can provide real-time data, the ability to translate this data into actionable insights and optimize biogas production remains a key gap. (3) there remains a deficiency in community-based applications equipped with validated datasets to facilitate the AI modeling process.

Integrating the IoT into biogas production in the Rwanda context presents significant potential impacts on both sustainability and scalability. From a sustainability perspective, IoT enhances operational efficiency by enabling real-time monitoring and control of key production parameters, and maximizes biogas yield, contributing to more efficient feedstock utilization. On the scalability front, IoT supports the expansion of biogas production by enabling centralized, remote management of multiple sites, which simplifies the supervision and coordination of large-scale operations.

2.9 Machine Learning in Biogas Generation Context

In the modern world, AI technology mimics human intelligence through the use of a computer. Numerous criteria exist for classifying AI due to field complexity applied. AI is categorized into four distinct types: symbolic AI, ML, heuristics, hybrids, and others [103], and it has been adopted in many sectors, such as agriculture, industry, education, health, and energy generation [104], [105].

ML, a branch of AI, enables machines to learn and enhance themselves autonomously using historical data, without requiring explicit programming [106]. This approach enables automatic learning by computers, operating independently of human intervention, with a primary focus on developing self-learning algorithms through continuous data inputs. An accurate ML model increases the predictive performance for the future, and it is achieved through investigating various learning algorithm models and selecting the best-suited for the problem [107]. The learning process starts with observing the patterns within the database based on the provided input, and it enhances its ability to make decisions by identifying these patterns in the data. Figure 9 illustrates the process involved in supervised and unsupervised learning.

ML is broadly classified into several categories based on the learning approach and the nature of the problems being addressed. The major classifications include:

- Supervised Learning: It trains using a known dataset to develop the learning algorithm and tests it with the testing dataset, comparing the results with the correct or intended data to assess the accuracy [108]. This process enables modifications and adjustments to the model if the results are unsatisfactory. Supervised learning is divided into two main sub-categories, classification, where the model learns to predict discrete class labels, and regression, where the model learns to predict continuous output values [108].
- Unsupervised Learning: In this approach, the model is not provided with labeled data, and it tries to uncover hidden patterns, structures, and relationships within the input data. Unsupervised Learning techniques include clustering, dimensionality reduction, and anomaly detection [109]. This learning process lacks an accuracy check, but it can label data based on identified patterns and derive its solutions.

- Reinforcement Learning: This involves an agent interacting with an environment, taking actions, and receiving rewards or penalties based on the outcomes. The agent then learns to optimize its behavior to maximize the cumulative reward [110]. Reinforcement learning enables machines and software agents to identify the optimal behavior to optimize their performance based on a given objective.

- Semi-Supervised Learning: This is a combination of supervised and unsupervised learning, and the model is trained on a dataset that includes both labeled and unlabeled data [111]. The key idea is to leverage the unlabeled data to enhance the performance of the model, as the unlabeled data can provide valuable insights into the underlying structure of the data distribution.

The application of ML techniques in the biogas industry is a rapidly evolving field, with promising results across various aspects of the process based on their learning style and data structures [112]. Supervised and unsupervised learning algorithms have demonstrated their utility in tasks such as classification, cluster analysis, and process optimization [108], [109]. For instance, ML models can be trained to predict biogas yields, identify optimal operating conditions, and detect anomalies in real-time, enabling proactive maintenance and improved resource management [22]. The prior classify ML in into three categories such as production, detection, and energy utilization.

In the digestion process, a successful prediction model for the downdraft gasification with a water gas shift unit to generate biohydrogen in the gasification was proposed using an ANN model [101]. However, there is a lack of comparison of alternative models to validate the improved accuracy.

In biomass quality measurement, researchers explored the chemical exergy of torrefied biomass compared to raw biomass [28], [113]. Their investigations revealed that the SVM method exhibited notably superior performance in estimating parameters such as HHV and energy recovery. The ANN models tend to yield more consistent and trustworthy prediction accuracy compared to the other models explored. However, several other ML techniques were not considered in the evaluation process.

For controlling the production quality, researchers employed partial least squares regression (PLSR), weighted regression, and an ANN model to forecast the biogas flow rate, using Carbon (C), Hydrogen (H), and Oxygen (O) as input-independent parameters are presented [29]. It included the test on various forecasting algorithms, but the method is only for the detection of flow rate without considering the major parameters affecting the biogas yield.

Similarly in [114], a study developed RF models to predict the biogas yield of bioenergy crops using image datasets acquired from drones. The random forest model demonstrated flexibility and effective scalability for large datasets. However, there has been a lack of studies comparing it with other ML methods, highlighting a deficiency for biogas yield forecasting systems specifically for environmental parameters data.

Prior research performed predictions based on single models, demonstrating their district dominance. However, the lack of integration of ML models with other modeling approaches, such as process-based models or hybrid models, which could potentially improve the accuracy and robustness of biogas prediction required for the complexity of biological processes involved in biogas production, which can be difficult to capture fully using a single ML model.

This study explores the use of hybrid modeling approaches, which combine data-driven ML techniques with an optimization algorithm strategy that incorporates the underlying biological processes. These hybrid models have the potential to provide more accurate and robust predictions of biogas production, as they can leverage the strengths of both modeling approaches.

2.10 Hybrid Modeling Approach

Modeling complex real-world phenomena often requires a balance between the flexibility of data-driven approaches and the interpretability of theory-driven models. In recent years, the field of ML has witnessed a growing interest in hybrid modeling strategies that combine the strengths of these two paradigms, offering the potential to achieve more accurate predictions. This review aims to provide a comprehensive overview of the state-of-the-art in hybrid modeling for ML-based prediction, highlighting the key concepts, and methodologies.

- Physics-informed models: One prominent hybrid modeling strategy is the integration of physics-informed models with data-driven ML techniques, known as physics-informed ML [107]. As explained in [115], this approach is particularly useful in scientific problems involving complex processes that are not fully understood. The integration of physical knowledge can be achieved through various means, such as data and feature engineering, and the development of hybrid architectures that leverage both physical and data-driven components.

- Assemble model: In hybrid modeling, the assembled model is another aspect of a combination of multiple ML algorithms to capitalize on their complementary strengths [116]. This ensemble method can lead to improved prediction performance by leveraging the unique capabilities of different models. This approach has been applied in a wide range of domains, including environmental modeling, energy systems, and predictive maintenance [117]. Figure 2.6 depicts the integration of multiple base models, each trained on a different subset of the data or using a different algorithm. The individual predictions from these base models are then combined, often through a weighted voting system or a meta-model, to generate the final ensemble prediction.

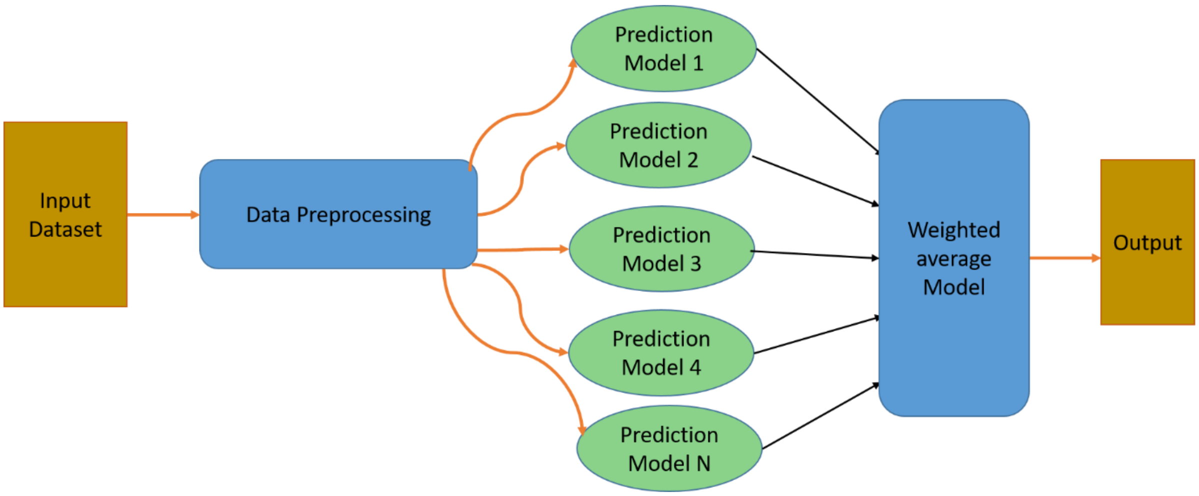


Figure 2.6: Typical ensemble model architecture

The literature review reveals a wide range of ensemble techniques, including bagging, boosting, and stacking [118]. These techniques differ in their underlying mechanisms and the way they combine the component models as follows:

- Bagging: Among the ensemble techniques, bagging involves training multiple models on random subsets of the data and aggregating their predictions to improve overall predictive accuracy [119]. Bagging has been widely used in various applications due to its simplicity, robustness, and ability to reduce the variance of individual models [120]. However, in the presence of very small or imbalanced datasets, the random sampling process might not be effective, leading to suboptimal performance and potential overfitting on the resampled data.

- Boosting algorithms: In ensemble modeling, boosting has emerged as a widely adopted approach, offering a powerful way to enhance the predictive capabilities of weak learners by iteratively training a sequence of base models, each of which focuses on the instances that were misclassified by the previous models [121]. However, one significant limitation is the potential for overfitting, the boosting model becomes too specialized to the training data and fails to generalize well to new, unseen instances.

- The stack-based model: It is an ML architecture that utilizes a series of interconnected layers, each with its specialized function, to process and analyze data. the stack-based approach is primarily concerned with the accuracy of predictions [122].

The ensemble models have demonstrated impressive results for improving prediction accuracy. However, the overfitting limitations are identified. Effective model selection and ensemble optimization techniques are needed to overcome the identified limitation. This study proposes a hybrid model that stacks data-driven boosting models (LightGBM, CatBoost) and an optimization algorithm (evolutionary strategy) to improve to accuracy of predicting biogas yield. LightGBM excels in handling large datasets with high efficiency and speed, while CatBoost provides robust performance with numerical data, which is often prevalent in biogas production datasets. The integration of these models ensures a comprehensive analysis of various data types. In an evolutionary strategy, the model can iteratively optimize and adapt the tuning of its parameters, leading to better convergence on optimal solutions. This hybrid approach improves the precision of biogas yield.

2.11. Conclusion

This chapter provided a detailed analysis of the current state and challenges of biogas production in Rwanda, emphasizing the importance of biogas as a renewable energy source. It outlined the biogas generation process, focusing on the anaerobic digestion process and key parameters that influence efficiency. Traditional monitoring technologies were reviewed, revealing their limitations in optimizing biogas production, thereby setting the stage for the introduction of modern technological solutions.

It explored the role of IoT and ML in improving biogas production. It discussed the architecture of IoT and the application of wireless sensor networks for real-time monitoring and remote management of biogas systems. The potential of ML to analyze complex data and improve operational efficiency was highlighted, leading to the introduction of a hybrid modeling approach to optimize biogas yield. This sets the foundation for understanding how these technologies can drive sustainability and scalability in Rwanda's biogas sector.

Chapter 3

Correlation Analysis with Multiple Regression Modeling of Environment Parameters Using Internet of Things Technology Applied in Biogas Energy Generation Context

Recently, the significant demand for biogas energy has increased. However, biogas operators lack automated and intelligent mechanisms to produce optimization. The IoT and ML have become key enablers for the real-time monitoring of biogas production environments. This research aims to develop an IoT-based architecture to gather environmental parameters for biogas generation. In addition, data analysis was performed to assess the effect of environmental parameters on biogas production. The edge-based computing architecture was designed comprising sensors, microcontrollers, actuators, and data acquired for the cloud Mongo database via MQTT protocol. Data were captured at a home digester on a time-series basis for 30 days. Further, Pearson distribution and multiple linear regression models were explored to evaluate environmental parameter effects on biogas production. The constructed regression model was evaluated using R^2 metrics, the result found 73.4% of the variability of biogas yield given the environmental parameters experimented. From a correlation perspective, the experimental result shows a strong correlation of biogas production with an indoor temperature of 0.78 and a pH of 0.6. On the other hand, outdoor temperature presented a moderated correlation of 0.4. This implies that the model had a relatively good fit and could effectively predict the biogas production process.

3.1 Introduction

The use of renewable energy is expanding globally due to resource availability and fluctuating energy prices, with efforts to mitigate the effects of climate change [123]. By 2015, its usage accounted for nearly 22% of the total energy consumed worldwide [124, 125]. Developed nations are advancing their use of renewable energy; for example, renewable energy sources are anticipated to produce adequate electricity in several United States over the following two decades [126]. Further, Africa as a continent presents the highest potential to be the first continent to base a major amount of its industrial and economic growth on clean and renewable energy sources [127, 128]. Unlikely other renewable energies, biogas is a promising solution since its characteristics are available and affordable to the local community. Biogas is a renewable gaseous fuel that is generated through the breakdown of organic materials without the presence of oxygen in a process called anaerobic digestion [129]. Domestic biogas is made from animal excrement from cow, or pig dung, coupled with food waste, agricultural waste, and occasionally human excreta. Biogas's major ingredients are CH_4 and CO_2, representing 50–60% and 35–45%, respectively [130]. Rwanda's government has invested extreme efforts over recent years to encourage the adoption of biogas usage through various initiatives, including the construction of biogas digesters for local communities and the Girinka project (One Cow per Poor Family) [131], which has gradually increased cattle dung which is a major source of biogas in Rwanda.

Despite various governments' support policies, these studies disclosed that the adoption and diffusion of biogas technology have been considerably low [132, 133]. This is not isolated to Rwanda; the lower adoption of domestic biogas technology has been recognized globally [134, 135]. These issues are not limited to technical challenges such as a lack of classification requirements, as well as insufficient raw materials, and a lack of precise technology for controlling the operating parameter. Several works of research on the parameters affecting the production of biogas in the AD process have increased in importance over the past few years. Feedstock characteristics, digester structures, continuous processing, operating, and environmental conditions present importance in biogas production [136].

Concerning environmental parameters like temperature, pH, moisture content, and humidity presented a high impact during biogas generation [137, 138]. For example, Toutian et al. [139] discussed a lab-based experiment on the effect of temperature on biogas production during the hydrolysis stage. It presented the efficiency of production in the mesophilic environment [140]. Abudi et al. [141] detailed the contribution of pH for optimum biogas generation. Optimal biogas production was obtained at a pH range between 6.8 and 7.4 [142]. Furthermore, the moisture content in the substrate allowed for the free and relaxing movement of microorganisms, resulting in high biogas production [143]. Therefore, to ensure the consistent and effective generation of biogas, it is essential to maintain proper continuous control over the environmental parameters. This research aimed to assess the impact of environmental parameters on biogas production using a combination of IoT technology and ML techniques.

In recent years, the global renewable energy robot market has been predicted to reach the United States dollar (USD) 75.82 Billion by 2030, with a growth rate of 27.9% during 2022–2030 [144]. Recent researchers have demonstrated the impressive contributions of AI and the IoT on renewable energy economies and how these can be implemented in the entire process from energy generation to transmission and use [145–147]. The most recent IoT trends indicate the potential application of the IoT in energy, including energy utilization monitoring in smart cities, solar plant health monitoring, and more [148, 149].

This has been commonly applied in other industries, where IoT technology has been adapted for designing and optimizing the anaerobic digestion process [150]. For example, research on the IoT performance of anaerobic digestion was performed, where an operating condition monitoring tool was developed; however, this article lacked intelligent control mechanisms to regulate the optimum condition [151]. In [152], technologies covering the IoT architecture, and data analytics modeling, were leveraged to explore the existing works in biogas supply chain management. It experienced limitations when implementing the proposed architecture in a specific case study. Additionally, the authors of [153] proposed an IoT-based biogas measurement monitoring system that could classify various gases. However, the authors did not consider monitoring environmental conditions and behaviors that could affect production.

Furthermore, in [154], the security mechanism for IoT applications in biogas generation was proposed focusing on cyber-physical systems. Among these related works, the works in [155-157] investigated the integration of the IoT and data analytics models approach for anaerobic digestion performance. The main emphasis of these two publications was primarily on the algorithms used for analysis. Overall, based on the state-of-the-art IoT in AD automation, IoT technology has already been integrated by proposing various applications to support people. However, there is still a lack of applications in the community with the available validated dataset to support the AI modeling process.

The main contribution of this thesis is to design and develop an IoT-based architecture for gathering data, monitoring, and controlling operating conditions in the biogas generation process while addressing some of these existing limitations. This section proposed the application of data validation algorithms to avail datasets for further prediction purposes and support the production control of biogas in the Rwanda context. The use of multi-linear regression and Pearson distribution models to perform a correlation analysis of biogas production by considering multivariant environmental parameters in the biogas generation context is proposed. These models were validated using data gathered by the developing IoT-based architecture.

The proposed IoT-based architecture comprises sensors to acquire the biogas digester's environmental parameters, such as indoor temperature, ambient temperature, humidity, pH, and moisture content. The edge-centric IoT-based architecture proposed here excels by enabling real-time data processing at the production site. This allows for immediate analysis of parameters, leading to quick adjustments and optimized yield [158]. Unlike cloud-centric models, which face latency and bandwidth issues from data transmission, edge-centric systems reduce delays by processing data locally [159]. This enhances system responsiveness and decreases dependence on constant internet connectivity, which is crucial for maintaining productivity in remote or challenging environments. In other words, Edge-centric IoT architecture outperforms Device-to-Devices (D2D), hybrid, and cloud-centric models by offering real-time processing with minimal latency and reduced dependence on internet connectivity [160]. Unlike D2D's scalability issues, hybrid's complexity, and cloud's latency, edge-centric systems enable immediate adjustments on-site, optimizing biogas yield efficiently and effectively.

Figure 3.1 presents the adopted edge-centric architecture with three main components: (1) the edge layer with mounted sensors, microcontrollers, and actuators; (2) the network layer made by an IoT Wi-Fi gateway; and (3) the cloud layer was made of a web-based platform with persistent data management hosted on the cloud server to acquire data from sensing devices.

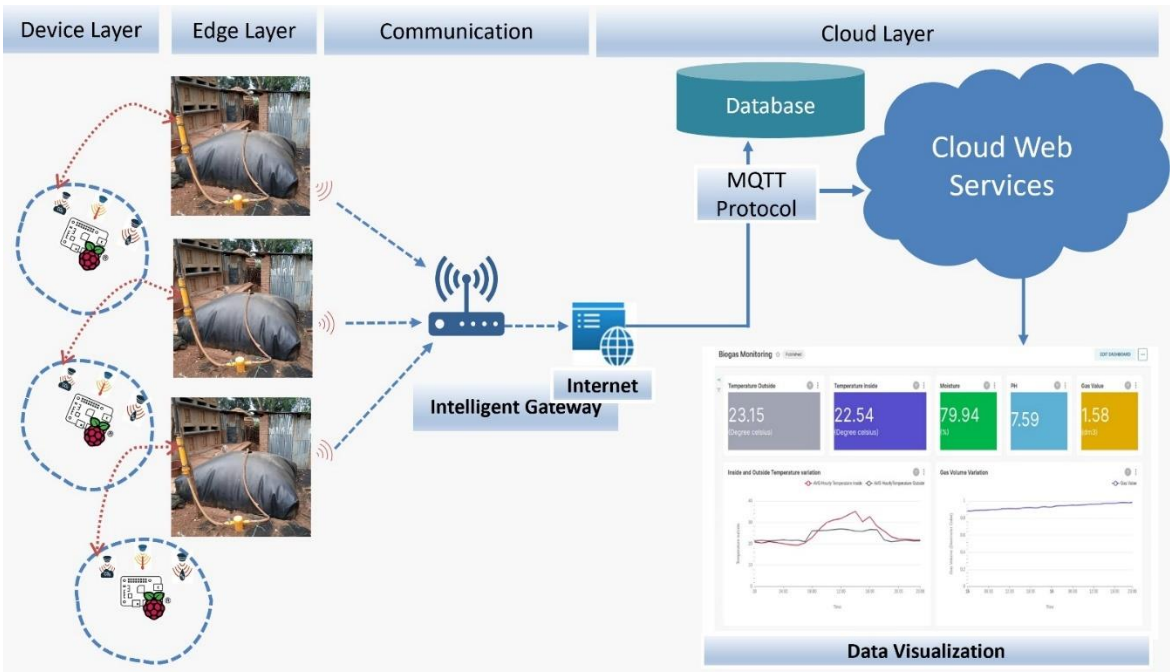


Figure 3.1: The Proposed IoT-based architecture applied in the biogas generation context.

The proposed system helps to increase biogas production by controlling environmental parameters affecting biogas production.

3.2 Materials and Methods

This section detailed the materials and methods adopted for implementing the IoT-based architecture, data preprocessing, and data analysis used in the study. Figure 3.2 describes the flow of the methodology used to validate the proposed architecture. From a material perspective, the IoT-based architecture was developed and deployed on the physical digesters to collect time series data. From a method perspective, a series of data pre-processing such as missing values, high peak values records removal, and datatype conversion was performed, and further data validation analysis was conducted and validated.

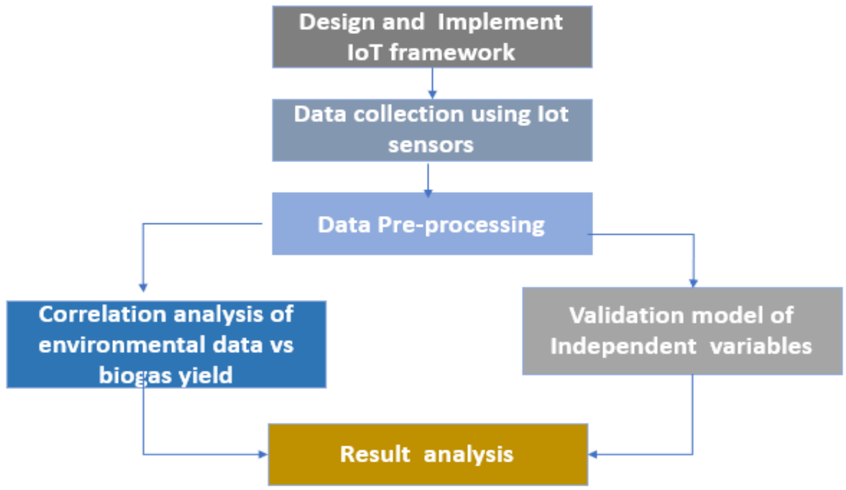


Figure 3.2: Research study methodology

3.2.1 Interned of Things-based Architecture Design and System Setup

During this research, a single-stage low-cost polyethylene tube digester of 4000 liters was experimented with, and cow manure and home wastes were considered as input materials. The research was conducted in the Eastern province of Rwanda specifically in the Rwamagana district. The reason for choosing the indicated area is that it is recognized as a hub for agriculture and animal husbandry [161] Therefore, it is a promising supply of biogas from crop residues and animal manure. Furthermore, the selected district experiences a high average temperature, which is a crucial factor in biogas production [162].

This research employs IoT techniques to create a smart digester control system. The proposed IoT-based architecture comprises of ground-based nodes mounted on a digester to periodically collect related environmental data. It integrates a microcontroller, sensor elements, and actuators, and sends notifications via a web platform to relevant stakeholders. Furthermore, the data is stored in a Mongo database, which is later used to display events on a dashboard, for future decision-making. The process diagram of the proposed architecture follows in Figure 3.3.

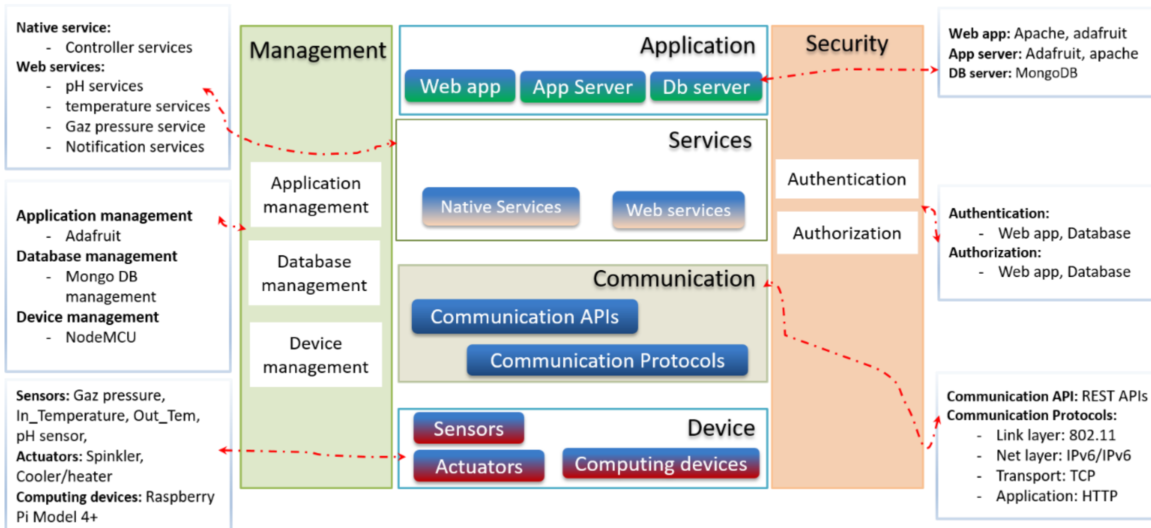


Figure 3.3: The process diagram of the proposed architecture

The system development consists of six components, as outlined in the IoT architecture structure illustrated in Figure 3.3. These components include the application layer, management layer, services, communication, security, and device layer. Table 3.1 presents a detailed role of three layers during the system implementation.

Table 3.1: Description of the proposed architecture

Component	Description	Criteria	Justification
Edger layer	- Comprises sensors, actuators, and microcontrollers. - Perform local data analysis for controlling actuators. - Ensure data security through authentication	- Real-Time Data Processing - Data Security - Compatibility	- **Real-time data processing**: Performs local data analysis and controls actuators to optimize biogas yield immediately. - **Data security**: Ensures data integrity and security through authentication methods. - **Compatibility**: Must be compatible with various

			sensors and actuators for accurate measurements and control.
Network layer	- Comprises Wi-Fi module. - Perform data routing and transmission	- Data Routing and Transmission. - Range and Reliability. - Bandwidth	- **Data Routing and Transmission**: Efficiently route data from the edge layer to the cloud. - **Range and Reliability**: Wi-Fi should provide reliable connectivity over the required range. - **Bandwidth**: Sufficient bandwidth to handle data transmission needs without delays.
Cloud layer	- Allow permanent data storage, Allow public data access - High data performance analysis	- Data Storage Capacity. - Performance Analysis. - Access Control	- **Data Storage Capacity**: Provides ample space for storing large amounts of historical data. - **Performance Analysis**: Enables high-performance data analysis and visualization. - **Access Control**: Allows controlled public access to data while protecting sensitive information.

The detailed system design and configuration of the edge and cloud node are discussed in the next subsections.

3.2.1.1 IoT Kit Design

The designed sensor kit comprises various sensors to acquire data such as moisture content, pH level, pressure, and temperature respectively. Sensing devices are connected to a customized Raspberry Pi 3 B+ microcontroller, with a built-in Wi-Fi module, used as an IoT gateway. The acquired data from the sensors can easily be sent to the database. Each node is connected to a solar panel power supply. Table 3.2 describes all the devices required to design the kit.

Table 3.2: Components of designed sensor kit.

Device	Description	Basis for Selection	Comparison with Alternatives
DS18b20	Indoor Temperature, humidity	Low power consumption; accuracy in temperature sensing	Compared to alternatives like TMP36, DS18b20 is more accurate and operates at lower power in active mode.
OAT-M-24	Ambient Temperature	Extremely low power; suitable for long-term monitoring	Alternative sensors may consume more power and may not offer the same low transition period.
700KPGPN	Gas pressure	Moderate power use; good sensitivity for	Compared to BMP180, it offers better power efficiency
DIY pH	pH	Low-cost; sufficient accuracy for pH monitoring in biogas systems	commercial pH sensors, it is cost-effective but may require more

			frequent calibration.
Capacitive Moisture	Moisture	Low power consumption; durable in soil environments	Compared to resistive sensors, it is more durable and has better power efficiency
Raspberry Pi .3	Microcontroller	Balance between performance and power consumption; broad community support	Compared to RPi 4, it consumes less power.

Figure 3.4 presents the IoT kit design setup, developed to connect, and control the IoT sensing devices, and Figure 3.5 illustrates the real sensor node kit development in the test bed stage.

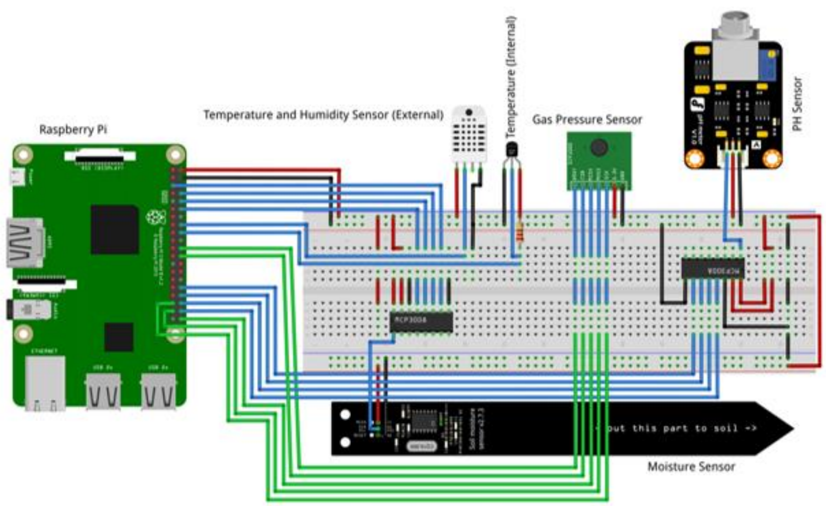


Figure 3.4: IoT Kit Design

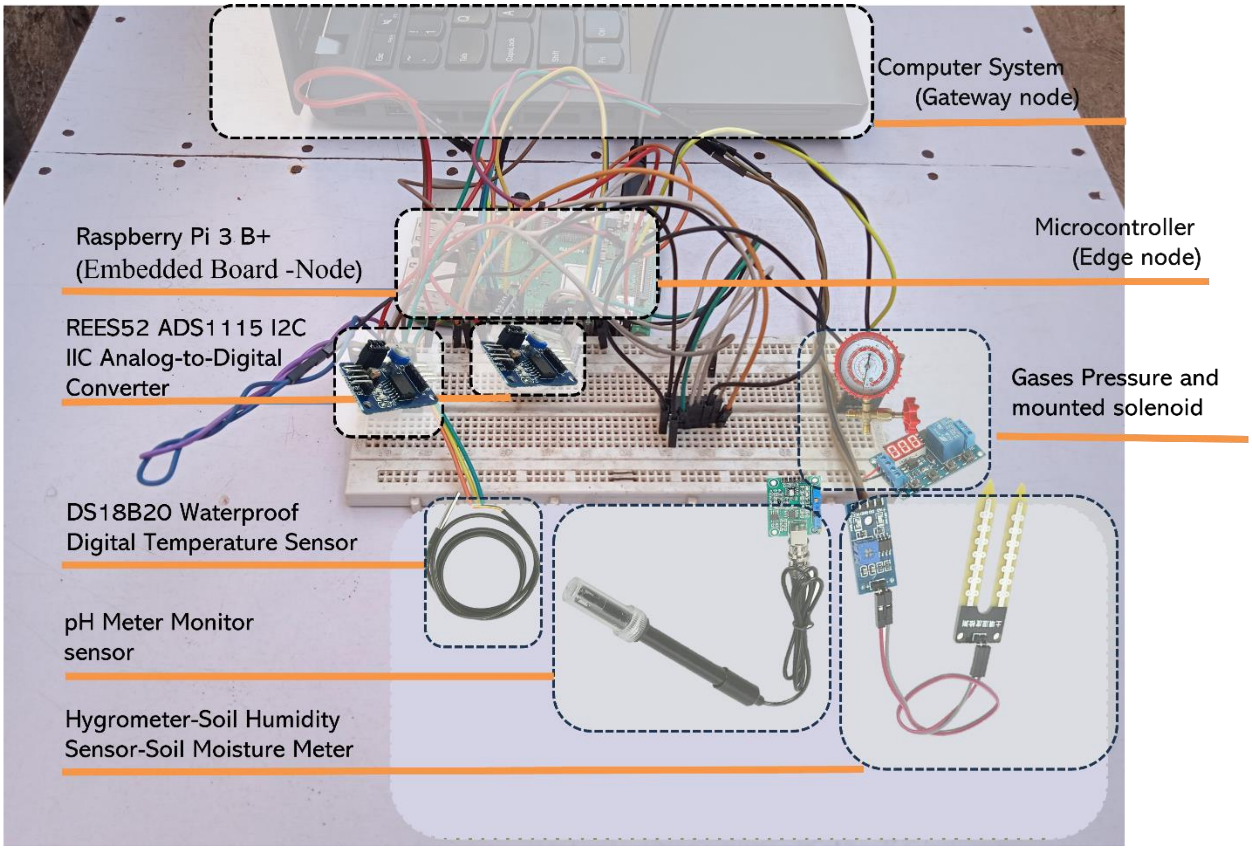


Figure 3.5: Detailed mounted sensor node with related components

3.2.1.2 The Actuator's Design

The designed IoT system is intended to provide an activation mechanism. A set of activation tools are developed to provide a stable environment. Figure 3.6 presents the programmed actuation logic of the system. It presents a set of actuator devices such as sprinklers, thermostatics, and thermal electric actuators proposed.

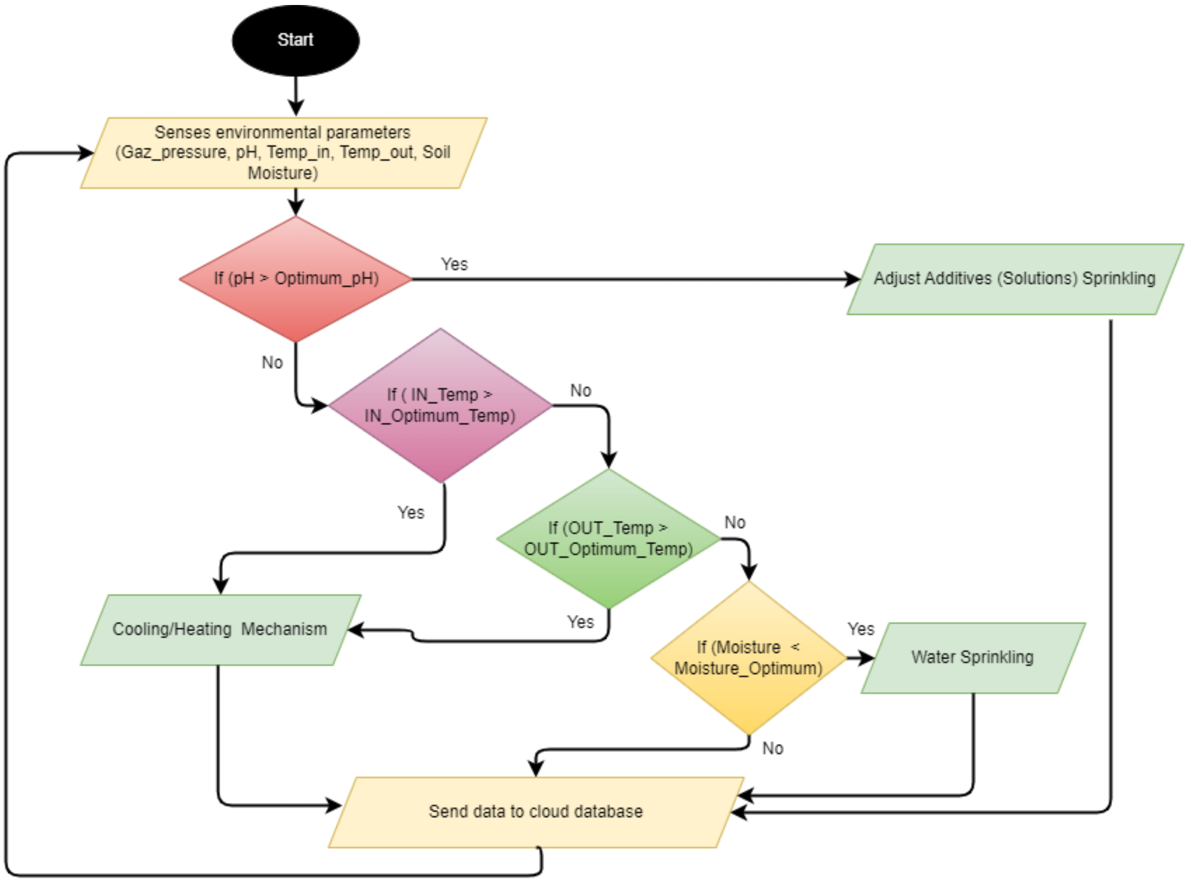


Figure 3.6: Actuation control logic

In the programmed logic, the environmental parameter threshold is defined, when the acquired data from the environment are out of the threshold, the microcontroller sends a notification signal to a specific actuator to take the necessary action. Table 3.3 presents a detailed description of the actuators explored in this study.

Table 3.3: The actuators Implemented in Smart biodigester.

Components	**Description**
Sprinkler	Discharge water when the effect of low moisture is detected.
A thermostatic	Provide heating in the environment when the temperature is below the threshold.
Thermal electric	Provide cooling in the environment when the temperature is above the threshold.

3.2.2 Deployment Architecture of the Designed Wireless Sensor Nodes

Figure 3.7 illustrates the deployment design architecture of the proposed biogas digester monitoring system. The system components are shown including Edge, Gateway, and Cloud components.

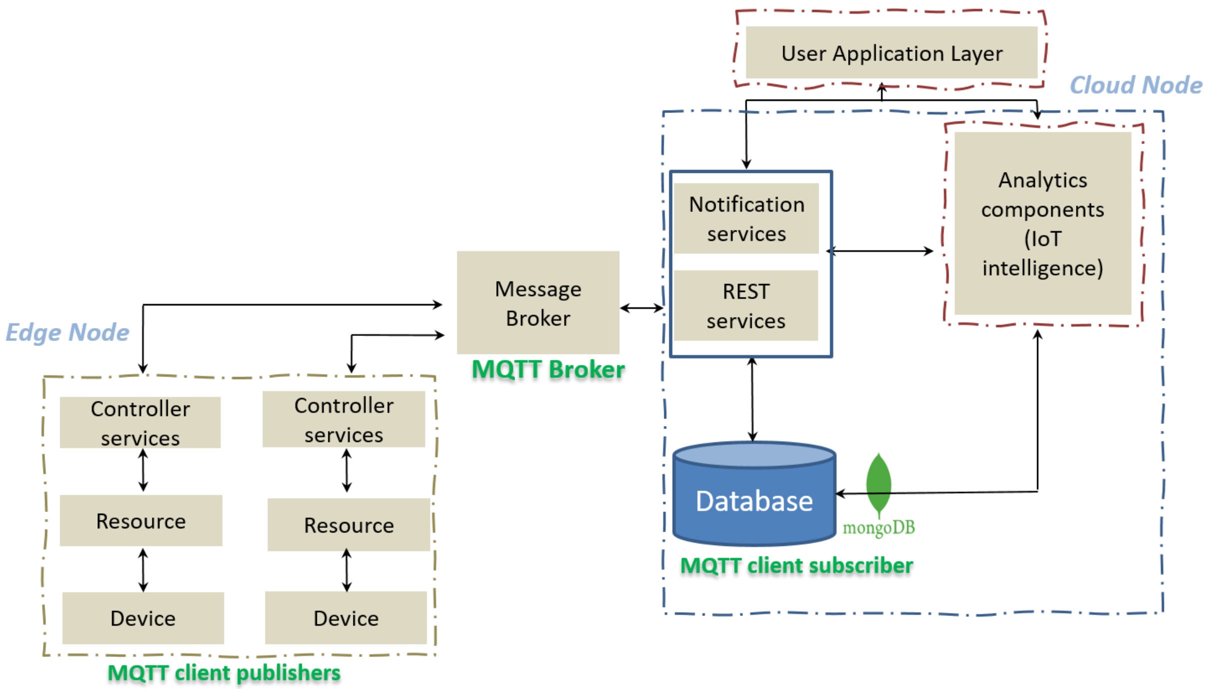


Figure 3.7: Deployment design of IoT-based digester monitoring system

In the system, the sensors will collect environmental data within a digester. This is recorded in a persistent database (MongoDB) and data is displayed in the dashboard. The system is intended to notify the biogas operators when there are anomalies.

MQTT Protocol

The MQTT is a communication protocol adapted at the application layer. It acts as a client-server publish/subscribe messaging protocol designed for machine-to-machine communication in a low bandwidth environment [163]. It has been adopted in many IoT-based applications such as manufacturing process management and healthcare [164, 165], energy generation and trading [166], and agriculture environment monitoring [167]. Developing an IoT system involves IoT data transmission with IoT communication protocols. MQTT presents a remarkable contribution to IoT applications due to low power consumption, small bandwidth, and less memory, commonly needed in IoT systems [168]. The sensor node publishes data to the Mongo database. The data is sent to the MQTT broker followed by the MQTT client subscriber as shown in Figure 3.8.

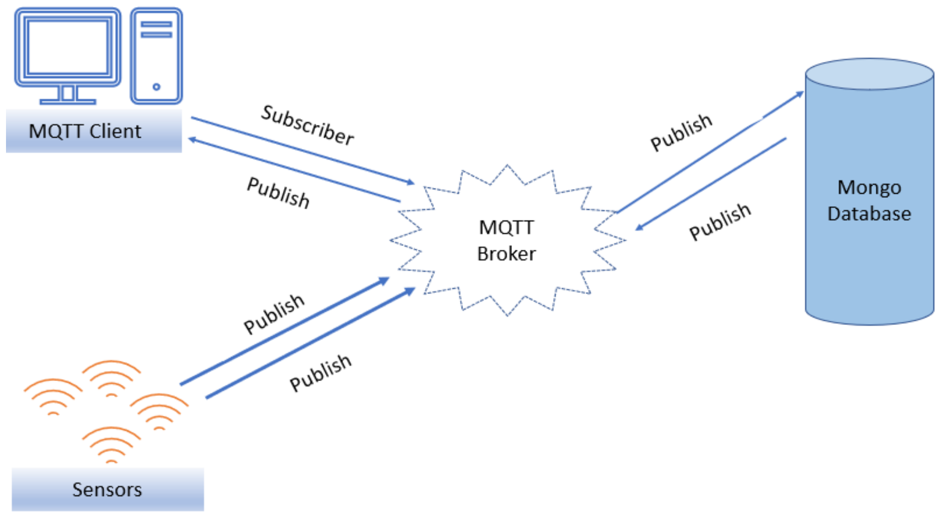


Figure 3.8: MQTT communication pathway

Data Storage

The sensor data is pushed to the MongoDB database hosted on the cloud server. MongoDB is an open source used to store semi-structured data (NoSQL). It is a database that saves and retrieves documents in either JavaScript Object Notification (JSON) or extensive Markup Language (XML). Mongo is recommended to be adapted to big data management due to its characteristics [169]. In this research, MongoDB was chosen due to its capability to store and process data in real-time. The application with MongoDB has key characteristics such as scalability and high processing speed [170] supporting our objectives.

Cloud Web Platform

The web application component is a major part of the developed IoT system. It enables intended users to view and understand the status of the biogas station in real-time. ReactJS Programming language was adapted to develop the front end, while the back-end web services are developed using the Laravel PHP framework. The web application is deployed on the cloud server. Table 3.4 presents summarized functionalities of the application.

Table 3.4: Web application capabilities

Non-function	Functional
Scalable design enables fast response	Dashboard for data visualization
Security via user authentication	Instant Notification
Responsiveness across multiple devices	Data Analytics

3.2.3 Data Acquisition

As discussed in previous sections, a novel contribution of this research is the development and integration of an IoT-based architecture that automates environmental data collection and data analysis. Data acquisition is a fundamental and essential task when conducting data analysis. The IoT kit was placed for 3 months to collect data in 60 minutes. Figure 3.9 presents the sensor data observations acquired on the on-cloud Mongo database and visualized on the web dashboard. The dashboard presents the last sensor records in widget form such as temperatures, moisture, pH, and biogas variation for a specified period. The developed web application intends to facilitate the biogas operators in decision-making processes.

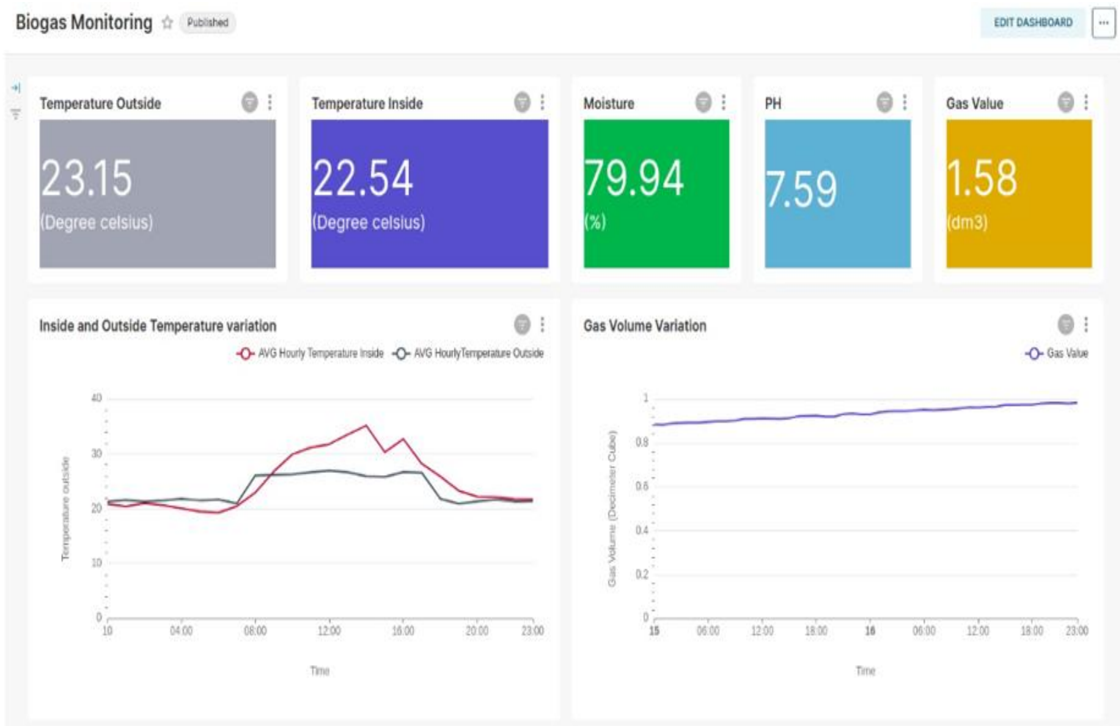


Figure 3.9: Captured data sample appearing on the web user interface.

During the data collection period, a dataset of 3000 records is acquired on the Mongo database, each record has 6 variables. Table 3.5 presents, the sample of the first five rows displayed using the head() method of Pandas library among 3000 records of the dataset. The description of variables within the dataset is the following: (1) moisture content of substances (moisture), presents the moisture content level of substances within a biogas digester presented in percentage (%). (2) the outdoor temperature (Temp_out), presents the outdoor temperature of the digester in degrees Celsius (°C). (3) The temperature of substances within the digester (Temp_in) presents the indoor temperature of Celsius. (4) The pH level of the substances (pH), represents the acidity level of the substances which ranges from 0-14. (5) gas generated (gaz_change), which presents the gas yield from the biogas generation process in a decimeter cube (dm^3) and (6) acquisition time (time_occur) presenting the timestamp when the data is captured.

Table 3.5: Dataset Sample

Moisture	Temp_out	Temp_in	pH	Gaz value	Time-occur
85.24	20.32	36.90	6.86	0.08	3/1/2023 0:01
85.95	19.59	36.60	7.62	0.07	3/1/2023 0:16
86.04	20.96	37.77	6.27	0.08	3/1/2023 0:31
83.31	19.67	35.00	7.31	0.06	3/1/2023 0:46
85.18	20.33	36.50	6.24	0.08	3/1/2023 1:01

3.2.4 Data Pre-Processing

The data acquired from the experiment are extracted in CSV file format to be used in the ML model as input data. Data preprocessing was performed using the Anaconda Python programming environment. Before preprocessing a set of Python libraries such as Matplotlib, Pandas, Scikit-learn library, and NumPy are imported for data preprocessing. The dataset is imported using the read_csv () function of the panda's library. The CSV file contains certain extra columns and rows with missing values and high peak values which are impractical. The rows with missing and high peak values are replaced with mean values of entire available values via the Imputer class of sklearn preprocessing library. In addition, timestamp values are converted from the 12h system to 24h using the strftime() function from the datetime library to easily employ time in our model. Table 3.6 shows the sample of the dataset after pre-processing.

Table 3.6: Dataset after Preprocessing

Moisture	Temp_out	Temp_in	pH	Gaz value	Date	Time (12)	Time (24H)	Day_hour
85.24	20.32	36.90	6.86	0.08	3/1/2023	12:01 AM	0:01	12.0
85.95	19.59	36.60	7.62	0.07	3/1/2023	12:16 AM	0:16	12.2
86.04	20.96	37.77	6.27	0.08	3/1/2023	12:31 AM	0:31	12.3
83.31	19.67	35.00	7.31	0.06	3/1/2023	12:46 AM	0:46	12.5
85.18	20.33	36.50	6.24	0.08	3/1/2023	1:01 AM	1:01	1.0

3.2.5 Data Analysis

A major objective of the IoT-based architecture is to improve the quality monitoring of environmental parameters in biogas generation context using sensors, and data analytics tools. In this regard, the multiple linear regression model was adopted to assess the fitness of environmental parameters, subsequently, the Pearson correlation analysis was utilized to examine the relationship between environmental parameters and biogas production.

A multiple linear regression model is a supervised ML model that employs two or more independent variables to forecast the outcome of a dependent variable [171]. This research adapted Equation (3.1) to validate the contribution of digester environmental parameters in biogas generation. The ordinary least squares (OLS) regression analysis technique is employed to find the best fitting. In essence, the OLS entails leveraging the parameter estimation from linear regression and considering the sum of squared discrepancies between the real sample value and the OLS estimation as the main point of reference for parameter estimation [172]. In the context, y represents the dependent variable, and βs are the regression coefficients. And a set of Xs presenting the independent variable [173].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \tag{3.1}$$

Multiple linear regression model can be evaluated by several metrics including RMSE, MAE, and R^2. This study explores R^2 to validate the model since it is very informative compared to others

Pearson correlation coefficient indicated in Equation (3.2), is a measure of linear correlation that can analyze the relationship between two or more variables [174]. In this context, it is adopted to express the degree of linear correlation between environmental parameters and biogas production.

$$\boldsymbol{r_{xy}} = \frac{\sum(\boldsymbol{x}-\overline{\boldsymbol{x}})(\boldsymbol{y}-\overline{\boldsymbol{y}})}{\mathbf{N}\cdot \boldsymbol{S_x S_y}} \tag{3.2}$$

The formula given defines the correlation coefficient for the two variables. In the formula, N represents the total number of data samples, while x and y represent the mean values of the two sets of variable data. S_x and S_y represent the standard deviations of the respective variable data samples.

3.3 Result

3.3.1 Model Validation Results

The validation of the models was assessed using R^2 metrics. the model was constructed by taking indoor temperature, ambient temperature value, and moisture content as independent variables and biogas production value as the dependent variable. Table 3.7 presents the OLS model result, R^2 indicates that 73.4% of the variability in biogas production is explained by the environment parameters explored in the regression model. This implies that the model has a relatively good fit and can effectively predict the biogas production process.

Table 3.7: OLS multiple regression model results

Dep. Variable	Gaz-value	R^2	0.734		
Model	OLS	Adj. R^2:	0.734		
Method	Least Squares	F-statistic:	2066		
	Coef	std err	T	P>|t|	[0.025 0.975]
Const	-0.3942	0.000	42.463	0.040	-0.376
Ph	0.6021	0.000	19.186	0.021	0.002
Temp_in	0.843	0.000	25.460	0.012	0.005
Temp_out	0.526	0.000	17.903	0.032	0.003
Moisture	0.0129	0.000	20.974	0.040	0.003

3.3.2 Correlation Analysis

The relationship between biogas production (y-variable) and environmental parameters (x-variables) was calculated by Pearson correlation coefficient r. The relationship between environmental variables and biogas yields has been constructed using the Seaborn heatmap Python data visualization library, which revealed the correlation between temperature, moisture, pH, and biogas yield.

Figure 3.10 presents the inter-variables correlation matrix, constructed. The value of r falls within the range of -1 to +1. When $r>0$ indicates a positive correlation between the two variables. It means that as one variable increases, the other variable increases as well. Conversely, when $r<0$ indicates a negative correlation, where if one variable increases, the other variable tends to decrease, and the larger the absolute value of r, the stronger the correlation. It is important to note that correlation does not imply causation. When $r=0$ indicates no linear correlation between the two variables.

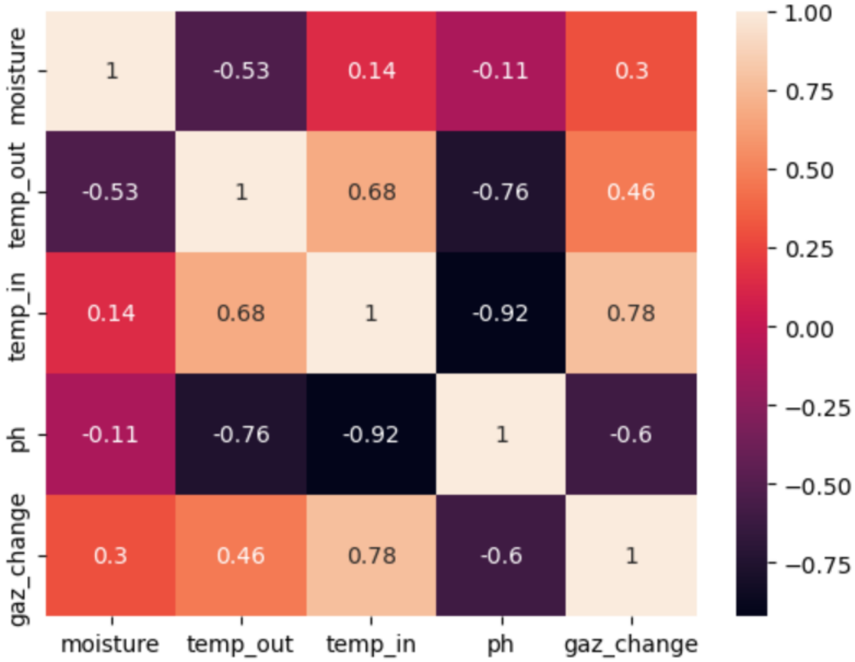
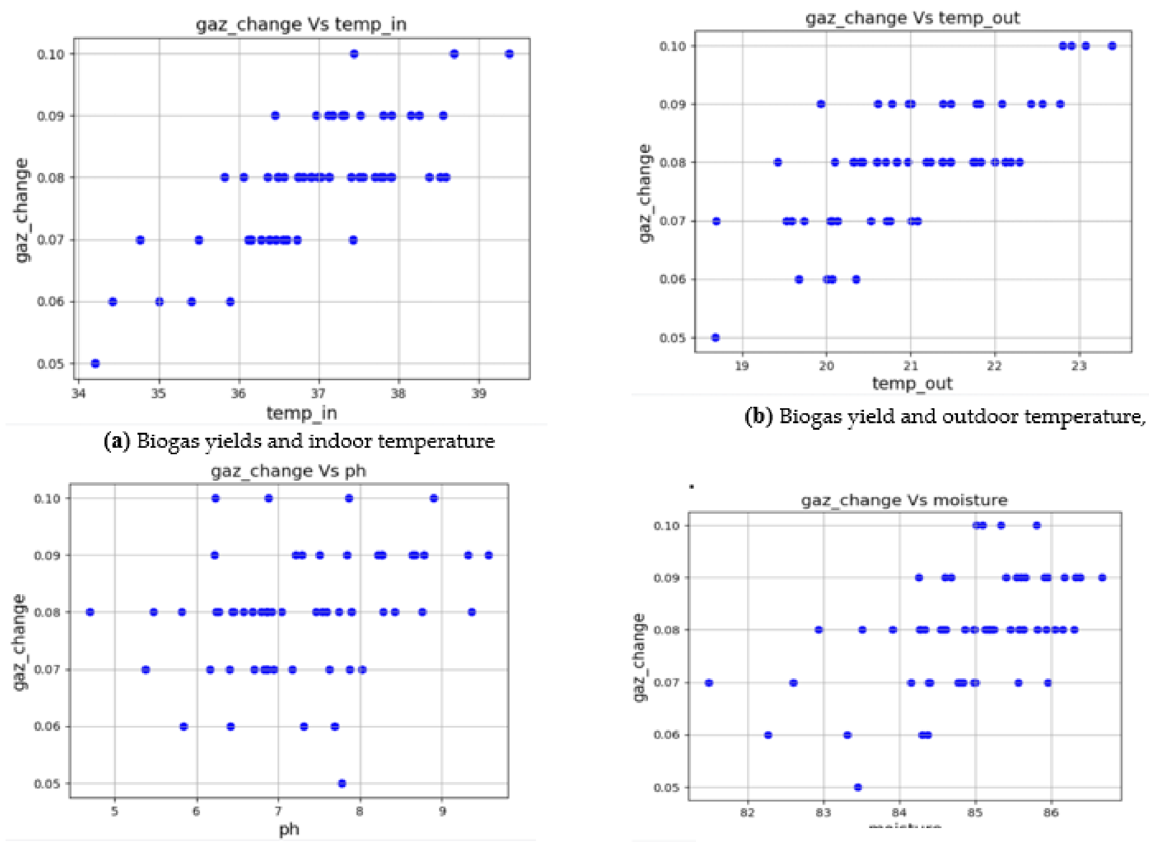


Figure 3.10: Inter-correlation matrix

The Pearson correlation result shows that indoor temperature and pH have a strong correlation of 0.77, and -0.6, while outdoor temperature and moisture present a moderate correlation of 0.46 and 0.3 respectively. As presented in Figure 3.10, throughout the experiment, the matplotlib python library is imported to create graphs representing the correlation between variables. According to its constraints, the premise of the correlation analysis is that the distribution of environmental parameters conforms to a normal distribution. Figure 3.11 provides information regarding our findings.



(a) Biogas yields and indoor temperature

(b) Biogas yield and outdoor temperature,

(c) Biogas yield and pH

(d)Biogas yield and moisture

Figure 3.11: Relationship between biogas yield and environmental parameters.

In Figure 3.11(a) a correlation between the biogas yield and indoor temperature indicates an indoor temperature range between 37–39 °C, inducing the maximum production of about 0.10dm^3. In Figure 3.11(b), the correlation between the biogas yield and outdoor temperature shows the maximum outdoor temperature of (23–24) °C, resulting in the maximum biogas generation of about 0.10dm^3. In Figure 3.11 (c), the correlation between the biogas yield and pH shows a maximum biogas production of 0.10dm^3 for a pH range (6-9). Figure 3.11 (d), a correlation between biogas yield and moisture indicates a moisture range of 85-86, presenting the maximum biogas production. The results indicate that the environmental parameters have a high impact on biogas production. Moreover, the temperature increased both outdoors and inside during the daytime while it decreased at night. Thus, it can be concluded that the daytime period also has a significant impact on biogas prediction.

3.4 Discussion and Conclusion

This research aims to improve Rwanda's biogas industry by developing an IoT-based architecture for collecting and managing data on biogas functionality. The prototype, tested in a home environment, involves sensor nodes equipped with multiple sensors to capture environmental parameters like temperature, pH, and moisture. Each node is uniquely identified and linked to specific biogas owners, allowing real-time monitoring and mapping of biogas setups across Rwanda. The data from these sensors are transmitted to a central database using the MQTT protocol, ensuring real-time, low-bandwidth communication. The system is managed through a secure web interface developed using the Laravel PHP framework, and the setup was tested in a production environment over 3 months.

The collected data were analyzed using statistical methods and supervised ML models to evaluate the impact of environmental factors on biogas production. Multiple linear regression analysis yielded an R^2 value of 0.734, indicating a strong predictive capability for biogas production. Pearson correlation analysis further revealed that indoor temperature and pH had strong correlations with biogas output, with coefficients of 0.78 and -0.6, respectively, while outdoor temperature showed a moderate correlation of 0.46. These findings suggest that the developed IoT-based architecture when integrated with the ML model can effectively predict biogas production and offer valuable insights for optimizing biogas systems in Rwanda.

The impact of IoT-based monitoring extends beyond the immediate benefits of improved production. However, IoT systems can play a vital role in ensuring the quality and accessibility of biogas, a crucial fuel source for many. While developing the abovementioned architecture, the following was found as a limitation: a lack of benchmarks for sensor calibrations, and there is a need for industrial sensors for precise accuracy.

Chapter 4

Analysis of Energy Harvesting for Self-Powered Internet of Things Edge Node Devices Applied in Biogas Generation Context

The IoT ecosystem has become a key enabling for the development of intelligent systems. Various IoT hardware components such as sensors, microprocessors, and communication devices require power to operate, and they are often deployed in an electrical power-constrained environment. Therefore, investing in energy harvesting as an alternative power source can improve the efficiency and lifetime of IoT applications. This section presents the designed IoT sensor node to monitor the biogas production environment and investigates the energy harvesting techniques suitable to power the WSN. The appropriate power harvesting sub-system capacity was analyzed. The result shows that the proposed IoT-based architecture can be powered by solar panel PV of 15W, 12V, 1.25A, directly connected to the output of maximum power point tracking (MPPT) type charge controller of 12V, 2A, the rechargeable battery of 12V, 12Ah. Furthermore, in this research, a mathematical model for predicting a solar panel size for a given global solar radiation is developed. The model result shows that the proposed sensor node can be powered by a solar panel size varied from 17.8 cm2 to 21.7 cm^2, for a given GHI available in the deployment environment. The model was validated for different values, and it is generic to be adopted anywhere. Decision-makers can adopt the result of this work to achieve precision energy.

4.1 Introduction

Recently, IoT technology has changed our daily living practices in various fields such as agriculture, manufacturing, smart cities, and energy management [175-179]. Regarding energy management, biogas is one of the most promising renewable energy sources affordable by local communities in developing countries. The benefits of adopting IoT in biogas generation are to improve productivity, and quality of biogas yield, through regular environmental parameters monitoring. The designed IoT system comprises a WSN with various sensors to monitor the environment such as temperature, humidity, pH various gases, and moisture levels, and a set of actuators to take necessary action in case a given threshold is surpassed [180]. Biogas energy is one of the alternative sources of energy available in electricity-constrained environments.

Considering this, the WSN is powered by rechargeable batteries, and the lifetime of the WSN is based on the powered battery capacity. Therefore, to maximize WSN utilization, there is a need to apply the appropriate energy harvesting technology.

Energy harvesting is the technique of turning ambient energy into electrical energy, and it is classified based on the source that is being harvested, they are grouped into five categories namely: mechanical, ambient, organic human, and hybrid [181-183].

The researchers adapted different renewable energies such as solar, wind, and hybrid as the primary source of supplying power for the remote network nodes [184,185]. In this research, solar energy is adopted due to its remarkable optimum power density providing adequate energy for powering systems [186]. Solar energy is harvested through three processes such as photosynthesis, photovoltaic converter, and helio-thermal, the photovoltaic source is our focus, the PhotoVoltaic (PV) effect, occurs when two different materials transform sun rays into direct current (DC) power when exposed to light [187]. Solar cells, often known as photovoltaic cells, are solid-state electrical junction device that uses sunlight to generate electricity [188].

To analyze an appropriate power harvesting method, various parameters such as energy source characteristics, storage device type, power management unit, sensor node operational mode, communication protocols, and sensor node specification must be considered [189]. The power harvesting prior research work focuses on analyzing power harvesting circuit design and energy harvesting computation [190-193]. Also, the state of the arts focuses on the analysis of power harvesting solutions that are generic for all IoT devices. However, the embedded systems behave differently, there is a need for a power harvesting solution focused on the power requirement of individual systems.

The contribution of this thesis is to analyze the energy harvesting system and develop a solar panel size prediction model for a designed sensor node. The proposed model suggests a solar energy PV size, of an optimum 17.8 cm^2 for a given solar irradiation available in the deployment environment. In addition, this research analyzed the specification of the energy harvesting sub-system based on the electrical parameters. The results propose that WSN can be powered by solar panel PV of 15W, 12V, 1.25A, and it is directly connected to the output of MPPT type charge controller of 12V, 2A, the rechargeable battery of 12V, 12Ah. The proposed solar energy harvesting capability makes it suited for wireless sensor nodes which require daily energy consumption of 9.8Wh.

4.2 Related Works

The proliferation of IoT devices has led to increasing demand for efficient and reliable power sources to support their continuous operation [187]. With the IoT network growth, traditional battery-powered solutions may not be viable due to the challenges of battery replacement.

The selection of an appropriate energy harvesting method depends on the specific deployment scenario and the energy requirements of the IoT devices. The piezoelectric harvesters could be advantageous in environments with significant mechanical motion, and piezoelectric materials are used for converting mechanical vibrations into usable energy [188]. Hybrid approaches combining multiple energy harvesting techniques have been explored to improve the reliability and resilience of the power supply [189]. Furthermore, Solar energy is well-suited for outdoor applications with sunlight. Solar energy harvesting uses photovoltaic cells to convert sunlight into electrical current. In this research, solar energy harvesting has been identified as the best to power outdoor applications biogas system.

Solar energy harvesting involving the conversion of solar radiation into electrical energy, has been proposed as a viable approach to address the energy constraints faced by IoT devices [191], particularly in regions with high solar irradiation. This technique allows IoT devices to be self-powered, reducing the need for frequent battery replacement. Several studies have explored the application of solar energy harvesting in IoT devices. For example in [192] an energy harvesting system specification was made focusing on its circuit design simulation. Similarly in [193], a hybrid sustainable energy system is designed, using solar photovoltaic and wind turbines. the proposed system uses a single-ended primary-inductance converter to boost the electrical energy generated to attain the required voltage level when charging the battery.

Solar energy harvesting current studies propose generic solutions that have made a significant impact on supporting IoT applications. However, each embedded system has unique power requirements that need to be considered. Furthermore, there is a need for a generic model that can predict the solar energy required for a device in any deployment location.

This research contributes by analyzing the solar energy system capacity needed to power the designed sensor node and developing a mathematical model for predicting the solar energy requirement given solar irradiation as the dependent variable.

4.3 Methodology of Power Harvesting System Design

This section presents the methodology for investigating the appropriate solar energy to power the designed wireless sensor node. In this research solar energy harvesting depends on the sensor nodes' energy budget, and solar irradiation is found in the deployment geographical location. The sensor node power harvesting capacity was computed referring to electrical parameters, and a mathematical model for predicting the solar PV size was derived from derivatives function that converge at a point theorem.

4.3.1 Power Harvesting System for the Wireless Sensor Node Block Diagram

Figure 4.1 presents the block diagram comprising an IoT-based system for monitoring the biogas digester status and the corresponding energy harvesting system. A sensor node unit is mounted with various sensors to measure moisture content, pH level, pressure, humidity, and temperature respectively. Sensing devices are connected to a customized Raspberry Pi 3 B+ microcontroller with a built-in Wi-Fi module. The reasons for component selection are detailed in Table 3.2

The sensor data are pushed to the MongoDB database hosted on the cloud server via the MQTT communication protocol. The designed wireless sensor node is intended to communicate the data to the biogas operator via a remote web application.

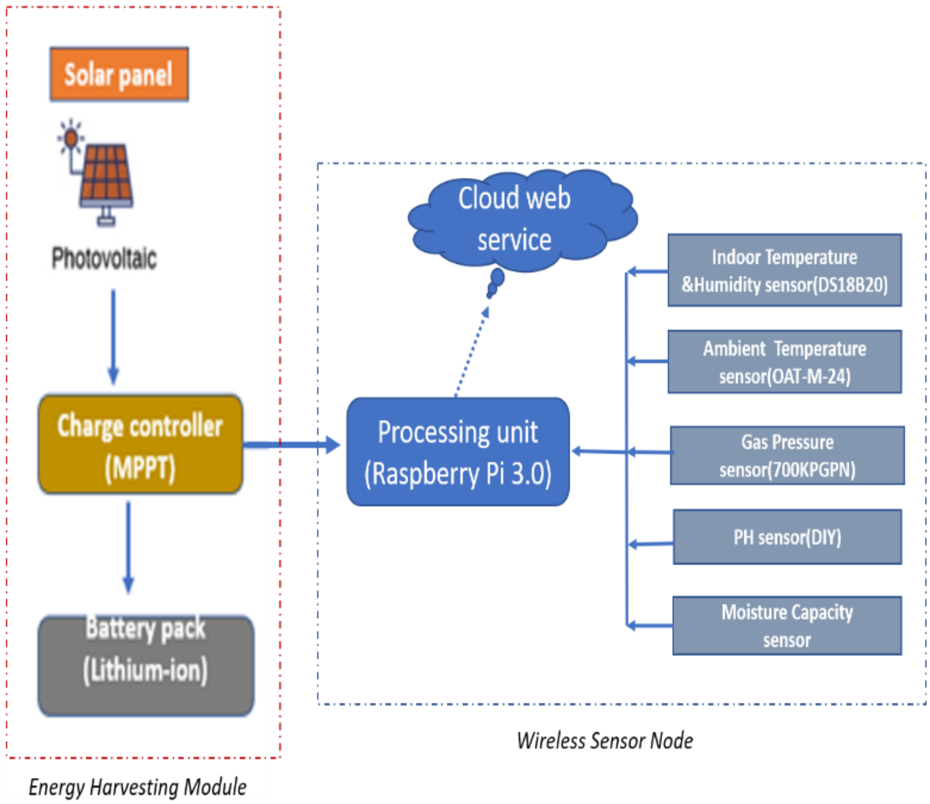


Figure 4.1: Power harvesting for the designed wireless sensor node block diagram.

To ensure the sustainability of a sensor node, in this research, solar energy is found the most convenient to power the designed Sensor, given that it is intended to operate as an outdoor application. The proposed solar energy harvesting system comprised of a collection of sun rays, solar panels that are used to convert light energy into electrical energy, and the charge controller's main function is to extract power from the panel, store it in the battery pack, and supply energy to the wireless sensor node/network. The charged battery is then used to power the sensor node through the charge controller as well, to avoid over-discharge of the battery pack.

4.3.2 Analysis of Solar Energy Harvesting

The solar energy harvesting analysis depends on the sensor nodes' energy budget, and solar irradiation found in the deployment geographical location. Initially, an analysis of sensor node energy consumption was made referring to the datasheet of each component. Furthermore, the polynomial-based mathematical model was developed for predicting the solar PV size. Finally, the power harvesting subsystem (solar panel, battery bank, charger controller) specification was computed based on electrical parameters.

4.3.2.1 Wireless Sensor Network Energy Consumption

In this research, a solar energy harvesting system for an IoT system depends on the sensor node energy consumption of each electrical load in the circuit. Table 4.1 illustrates, the energy consumption of each component in three different states namely: sleep, active, and transition mode according to the datasheet of each component.

Table 4.1: Energy consumption of each node component in three states.

Device	Sleep Current (mA)	Active: Wake-Up Current (mA)	Active: Processing Current (mA)	Transition Period (ms)	Energy usage in the transition period (mWh)
DS18b20	0.75	1.5	2	10	0.000021
OAT-M-24	0.1	0.2	0.3	5	0.000001
700KPGPN	0.5	1	1.5	8	0.000011
DIY pH Sensor	0.2	0.4	0.6	7	0.000004
Moisture Capacitive Sensor	0.3	0.6	0.9	12	0.000010
Raspberry Pi 3	80	200	700	50	0.013889
Energy usage in the transition period (mWh)					0.013936

At this point, further calculations are based on sensor node data sending rating represented by circle/16ms. The data processing and communicating are 85ms, assuming that the sensor node control board will be powered by 5V. Therefore, during 24hrs data will be processed in 7.6 seconds. Therefore, knowing the sensor node processing time, and the power consumption as computed in Equation (4.1), the active energy of 7.45mWh was calculated from Equation (4.2).

$$APower = Rated\ volt\ *\ Current\ consumed \tag{4.1}$$

$$AEnergy = APower(mW) * ATime(h) \tag{4.2}$$

Where: AEnergy, APower, ATime is energy, power, and time spent in active mode. Additionally, to ensure energy optimization, the sleeping model must be longer than other models. The daily node sleeping time of 23.87hrs was derived from to sensor node active time, thus the Energy consumption in sleeping mode was computed from Equation (4.3).

$$SEnerg = SPower(mW) * STime(h) \tag{4.3}$$

Where: SEnergy, SPower, STime is energy, power, and time spent in sleep mode. Finally, the overall sensor node energy of 9.8Wh results, from summations of transition, sleeping, and active energy as seen in Equation (4.4).

$$Tenergy = Tenergy + SEnergy + AEnergy \tag{4.4}$$

Where Tenergy is the total energy consumed by the sensor node.

4.3.2.2 Battery Size

Referring to the daily sensor node's energy consumption, an optimum battery capacity required for a certain number `n` of days was obtained by Equation (4.5):

$$Bn = (DC * nACF\ )/DoD \tag{4.5}$$

Where: DC, ACF, and DoD are daily consumption, annual correction factor, and depth of discharge respectively. An annual correction factor is a multiplier used to account for the expected decrease in a battery's capacity over time due to aging and environmental factors as the batteries degrade over a year, an annual correction factor is required to ensure that your system has enough battery capacity to meet the needs. For example, lead-acid batteries, have a higher annual correction factor (around 1.2 to 1.3) than Lithium-ion batteries (around 1.1 to 1.15) The DoD is the amount of battery energy you can use from the battery lead-acid, and Lithium-ion batteries have a range of 50%-65%, and 80%-90% respectively [194]. Assume, a battery that can supply the sensor node for 3 days, and the sensor node control board will use 5Vdc for the applied Lithium-ion battery.

4.3.2.3 Charge Controller (CC)

A charge controller panels the voltage and current that the solar panel sends to the battery to prevent battery overcharging [195]. In this research context, to analyze the charger controller capacity, three important parameters such as solar array generated energy, power of the battery bank, and correction factor are taken into consideration as given in Equation (4.6). The correction factor of 1.2 to 1.3 value is commonly adapted to account for various losses and fluctuations, however, it can be adjusted based on factors like temperature, and cable length between CC and solar panel [196].

$$\text{CC} = (\text{solar energy/power battery}) * CF \qquad (4.6)$$

In this context, the correction factor of 1.3 is considered, and with sensor node energy consumption, a 15W solar panel is safe enough to power the sensor node and the storage battery simultaneously.

4.3.2.4 Expected Solar Panel Size

The expected solar panel size expressed in the solar panel area was computed by exploring simultaneous Equations (4.7) and (4.8) to get the polynomial equation that satisfies the behaviors of the derivative that converges at the point theorem. "According to the behavior of derivatives that converge at a point, the value of that converging point is the amount that can be added to or subtracted from the gradient at the present input point to get the gradient at the next adjacent input point, but the difference of that two adjacent input points must be equal to 1" [197]. Additionally, daily energy consumption and GHI are given as inputs. Table 4.2 presents sample data used to develop the model.

Table 4.2: Explored input sample data.

GHI (kWh/m^2)	Sensor node energy consumption((kWh)	Expected Solar panel size (cm^2)
5.5kWh/m^2	9.8*10-3 kWh	17.8182cm^2
4.5kWh/m^2	9.8*10-3 kWh	21.778cm^2
3.5kWh/m^2	9.8*10-3 kWh	28cm^2

With sample data presented in Table 4.2, we can apply Equations (4.7) and (4.8) to solve the derivative converging point noted as "a" and the first integration constant noted as C1.

$$y = ax\Delta x + y(x) \qquad (4.7)$$

$$2y = 2ax\Delta x - a(2x+1) - 2C1 + 2y(x+1) \qquad (4.8)$$

Where y stands for solar panel area, x stands for GHI obtained in Rwamagana, Δx is an infinitesimal change in x, $y(x)$ is y as a function of x, this means that it's the value of the function y at a specific value of x, and $y(x+1)$ is the value of the function y at a specific value of X_{+1}

Starting from $x = 3.5kWh$ and $y = 28\ cm^2$, (4.7) equals:

$$y = 3.5a\Delta x + 28 \qquad (a)$$

Substituting (a) in (4.8),

$$2(3.5a\Delta x + 28) = 2(3.5a\Delta x) - a(2x + 1) - 2C1 + 2(21.7778)$$

$$7a\Delta x + 56 \; = \; 7a\Delta x \; - \; 8a - 2C1 + 43.5556$$

$$12.4444 \; = \; -8a \; - 2C1 \qquad \text{(b)}$$

Again considering $x = 4.5kWh \; and \; y = 21.7778 \; cm^2$, (4.7) equals:

$$y = 4.5a\Delta x + 21.7778 \qquad \text{(c)}$$

Substituting (c) in (4.8),

$$2(4.5a\Delta x + 21.7778) = \; 2(4.5a\Delta x) - \; 10a - 2C1 + 2(17.8182)$$

$$7.9192 = -10a - 2C1 \qquad \text{(d)}$$

Solving s (b) and (d) for “a” and “C1” simultaneously using the elimination method we get, $a = 2.2626 \; and \; C1 = -15.2726$

As stated, before in this reseach, if \`$a$\` is the final converging point of derivative for the actual equation we are looking for and $C1$ is the constant of the first integration of that constant, we get:

$$\int a \; dx \; = \; ax \; + C1 = 2.2626x - 15.2726 \qquad (4.9)$$

Therefore, daily energy consumption, GHI, and solar panel size relationship doesn’t exhibit linear characteristics, hence Equation (4.9) does not satisfy the expected equation that gives the solar size receiving GHI as input. According to the concept stated above, Equation (4.9) is a derivative of the equation we’re looking for. For that, we need to do a second integral.

$$y = \int (2.2626x - 15.2726) \; dx = 1.1313x^2 - 15.2726x + C2 \qquad (4.10)$$

Finally, There is no doubt that if we apply this explicitly stated limit of GHI found in Rwamagana \`$4.5 \leq x \leq 5.5$\` and keep in mind that $x \neq 0$, then $y = 1.1313x^2 - 15.2726x + C2$ the equation can accurately provide the area of solar panel size (in cm^2) which can be used to power/supply our sensor node will have an energy consumption of $9.8 * 10^{-3} kWh \; per \; day$.

Using the initial data, we used, we can get the value of C2 and prove that our equation is accurate,

- If $y = 28, when \; x = 3.5, C2 = ?$

$$C2 = 28 - (1.1313 * 3.5^2) + (15.2726 * 3.5) = 67.595675$$

- If $y = 17.8182, when \; x = 5.5, C2 = ?$

$$C2 = 17.8182 - (1.1313 * 5.5^2) + (15.2726 * 5.5) = 67.595675$$

Thus, the final equation is,

$$y = 1.1313x^2 - 15.2726x + 67.595675 \quad (4.11)$$

Finally, if we apply this explicitly stated limit of GHI found in the study area `$4.5 \leq x \leq 5.5$` and keep in mind that $x \neq 0$, Equation (4.11) can accurately provide the area of solar panel size (in cm^2) which can be used to power/supply our sensor node.

4.4 Result

4.4.1 Analysis of Solar Harvesting System Specification

To design an energy harvesting system for a given device requires key knowledge. It is very important to first know the energy consumption of that device. In this research, the WSN daily energy consumption was preliminary computed in Equations (4.2), (4.3), and (4.4). Table 4.3, presents the energy consumption of sensor nodes in three states, showing high energy usage in a sleeping state. To minimize power consumption, in this research we assume that the active mode is taking less time, and the overall sensor node energy consumption of 9.8Wh was achieved.

Table 4.3: Daily energy consumption of a sensor node.

State	**Power (mwatts)**
Sleep	9727.025
Active	7.45267
Transition	0.0139

Furthermore, we referred to the sensor node energy consumption to obtain the corresponding solar harvesting sub-system. The capacity of the solar panel, battery, and respective charger controller was computed in Equations (4.5), and (4.6) respectively. Table 4.4 presents, solar harvesting components specifications, suitable to power the sensor node.

Table 4.4: Solar power harvesting system component specifications.

S/N	Name	Specifications
1.	Solar panel (PV)	15W,12V, 1.25A
2.	Charge Controller	MPPT type Charge Controller. 12V, 2A,
3.	Battery	12V, 12Ah

4.4.2 Solar Panel Size Prediction Model Result

The contribution of this research paper is to develop a mathematical model for predicting the solar panel size. The model presented in Equation (4.11) was derived from Equations (4.7), and (4.8). Y represents the dependent valuable "The expected panel size in cm^2", and X represents the dependent valuable "Global Horizontal Irradiation". The model presented is a polynomial-based equation fitting the sample data explored in Table 4.2.

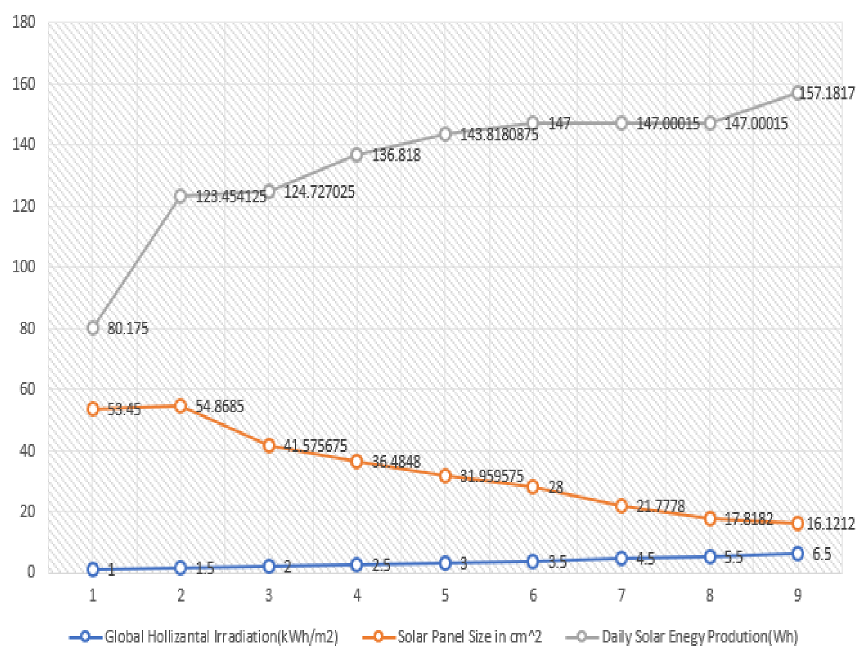


Figure 4.2: Correlation between GHI, solar panel size, and daily solar energy production.

As shown in Figure 4.2, the model result shows that the sensor node can be powered by an optimum panel size of 17.8 cm^2 for given an optimal GHI of 5.5 kWh/m^2 available in the deployment environment. However, the provided data reveals a systematic reduction in solar panel size with increasing levels of GHI, suggesting an intriguing relationship between solar exposure and panel dimensions. As GHI rises from 1 kWh/m^2 to 6.5 kWh/m^2, the corresponding solar panel size diminishes consistently. This inverse correlation hints at the potential for leveraging higher irradiance to achieve effective energy capture with smaller physical footprints. Despite fluctuations, solar energy production remains relatively stable from 2 kWh/m^2 to 6.5 kWh/m^2, suggesting a threshold effect in energy output. This highlights the importance of optimizing panel size relative to GHI to maximize energy yield while minimizing physical space requirements. The findings prompt consideration of the trade-offs between panel size, efficiency, and energy output, highlighting opportunities for the design and deployment of solar energy systems tailored to varying irradiance levels.

4.5 Discussion and Conclusion

The IoT platforms play a critical role in monitoring and optimizing the biogas production process. However, the sustainability of those platforms requires a consistent and reliable power supply. To address this need, this research has proposed a solar energy harvesting technique that can effectively power the designed IoT sensor nodes.

In this context, solar energy harvesting leverages the principles of enhanced energy power capacity, as discussed in the literature, which incorporates sunlight as an appropriate energy source to power outdoor applications. The solar energy harvesting system specifications are derived from sensor node energy consumption. Furthermore, experimental analysis is conducted to develop a mathematical model predicting the solar panel size required to power the designed sensor node given global irradiation as a dependent variable. The model was validated for different values, and it can applied anywhere

The research findings are crucial for managing the power budget by optimizing maximum power harvesting, which helps sustain IoT systems, reduce unnecessary expenses, and prevent power shortages. This outcome presents a promising solution for achieving sustainable and precise energy for the IoT ecosystem in Rwanda and Africa, particularly in remote and off-grid areas with abundant sunlight. To enhance solar energy prediction models, it is essential to consider factors like weather conditions and temperature, in addition to solar irradiation, as these can significantly affect solar energy production.

Chapter 5

Maximizing Biogas Yield Using an Optimized Stacking Ensemble Machine Learning Approach

Biogas is a renewable energy source that comes from biological waste. In the biogas generation process, various factors such as feedstock composition, digester volume, and environmental conditions play a vital role in ensuring promising production. Accurate prediction of the biogas yield is crucial for improving biogas operation and increasing energy yield. The purpose of this research is to propose a novel approach to improve accuracy in predicting the biogas yield using the stacking ensemble ML approach. This approach integrates two machine-learning algorithms: LightGBM, CatBoost, and evolutionary strategy as an optimization algorithm to attain high performance and accuracy. During the training phase of the base learners, optimal hyperparameter values were identified using a random search optimization technique coupled with cross-validation. The ML and proposed triadic models were experimented on environmental data collected from biogas production facilities. The comparative analysis of the proposed model with others such as KNN, RF, and DT was performed. The study's findings demonstrated that the proposed model outperformed the existing models with a RMSE of 0.004, and an MAE of 0.0024 for accuracy metrics. In conclusion, an accurate predictive model cooperating with a fermentation control system can significantly increase biogas yield. The proposed approach stands as a pivotal step toward meeting escalating global energy demands.

5.1 Introduction

Biogas is indeed a renewable energy source that is produced through the decomposition of organic matter in an anaerobic environment [198]. It is primarily composed of CH_4 and CO_2, along with small amounts of other gases such as hydrogen sulfide H_2S, and trace compounds [199] [200]. Biogas can be used as a versatile fuel for various purposes, including electricity generation, heating, and even as a transportation fuel. Biogas production is a complex process influenced by multiple interconnected factors including feedstock composition, environmental parameters, and organic loading rate [201]. Different feedstocks have varying levels of biodegradability and methane potential. The common feedstocks include animal manure, agricultural residues, food waste, and wastewater sludge [202] [203]. Further, environmental parameters such as temperature, humidity, pH, and moisture level play a vital role during the biogas production process where the optimal temperature range is typically between 35°C and 55°C, Higher temperatures can accelerate the digestion process, but extreme temperatures can inhibit microbial activity [204], [205]. The pH level of the digester is crucial for maintaining optimal microbial activity. Most biogas production occurs in a slightly acidic to neutral pH range of 6.5 to 7.5 [206]. The length of time the organic matter remains in the digester, known as the retention time, affects biogas production. Longer retention times allow for more complete degradation of the feedstock and increased gas production The availability of essential nutrients, such as nitrogen and phosphorus, plays a role in microbial activity and biogas production where the carbon to nitrogen (C/N) ratio must be maintained in the optimum range for efficient biogas production [207].

With technology evolution, AI, and IoT technology, it is feasible to predict the biogas generation referring to the available influencing parameters dataset. Feedstock composition can vary significantly, even within the same category. This makes it difficult to establish a standardized prediction model that applies to all types of organic matter. However environmental parameters are a common factor that contribute to the overall biogas generation process. This research aims to investigate the contribution of environmental parameters in biogas prediction and propose a new prediction algorithm that guarantees high accuracy in estimating biogas output compared to the existing methods. Recent studies have highlighted the remarkable advancements made by AI and IoT techniques in enhancing renewable energy sectors [208], [209]. From a biogas perspective, research studies on AI in biogas prediction are explored to enhance the biogas generation process [210, 211]. For example, SVMs are presented as the most popular ML algorithm to predict the biogas output in several studies from wastewater treatment plants. The study findings showed that SVMs were able to achieve an accuracy of 95% [212]-[214]. Another researcher explored the contribution of artificial neural networks (ANNs) algorithm in biogas prediction, the research finding presents the highest accuracy of 92% [215]. Another paper investigated the application of the decision trees algorithm in biogas prediction by dividing the data into smaller groups until each group can be predicted with a high degree of accuracy, thus 89% of model accuracy was achieved. Further, the RF algorithm was explored by combining multiple decision trees to improve the accuracy of the prediction, however, 91% of the accuracy was presented [216].

Prior research, performed predictions based on single ML models, demonstrating their district dominance. Although a single prediction model can enhance production accuracy by adjusting parameters and choosing forecasting variables in the prediction process, it also carries uncertainties related to its model structure and faces, challenges when adapting to various environments [217, 218].

In the ML context, the continuous evolution of algorithms and techniques has opened up new avenues for improving the accuracy and efficiency of predictive models. The hybrid learning technique is widely adopted to reduce bias and variance by blending less powerful models to form a robust model [219]. Recent research studies have indicated the integration of multiple models to construct an ensemble model can effectively leverage the strengths of these diverse models, ultimately enhancing the dependability and precision of biogas prediction [220, 221]. However, the adoption of the ensemble model for biogas prediction is leaving significant unexplored potential.

This research proposes a novel stacking ensemble learning model for the prediction of biogas yield aims to optimize the robustness and accuracy. The stacking approach distinguishes itself from other assembling methods due to its hierarchical structure [222]. In stacking, predictions from multiple base models are combined and fed into a meta-model that learns to optimize the outcome. This approach allows the meta-model to leverage the strengths and address the weaknesses of each base model, potentially enhancing generalization and accuracy [223][224] In this research, the stacking process combines the LightGBM, and CatBoost ML models, and an evolutionary strategy algorithm was adopted to optimize the meta-model to attain high performance and accuracy.

Integrating LightGBM and CatBoost for biogas yield prediction is justified by their complementary strengths in handling diverse data types and improving model accuracy. LightGBM is chosen for its efficiency and scalability with large and diverse data types, making it a suitable choice for large-scale biogas production data analysis [225]. while CatBoost is selected for its superior handling against overfitting [226]. In the biogas forecast context, which involves complex interactions between various factors, the integration of these algorithms can enhance prediction accuracy by combining LightGBM’s speed and CatBoost’s precision, leading to a more robust and accurate predictive model.

To improve the performance of the integrated model (LightGBM and CatBoost), an optimization algorithm called evolutionary strategy is adopted to fine-tune the meta-model's hyperparameters hence improving its performance. Evolutionary strategy is a class of optimization algorithms inspired by natural selection, capable of efficiently navigating complex, high-dimensional search spaces to find optimal solutions [227]. Given the complex, nonlinear relationships inherent in biogas yield prediction parameters, where various numerical and categorical factors interact, evolutionary strategies can explore a wide range of potential model configurations [228],[229]. This approach helps discover the optimal settings that might be difficult to identify through traditional grid search or random search methods. Adopting evolutionary strategies features, the integrated model (LightGBM and CatBoost), can better capture the underlying patterns in the biogas experimental data, leading to improved generalization, robustness, and accuracy in predicting biogas yield.

In the biogas production process, various operational parameters can be adopted for forecasting biogas yield. This study focuses on developing a biogas yield prediction model based on the environmental parameters dataset. The accuracy metrics such as MAE and RMSE were adopted to evaluate the performance improvement of the hybrid model compared to other models explored in the research.

5.2 Materials and Methods

This section describes the material and methodology used for the implementation of this study. Regarding material, data were collected through an IoT framework designed and deployed at a home digester in a previous study. The data was subjected to pre-processing procedures, involving the elimination of errors or outliers, imputation of missing values, and normalization to ensure consistent scaling across all features. The proposed model was developed by combining two based classifies models, including LightGBM, CatBoost, and evolutionary strategy optimization. Finally, the proposed model was compared to other machine-learning techniques. Figure 5.1 presents the proposed method for the implementation of the research.

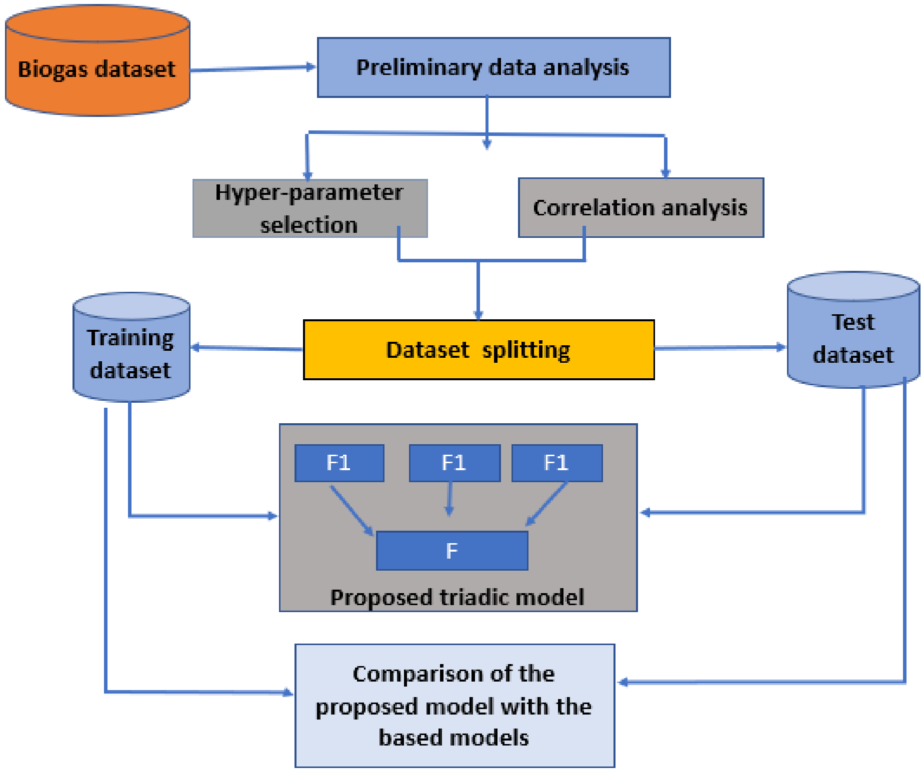


Figure 5.1: Proposed triadic ensemble model.

5.2.1 Data Collection

This research is part of other ongoing research works, in the previous study, an IoT-based architecture was developed to monitor and control the biogas digester status, and it is considered a data collection tool in this study. The data were collected from a home digester, and up to 3000 records were accumulated, encompassing operational parameters data, such as digester temperature (T), digester pH, moisture level, pH level, and gas volume. The description of the IoT platform and data collection process was explored in our previous study [230]. The temperature is important as it affects the production rate. The pH level is vital for determining the stability and corrosiveness of the biogas. Gas volume is a factor providing insights into its energy content. Moisture level measurement subtracts moisture levels and influences the movement of microorganisms. Table 5.1 presents variables considered in the proposed biogas prediction model.

Table 5.1: Explanatory variables in the biogas prediction model

Variables	Description	Unit
Moisture level	Moisture level of substances	%
temp_out	The ambient temperature of the digester	ºC
temp_in	Temperature inside the digester	ºC
pH	The quantitative scale of acidity and alkalinity of solutions of chemical compounds	log10[a(H+)
gaz_change	Unit of volume of process gasses	dm^3

5.2.2 Machine Learning Models Experimentation

In this research, we propose a triadic ensemble ML model that integrates three distinct algorithms: LightGBM, CatBoost, and evolutionary strategy. The model engages supervised ML regression models, where a set of input data is employed to predict the output data [231]. Subsequently, the proposed model is compared with existing regression models namely, Random Forest, KNN, and DL. The most effective model is recommended to predict biogas production. These predicted values can be utilized to optimize biogas plant operations or devise strategies for future biogas production endeavors.

5.2.2.1 Classical Models

Classical ML models are applied in various scenarios, depending on the characteristics of the problem and the data. In this research, The KNN, RF, and DT were chosen, due to their outperformance as the best model applied in prediction in the prior research. In this research, they will be adopted for comparative analysis with the proposed model.

- **The KNN Algorithm** is an ML technique used for regression tasks. It relies on the idea that similar data points tend to have similar values [231]. Throughout the training process, the KNN algorithm stores the whole training dataset as a reference, to perform prediction, a calculation of the distance between the input data point and the trained data by referring to the euclidean distance [232].
- **Random Forest** is a powerful ML algorithm, that can handle both classification and regression problems. Figure 5.2 shows how a random forest combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, the random forest algorithm is a bagging method expansion that employs

both bagging and feature randomness to produce an uncorrelated forest of decision trees [23].

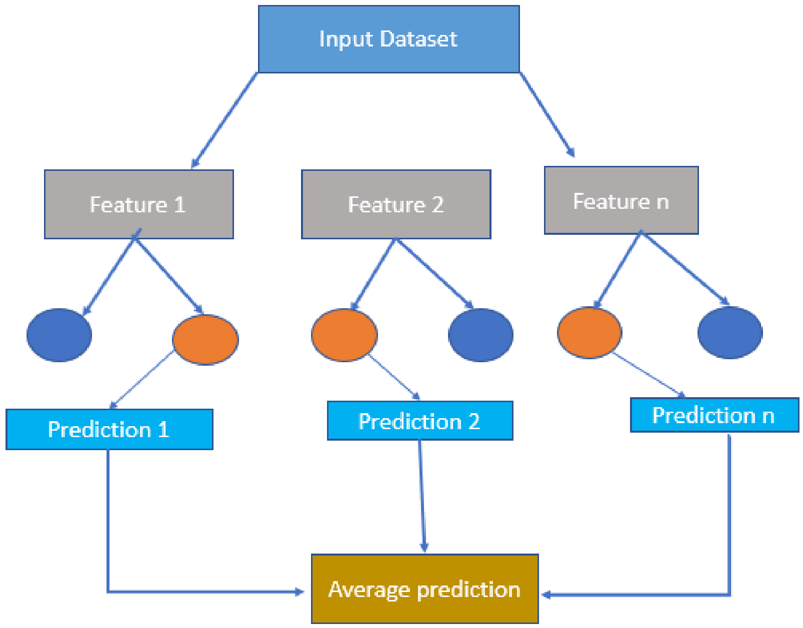


Figure 5.2: The Random forest model.

- **Decision Trees** are a popular supervised learning method that can work for both regression and classification problems. DT builds a model that predicts the value of a target variable by inferring basic decision rules from data features. It is a hierarchical decision support model that displays options and their probable outcomes, including chance occurrences, resource expenses, and utility [234].

5.2.2.2 Assembler Learning Models

Ensemble learning is a method that combines multiple weak models to enhance predictive accuracy, rather than relying on a single robust model. These weak models are collectively integrated to produce the final prediction. The descriptions of the algorithms adopted in this study are provided as follows.

Boosting

Boosting involves modifying the target value to the extent that the previous model did not fit, at each stage, a weak learner is incorporated to address the limitations of the current ensemble of weak learners [235]. Boosting seeks to enhance model accuracy through a sequential integration of weak learners. Regarding boosting models, improved algorithms were explored in this study such as LightGBM and CatBoost.

- **LightGBM**: employs various techniques such as gradient-based one-side sampling and exclusive feature bundling. LightGBM reduces computation time, and it achieves this by implementing gradient-based one-side sampling, focusing on objects with significant gradients while excluding those with small gradients from sampling [236]. LightGBM model suggested at Microsoft [237], that this advanced supervised algorithm is built on the foundation of gradient-enhanced decision trees. It has found applications in various domains, including medicine, economy, and agriculture applications [238]. As indicated in Equation (5.1) LightGBM is a gradient-boosting framework that uses tree-based learning algorithms and relies on a loss function that measures the discrepancy between the predicted and actual values of the target variable [239][240].

$$L(\Theta) = \Sigma\, l\big(y_i, F(x_i)\big) + \Omega(F) + \Psi(\Theta) \tag{5.1}$$

  Where $L(\Theta)$ is the loss function that depends on the model parameters Θ. The goal of ML is to find the optimal values of Θ that minimize the loss function. $F(x_i)$ is the model output or prediction for the input x_i. F is a function that maps the input space to the output space and is determined by the model parameters Θ, and the sum of all training samples (x_i, y_i) is denoted by Σ. The loss function, $l\,(y_i, F(x_i))$, measures the difference between the predicted value $F(x_i)$ and the true value y_i. The regularization term, $\Omega(F)$ is a function of the model output F, and it penalizes the complexity of the model. Additionally, there is an optional regularization term, $\Psi(\Theta)$ is a function of the model parameters Θ, and it penalizes the magnitude of the parameters.

- **CatBoost**: is a gradient-boosting framework invented in 2017, with the ability to handle regression features effectively, it introduces methods like ordered target statistics and ordered boosting to enhance traditional boosting techniques [241]. CatBoost relies on a loss function that measures the discrepancy between the predicted values and the actual values of the target variable CatBoost algorithm minimizes the loss function by updating the ensemble in each iteration. As indicated in equation (5.2), at the tenth iteration, the predicted value of the ensemble for a specific sample x_i is denoted by $F_{t^-1}(x_i)$, and the update equation for $F_t(x_i)$ is:

$$F_t(x_i) = F_{t^-1}(x_i) + \gamma_t h_t(x_i) \tag{5.2}$$

The learning rate γ_t in equation (5.2) corresponds to the learning rate for the t-th iteration, while $h_t(x_i)$ represents the prediction made by the t-th decision tree for the sample x_i [242].

Stacking

In contrast to other ensemble learning algorithms, stacking integrates various learning algorithms on a single dataset. Initially, a collection of base-level learning models is created. Then, a meta-level model is trained using the outputs from the base-level models. By combining the prediction results of multiple models in the first step as input for the meta-learner, stacking enhances prediction accuracy while mitigating biases.

An important consideration regarding stacking is the quantity of base learners [243]. Simply increasing their number doesn't consistently enhance prediction accuracy. To improve the stacking ensemble's accuracy depends on learning strategies were adopted rather than only focusing on expanding the base learner models [244]. Thus, identifying the best combination of base learner and meta learner is very important through an iterative experiment, and it is necessary to determine the optimal hyperparameter values for the base learners. In this research, a stack-based model is developed by combining two powerful boosting models such as LightGBM, and CatBoost.

5.2.2.3 **Optimization Algorithm**

The ML optimization algorithms are techniques used to minimize a loss function during the training process of the model. These algorithms adjust the parameters of the model to find the optimal set of parameters that lead to the best performance. Evolutionary strategy is found the best suit this research study.

Evolutionary Strategy: is a global optimization algorithm that incorporates stochastic elements, inspired by the biological principle of evolution through natural selection [245]. The evolutionary strategy algorithm optimizes the parameters $\theta_1, \theta_2, ..., \theta_n$ of model M to minimize the loss function $L(M,\theta)$. It generates a population of M models with random parameters: $\theta_1, \theta_2, ..., \theta_N$. It evaluates the fitness of each model in the population based on the loss function: $f(\theta_i)=L(M,\theta_i)$. Then select the top-performing models. Select the top k models from the population based on their fitness scores [246].

5.2.3 Hyperparameter Turning using Random Search

In ML, it's essential to set hyperparameters before initiating the model training process. Each ML algorithm has its own set of hyperparameters, and optimal performance is achieved only when these hyperparameters are appropriately configured during the running process. It is practical and challenging to manually set and find the optimal parameter for the best performance. While constructing an ensemble model, techniques such as grid search, random search, and Bayesian optimization are employed to tune the hyperparameters of each base model.

This research used RandomizedSearchCV hyperparameter optimization techniques for the performance of the proposed model. RandomizedSearchCV is a tool used to evaluate the hyperparameter computation. It is used to improve the prediction accuracy. RandomizedSearchCV discovers the following hyperparameters for the LightGBM model:

- Learning rate: [0.01, 0.1];
- Number of estimators: sp_randint(100, 1000);
- Maximum depth: sp_randint(3, 8).

The 'n_iter' parameter of RandomizedSearchCV is set to 10, indicating that 10 sets of hyperparameters will be randomly sampled from the identified parameter space. Moreover, the 'cv' parameter is set to 5, suggesting that nested 5-fold cross-validation will be used to measure the performance of each hyperparameter arrangement. Once the best hyperparameter configuration is known, the LightGBM model is trained on the entire training dataset using these settings. The trained model is then employed to make predictions on the unseen test dataset. Figure 5.3 shows how nested 5-fold cross-validation is performed, for each iteration, one of the 5folds is considered as a testing dataset.

Figure 5.3: The 5-fold cross-validation.

5.2.4 Proposed Triadic Ensemble Model

As previously stated, the systematic approach of combining base models in stacking ensemble learning, emphasizes the need to structure the model to enhance performance and diversity. The objective was to select three base learners that deliver high performance. The proposed stacking model structure is illustrated in Figure 5.4.

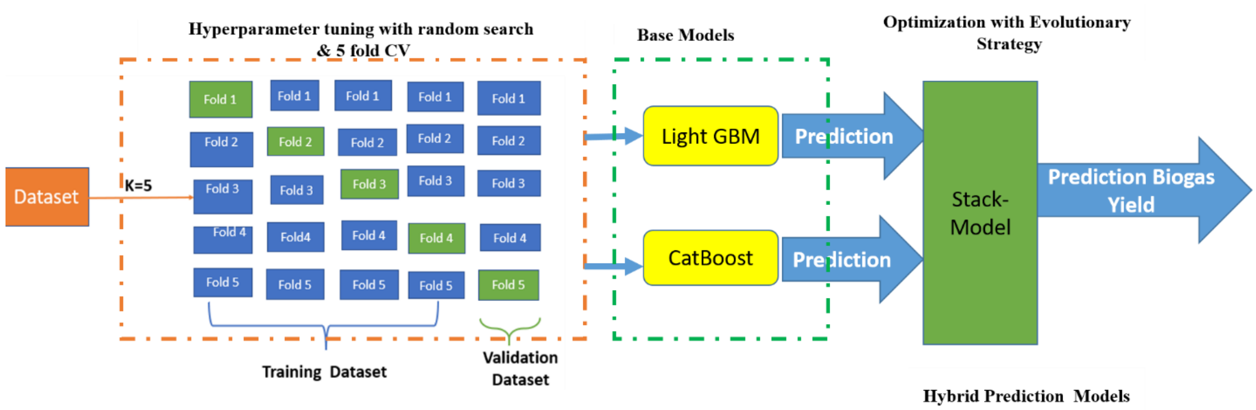


Figure 5.4: The Proposed model structure

Initially, we assess the performance of each ML model, followed by the selection of a base learner for the initial training phase of the stacking ensemble model. We used a five-fold cross-validation technique to evaluate each model independently. The hybrid model was derived from CatBoost, and Lightgb as base models. the evolution strategies optimization was adopted to optimize the learner model.

During the base learner learning phase, it's vital to optimize the hyperparameters. We achieved this by employing random search optimization with cross-validation to determine the optimal hyperparameter values for the base learners. The hyperparameters not only enhance the accuracy at the base model but also boosts the overall accuracy of the stacking model, which integrates these individual models.

During the initial phase, the original dataset is partitioned into a training set and a testing set. The training set is then used for training through k-fold cross-validation. In k-fold cross-validation, the training set is divided into k subsets, with each subset serving as a validation set while the remaining (k − 1) subsets are utilized for training the model and generating predictions for that specific validation subset. The details of the proposed triadic assembler algorithm are presented in Table 5.2.

Table 5.2: Proposed triadic essemble algorithm

Step.1 : *Train the base model with LightGBM:*

- *Initialize the LightGBM model (M1).*
- *Split the data into training and testing sets.*
- *Fit the model to the training data: M1.fit (X_train, y_train), where X_train represents the input data and y_train represents the biogas production output.*

Step.2 : *Refine the model using CatBoost:*

- *Initialize the CatBoost model (M2).*
- *Fine-tune the model parameters: M2.set_params (params).*
- *Fit the model to the training data: M2.fit (X_train, y_train).*

Step.3 : *Optimize the model parameters using Evolutionary Strategy:*

- *Set the population size (N) and maximum number of generations (G).*
- *Initialize the population of models with random parameters:*
- *P = [M1, M2... MN].*
- *For each generation (g = 1 to G):*
- *Evaluate the fitness of each model in the population based on prediction accuracy.*
- *Select the top-performing models (e.g., based on the highest fitness scores) for reproduction.*
- *Generate offspring models through variation and crossover operations.*

- *Replace the least fit models in the population with the offspring.*
  - *Select the best model from the final population based on fitness.*

Step.4 : *Prediction with the trained model:*

- *Use the best model to predict biogas production for new data inputs:*

y_pred = best_model.predict(X_new),

Where X_new represents the new data inputs.

Step.5 : *Utilize the predictions for optimization and planning:*

5.2.5 Evaluation Metrics

The evaluation metrics explored in the paper include the RMSE, MAE, and R^2. These metrics are used to assess the performance of regression models. The RMSE measures the average squared difference between predicted and actual values, with a lower RMSE indicating a better fit, and the MAE measures the average absolute difference between predicted and actual values, with a lower MAE also indicating a better fit [247, 248]. The R^2 gauge how well the model fits the data, with a higher R^2. We compare different regression models based on these metrics, the model achieving the lowest RMSE and MAE, as well as the highest R^2 is considered the best model for the task. The mathematical equations are as follows:

Mean Absolute Error

$$MAE(y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}|y_i - \hat{y}_i| \tag{5.3}$$

Coefficient of determination (R^2)

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{5.4}$$

Root Mean Squared Error (RMSE)

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2} \tag{5.4}$$

$\hat{y_i}$ is the predicted value of the i^th sample, and y_i is the corresponding true value for the total n samples.

5.3 Result

This section presents the prediction results from modeling biogas digester environmental data using the proposed stacked ensemble model, which is compared with other models in the study through performance metrics such as RMSE, MAE, and adjusted R^2. Additionally, the research explores the correlation among various variables.

5.3.1 Proposed Model Prediction Results

The dataset used in this study comprises 3,000 records, focusing on environmental parameters that impact biogas yield. The model is designed to predict the volume of biogas yield in the next hours based on previous measurements of five input values: ambient temperature, indoor temperature, moisture, pH level, and time. To select the training and testing dataset, the k-fold cross-validation approach was employed, with k=5 chosen to balance computational cost and prevent bias associated with lower k values.

Nested cross-validation was selected to optimize hyperparameters and reduce bias, with 5-fold cross-validation performed within each fold. In each iteration, the dataset was divided into 4 folds (80%) for training and 1 fold (20%) for testing. The models' performance was evaluated using three metrics, as mentioned in the previous section. The cross-validation results showed minimal variation with different k values, as indicated in Table 5.3.

Table 5.3: Model results through cross-validation test

Fold	**RMSE**	**MAE**	**R^2**
1	0.0043	0.0020	0.7670
2	0.0044	0.0026	0.7951
3	0.0041	0.0021	0.8153
4	0.0037	0.0025	0.7702
5	0.0035	0.0027	0.7899
Average	0.0040	0.0024	0.7875

5.3.2 Comparative Analysis of Machine Learning Models

The performance analysis compares the results of the proposed model to three ML models—KNN, DT, and RF—using the same 5-fold cross-validation on the identical dataset. Results from different values of K Are computed. However, Table 5.4 shows only the average results in terms of R^2, RMSE, and MAE values.

Table 5.4: Models' results comparison through cross-validation

Model	RMSE	MAE	R^2
KNN model:	0.0059	0.0048	0.6541
Decision Tree	0.0062	0.0050	0.6241
Random Forest:	0.0056	0.0045	0.6863
Proposed Model	0.0040	0.0024	0.7875

In Table 5.3, we examine the overfitting of the models using cross-validation consistency. the results are derived from 5-fold cross-validation, the Proposed Model's metrics (RMSE, MAE, and R^2) are consistent across all folds indicating the model is stable and likely to generalize well to new data therefore no overfitting.

Additionally, Figure 5.4 presents The results obtained through 5-fold cross-validation of all the models experimented. the dataset is split into five parts, with the model being trained on four parts and tested on the fifth, rotating through all parts. The Proposed Model's strong performance across these folds—reflected in its low RMSE (0.0040) and MAE (0.0024), and high R^2 (0.7875)—indicates that it performs well consistently across different subsets of the data, suggesting that its accuracy is not due to overfitting to specific data points.

Accuracy of the Model with RMSE and MAE

RMSE measures the average distance between the predicted values and the actual values. A lower RMSE indicates better accuracy. The MAE measures the average absolute difference between the predicted values and the actual values. Like RMSE, a lower MAE suggests better accuracy. The average difference between the predicted values and the actual values for the biogas yield is 0.0040, 0.0055, 0.0062, and 0.0059 for the proposed model, RF, DT, and KNN respectively. The average absolute difference between the predicted values and the actual biogas yield is 0.0024, 0.0044, 0.0049, and 0.0047 for the respective models as presented in Figure 5.5. Overall, the proposed method demonstrates the highest accuracy with the lowest RMSE and MAE values, followed by the RF and KNN models, while the DT model shows relatively lower accuracy.

Model Fit with R^2

Figure 5.6, presents a comparative analysis, using R^2 metrics. The graphs illustrated that different models had varying R^2 values. The RF model achieved an R^2 value of 0.6863, indicating that approximately 68.63% of the variance in the target variable can be explained by the model. The DT model obtained an R^2 value of 0.6240, indicating that approximately 62.40% of the variance in the target variable can be explained by this model. The KNN model achieved an R^2 value of 0.6540, indicating that approximately 65.40% of the variance in the target variable can be explained by this model. The proposed method obtained the highest R^2 value of 0.7808, indicating that approximately 78.08% of the variance in the target variable can be explained by this model. Overall, the proposed method demonstrates the highest model fit, with the highest R^2 values, indicating that it can better explain the variance in the target variable compared to the other models.

5.3.3 Variable Importance

The scatterplot provides valuable visualization of the relationships between various biogas parameters, though there is room for improvement in its design. Enhancements could include clearer labels for the axes and a reduction in the number of data points in Figure 5.5. The ordinate axis represents the moisture content of the biogas, measured in percentage, where a higher percentage indicates more moisture. The abscissa axis represents the temperature of the biogas, measured in degrees Celsius, with higher temperatures indicating hotter biogas. Each data point on the scatterplot corresponds to a single measurement of moisture and temperature for a specific biogas sample, with the color representing the gaz_change value (darker shades indicate higher values) and the size indicating the pH value (larger points denote higher pH).

From the scatterplot analysis, several observations emerge. Firstly, there is a general trend of increasing moisture content with rising temperature. Additionally, the gaz_change value appears to be negatively correlated with moisture content, suggesting that biogas with higher moisture content tends to have a lower gaz_change value. Lastly, there seems to be a positive correlation between pH value and temperature, indicating that biogas with higher temperatures tends to have higher pH values.

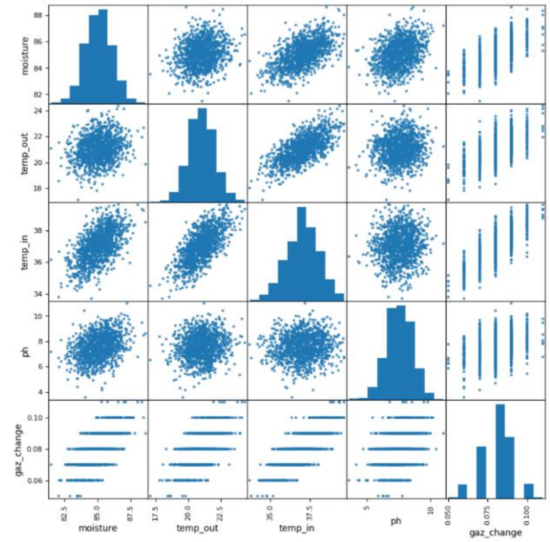


Figure 5.5: Comprehensive Insights: Scatterplot matrix analysis of the biogas dataset.

5.3.4 Deployment of the Proposed Model.

The research aims to integrate a robust and multi-dimensional methodology by combining IoT, and ML techniques to enhance biogas production. This comprehensive approach seeks to create a solution that is both practical and cost-effective for biogas operators, addressing their needs at local and industrial levels. By leveraging IoT for real-time data collection, mathematical modeling for energy harvesting optimization, and ML for predictive analytics, the proposed system aims to significantly improve the efficiency and yield of biogas production processes.

To achieve this, the research deploys an ML model on a Raspberry Pi, which is used to predict biogas production levels based on input parameters from the biodigester, such as temperature in/out, and pH. This setup in Figure 5.6, enables real-time processing and predicting at the edge, allowing for prompt adjustments to optimize biogas yield. The Raspberry Pi’s ability to handle these tasks locally ensures reduced latency and enhances the system’s responsiveness. In cases where the biodigester's performance deviates from expected patterns, the model can quickly identify and address non-functional issues, improving overall system reliability and efficiency in biogas production.

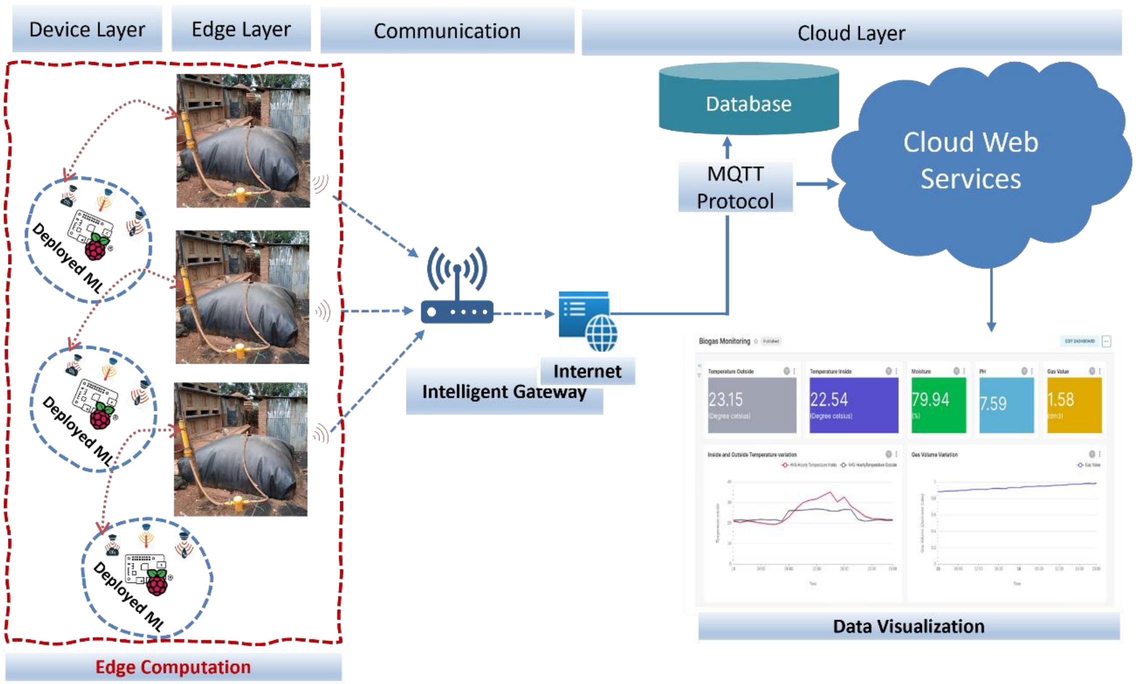


Figure 5.6: Model deployment on the edge node

5.3 Discussion and Conclusion

Biogas is a renewable and environmentally friendly energy source and has become increasingly important in the face of growing energy demands and the need for sustainable solutions. To effectively harness the potential of biogas, accurate prediction models are crucial for optimizing the production process.

This research has explored the use of hybrid ML techniques, as a promising approach for capturing the nonlinear relationships and dynamic effects inherent in the biogas production process. The proposed hybrid model adopted a stack assembling approach that combines the strengths of two modeling techniques LightGBM, CatBoost, and evolution strategy optimization algorithm to provide more accurate and robust predictions. The proposed model was trained from a real-world biogas production dataset. the model demonstrates improved performance compared to existing ML techniques. The findings indicated that the triadic ensemble model proposed significantly improves the accuracy. The proposed method outperforms the experimented models, achieving the lowest RMSE and MAE values. It also presents the highest R^2 value, indicating superior predictive accuracy.

This research contributes to the development of biogas forecasting models which represents a promising approach to improving the viability and sustainability of this renewable energy source. This advancement has significant implications for enhancing biogas operation and increasing energy output while addressing environmental challenges. As the demand for clean and reliable energy continues to grow, the continued advancement of these models will be crucial in unlocking the full potential of biogas generation.

During this research study, the following limitations have been identified: The accuracy is not as excellent as expected due to the limited dataset, The research experimented with the environmental factors as dependent variables, and there is a need to explore the combination of many other factors.

Chapter 6

Conclusions and Recommendations

6.1Conclusions

This thesis presents an integrated Intelligent Biogas Yield Optimization with a Control System to address biogas production inefficiency affected by inaccurate prediction of the input materials as well as inefficient control of environmental parameters during the AD process.

In this study, an edge computing IoT system was designed and developed within biogas production systems to manage biogas production and regulate environmental parameters. The IoT system has real-time monitoring capabilities enabling automated parameter adjustments while addressing operational challenges and improving the stability and efficiency of biogas production. This integration aligns with the current needs of biogas operators by providing a more responsive and adaptive approach to managing the biogas digestion process.

Additionally, an experimental analysis of solar power harvesting methods has identified effective strategies for powering IoT sensor nodes in biogas generation contexts. Approaches such as energy harvesting from environmental sources and implementing low-power sensor designs have proven suitable for ensuring reliable and sustainable operation. This analysis highlights the importance of sensor node energy budget and solar irradiation for the mathematical modeling of solar panels required for a particularly embedded system. The research results demonstrate that the solar energy proposed can facilitate the continuous operation of sensor nodes without relying on battery replacements.

Furthermore, the thesis includes development of an accurate forecasting system for biogas yield called the triadic hybrid prediction model which is a self-learning algorithm to customize the decision-making based on the environmental parameters behaviors. These results were simulated based dataset collected by the IoT device at home digester in Rwanda. The proposed model demonstrates high accuracy compared to other experimented models confirming its practical applicability.

In summary, this research attempts to design and develop an IoT-driven intelligent system equipped with an energy harvesting mechanism to manage the impact of environmental factors on biogas production. The project leverages various technologies, including ML, cloud services, web technology, and IoT, to make this development possible. This study developed the general solution in three connected studies:

(1) Design and develop the IoT-based architecture adopted in the biogas domain and ensure data acquisitions along with correlation analysis of the biodigester;

(2) Modeling the outdoor-based power harvesting mechanisms due to the countryside context where the developed model was deployed.

(3) The modeling, validation, and selection of the performing ML model along with integration into the IoT developed in (1).

The full-functioning solution is capable of gathering environmental parameters inside and outside of the biodigester at home digester with the integrated ML model to the edge node, the farmer can directly know the prediction of what shall be done or improved. At the same time, the data are sent to the authenticated cloud platform to support decision-makers. The key scope of the platform is as follows: (1) performing under internet-pruned conditions to send data to the cloud, data are sent in 16 seconds, which may dummy the server and consume many networks, we shall update based on the reason learned. (2) the prediction period of the biogas yield is in the month from the last month's historical data, which needs to be adjusted for the quarter, annual, and so on, (3) the model training was done once, which can’t consider the new upcoming cases in the future, we are planning to use pre-trained models to ensure the generalized cases.

6.2 Recommendations

From a biogas operation research perspective, there is no historical data that could facilitate research in the field of data modeling. Rwanda like other developing should rely on historical data for a better prediction of the future. Moreover, in the research journey, collaboration with industry partners is recommended as a crucial key to ensuring the availability and validation of emerging technology that brings solutions to the community.

The field of ML for energy management systems is still emerging, it introduces several aspects to intelligent systems. Here are some recommendations for future research works:

- Future research should focus on developing methodologies for sensor calibration to ensure consistency and reliability.
- A wider range of smart devices is required for monitoring other parameters.
- The energy harvesting model considers the combination of various factors.
- Exploring advanced prediction models that could consider a wider range of variables.

References

[1] "An updated roadmap to Net Zero Emissions by 2050 – World Energy Outlook 2022 – Analysis - IEA." Accessed: Aug. 31, 2024. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2022/an-updated-roadmap-to-net-zero-emissions-by-2050.

[2] S. Achinas, V. Achinas, and G. J. W. Euverink, "A Technological Overview of Biogas Production from Biowaste," Engineering, vol. 3, no. 3, pp. 299–307, Jun. 2017, doi: 10.1016/J.ENG.2017.03.002.

[3] E. Cervelli, E. Scotto di Perta, and S. Pindozzi, "Energy crops in marginal areas: Scenario-based assessment through ecosystem services, as support to sustainable development," Ecol. Indic., vol. 113, p. 106180, Jun. 2020, doi: 10.1016/J.ECOLIND.2020.106180.

[4] Z. Tshemese, N. Deenadayalu, L. Z. Linganiso, and M. Chetty, "An Overview of Biogas Production from Anaerobic Digestion and the Possibility of Using Sugarcane Wastewater and Municipal Solid Waste in a South African Context," Appl. Syst. Innov. 2023, Vol. 6, Page 13, vol. 6, no. 1, p. 13, Jan. 2023, doi: 10.3390/ASI6010013.

[5] C. H. Pham, C. C. Vu, S. G. Sommer, and S. Bruun, "Factors Affecting Process Temperature and Biogas Production in Small-scale Rural Biogas Digesters in Winter in Northern Vietnam," Asian-Australasian J. Anim. Sci., vol. 27, no. 7, p. 1050, 2014, doi: 10.5713/AJAS.2013.13534.

[6] M. M. Uddin and M. M. Wright, "Anaerobic digestion fundamentals, challenges, and technological advances," Phys. Sci. Rev., vol. 8, no. 9, pp. 2819–2837, Sep. 2023, doi: 10.1515/PSR-2021-0068/MACHINEREADABLECITATION/RIS.

[7] P. Onu, C. Mbohwa, and A. Pradhan, "Artificial intelligence-based IoT-enabled biogas production," 2023 Int. Conf. Control. Autom. Diagnosis, ICCAD 2023, 2023, doi: 10.1109/ICCAD57653.2023.10152349.

[8] S. Badri, M. Saini, and N. Goel, "Design of Energy Harvesting based Hardware for IoT Applications," Jun. 2023, Accessed: Jun. 19, 2024. [Online]. Available: https://arxiv.org/abs/2306.12019v1

[9] A. S. Adila, A. Husam, and G. Husi, "Towards the self-powered Internet of Things (IoT) by energy harvesting: Trends and technologies for green IoT," 2018 2nd Int. Symp. Small-Scale Intell. Manuf. Syst. SIMS 2018, vol. 2018-January, pp. 1–5, May 2018, doi: 10.1109/SIMS.2018.8355305.

[10] M. Bathre and P. K. Das, "Review on an energy efficient, sustainable and green internet of things," 2nd Int. Conf. Data, Eng. Appl. IDEA 2020, Feb. 2020, doi: 10.1109/IDEA49133.2020.9170736.

[11] H. Singh and G. Nanak, "Enhanced Energy Harvesting for IOT based Fuzzy Logics by using Gaussian Membership Functions Chanpreet Kaur," Int. J. Comput. Appl., vol. 180, no. 25, pp. 975–

8887, 2018.

[12] M. Pathak and R. Kumar, “Synchronous Inductor Switched Energy Extraction Circuits for Triboelectric Nanogenerator,” IEEE Access, vol. 9, pp. 76938–76954, 2021, doi: 10.1109/ACCESS.2021.3082499.

[13] S. Li et al., “Harvesting Thermal Energy through Pyroelectric and Thermoelectric Nanomaterials for Catalytic Applications,” Catal. 2024, Vol. 14, Page 159, vol. 14, no. 3, p. 159, Feb. 2024, doi: 10.3390/CATAL14030159.

[14] K. Wan et al., “Toward Self-Powered Sensing and Thermal Energy Harvesting in High-Performance Composites via Self-Folded Carbon Nanotube Honeycomb Structures,” ACS Appl. Mater. Interfaces, vol. 15, no. 37, pp. 44212–44223, Sep. 2023, doi: 10.1021/ACSAMI.3C08360/SUPPL_FILE/AM3C08360_SI_001.MOV.

[15] H. H. Ibrahim et al., “Radio Frequency Energy Harvesting Technologies: A Comprehensive Review on Designing, Methodologies, and Potential Applications,” Sensors (Basel)., vol. 22, no. 11, Jun. 2022, doi: 10.3390/S22114144.

[16] A. Mouapi, “Radiofrequency Energy Harvesting Systems for Internet of Things Applications: A Comprehensive Overview of Design Issues,” Sensors 2022, Vol. 22, Page 8088, vol. 22, no. 21, p. 8088, Oct. 2022, doi: 10.3390/S22218088.

[17] D. Hao et al., “Solar energy harvesting technologies for PV self-powered applications: A comprehensive review,” Renew. Energy, vol. 188, pp. 678–697, Apr. 2022, doi: 10.1016/J.RENENE.2022.02.066.

[18] V. Gopalakrishnan et al., “Energy Harvesting System with Solar Panels to Supply Low Power Electronic Devices,” IOP Conf. Ser. Earth Environ. Sci., vol. 1141, no. 1, p. 012008, Feb. 2023, doi: 10.1088/1755-1315/1141/1/012008.

[19] K. Malek et al., “Design and implementation of sustainable solar energy harvesting for low-cost remote sensors equipped with real-time monitoring systems,” J. Infrastruct. Intell. Resil., vol. 2, no. 3, p. 100051, Sep. 2023, doi: 10.1016/J.IINTEL.2023.100051.

[20] M. E. Thene and T. N. D. Mathaba, “Analysing the Design of a Solar Energy Harvesting Wireless Sensor Node for Agricultural Applications,” 5th Int. Conf. Artif. Intell. Big Data, Comput. Data Commun. Syst. icABCD 2022 - Proc., 2022, doi: 10.1109/ICABCD54961.2022.9856301.

[21] H. Sharma, A. Haque, and Z. A. Jaffery, “Maximization of wireless sensor network lifetime using solar energy harvesting for smart agriculture monitoring,” Ad Hoc Networks, vol. 94, p. 101966, Nov. 2019, doi: 10.1016/J.ADHOC.2019.101966.

[22] N. A. M. Amran, H. Mohamed, Z. F. M. Rafaai, N. S. Yacob, H. Junoh, and A. H. Shamsuddin, “A review of machine learning models in predicting biogas production,” AIP Conf. Proc., vol. 2934, no.

1, Mar. 2024, doi: 10.1063/5.0181016/3269825.

[23] J. Y. X. Ling et al., “Machine learning methods for the modelling and optimisation of biogas production from anaerobic digestion: a review,” Environ. Sci. Pollut. Res., vol. 31, no. 13, pp. 19085–19104, Mar. 2024, doi: 10.1007/S11356-024-32435-6/METRICS.

[24] S. Cinar, S. O. Cinar, N. Wieczorek, I. Sohoo, and K. Kuchta, “Integration of artificial intelligence into biogas plant operation,” Processes, vol. 9, no. 1. MDPI AG, pp. 1–18, 2021. doi: 10.3390/pr9010085.

[25] P. Sakiewicz, K. Piotrowski, J. Ober, and J. Karwot, “Innovative artificial neural network approach for integrated biogas – wastewater treatment system modelling: Effect of plant operating parameters on process intensification,” Renew. Sustain. Energy Rev., vol. 124, p. 109784, May 2020, doi: 10.1016/J.RSER.2020.109784.

[26] I. Angelidaki et al., “Biogas upgrading and utilization: Current status and perspectives,” Biotechnol. Adv., vol. 36, no. 2, pp. 452–466, Mar. 2018, doi: 10.1016/J.BIOTECHADV.2018.01.011.

[27] T. Beltramo, M. Klocke, and B. Hitzmann, “Prediction of the biogas production using GA and ACO input features selection method for ANN model,” Inf. Process. Agric., vol. 6, no. 3, pp. 349–356, Sep. 2019, doi: 10.1016/J.INPA.2019.01.002.

[28] S. Rasam, F. Talebkeikhah, M. Talebkeikhah, A. Salimi, and M. K. Moraveji, “Physico-chemical properties prediction of hydrochar in macroalgae Sargassum horneri hydrothermal carbonisation,” Int. J. Environ. Anal. Chem., vol. 101, no. 14, pp. 2297–2318, Nov. 2021, doi: 10.1080/03067319.2019.1700973.

[29] T. Beltramo and B. Hitzmann, “Evaluation of the linear and non-linear prediction models optimized with metaheuristics: Application to anaerobic digestion processes,” Eng. Agric. Environ. Food, vol. 12, no. 4, pp. 397–403, Oct. 2019, doi: 10.1016/J.EAEF.2019.06.001.

[30] Y. Yang, S. Zheng, Z. Ai, and M. M. M. Jafari, “On the Prediction of Biogas Production from Vegetables, Fruits, and Food Wastes by ANFIS- And LSSVM-Based Models,” Biomed Res. Int., vol. 2021, 2021, doi: 10.1155/2021/9202127.

[31] Y. Zhang et al., “Plant-scale biogas production prediction based on multiple hybrid machine learning technique,” Bioresour. Technol., vol. 363, p. 127899, Nov. 2022, doi: 10.1016/J.BIORTECH.2022.127899.

[32] M. Asadi and K. McPhedran, “Biogas maximization using data-driven modelling with uncertainty analysis and genetic algorithm for municipal wastewater anaerobic digestion,” J. Environ. Manage., vol. 293, p. 112875, Sep. 2021, doi: 10.1016/J.JENVMAN.2021.112875.

[33] K. Yoshida and N. Shimizu, “Biogas production management systems with model predictive control of anaerobic digestion processes,” Bioprocess Biosyst. Eng., vol. 43, no. 12, pp. 2189–2200, Dec.

2020, doi: 10.1007/S00449-020-02404-7/FIGURES/11.

[34] K. Solaun and E. Cerdá, "Climate change impacts on renewable energy generation. A review of quantitative projections," Renew. Sustain. Energy Rev., vol. 116, p. 109415, Dec. 2019, doi: 10.1016/J.RSER.2019.109415.

[35] "Improving energy access key to meeting development goals in Africa | UNCTAD." Accessed: May 04, 2024. [Online]. Available: https://unctad.org/news/improving-energy-access-key-meeting-development-goals-africa

[36] "Renewable energy jobs double to 13.7m in 10 years, IRENA finds | World Economic Forum." Accessed: May 04, 2024. [Online]. Available: https://www.weforum.org/agenda/2023/10/irena-renewable-energy-jobs/

[37] ["An introduction to biogas and biomethane – Outlook for biogas and biomethane: Prospects for organic growth – Analysis - IEA." Accessed: May 04, 2024. [Online]. Available: https://www.iea.org/reports/outlook-for-biogas-and-biomethane-prospects-for-organic-growth/an-introduction-to-biogas-and-biomethane

[38] [S. Sahota et al., "Review of trends in biogas upgradation technologies and future perspectives," Bioresource Technology Reports, vol. 1. Elsevier, pp. 79–88, Mar. 01, 2018. doi: 10.1016/j.biteb.2018.01.002.

[39] "Global production of biogas | Statista." Accessed: Jun. 16, 2024. [Online]. Available: https://www.statista.com/statistics/481791/biogas-production-worldwide/

[40] G. A. Mabea, "Electricity market coupling and investment in renewable energy: East Africa Community power markets," Int. J. Sustain. Energy, vol. 39, no. 4, pp. 321–334, Jan. 2020, doi: 10.1080/14786451.2019.1709461.

[41] U. Nations, Special issue on access to energy in s ub-Sahara n Africa Special issue on access to energy, no. 17.

[42] AfDB, "Rwanda Energy Sector Review and Action Plan," African Dev. Bank Gr., pp. 1–108, 2013, [Online]. Available: https://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/Rwanda_-_Energy_Sector_Review_and_Action_Plan.pdf

[43] E. U. E. I. EUEI, "Biomass Energy Strategy (BEST)," Biomass Energy Strateg. Rwanda-Volume 2backgr. Anal., vol. 2, no. June, pp. 1–114, 2009, [Online]. Available: https://www.undp.org/content/dam/uganda/docs/UNDPUg2014 - Biomass BEST Strategy(compressed).pdf

[44] "Low-Cost Polyethylene Tube Digester - energypedia." Accessed: Mar. 29, 2023. [Online]. Available: https://energypedia.info/wiki/Low-Cost_Polyethylene_Tube_Digester

[45] J. N. Mungwe, D. A. Asoh, and E. Mbinkar, "Design Considerations of a Flexible Biogas Digester

System for Use in Rural Communities of Developing Countries," J. Sustain. Bioenergy Syst., vol. 11, no. 04, pp. 260–271, 2021, doi: 10.4236/jsbs.2021.114016.

[46] M. H. Masud, M. Rashid, M. N. Hossan, and M. M. Ahmed, "Domestic Waste To Energy, Technologies, Economics, and Challenges," Ref. Modul. Earth Syst. Environ. Sci., 2023, doi: 10.1016/B978-0-323-93940-9.00026-8.

[47] F. Mapelli and J. N. Mungwe, "Modern Energies Services for Cooking: from Improved Cook-Stoves to Domestic and Community Biogas Based Systems," Renew. Energy Unleashing Sustain. Dev., pp. 43–74, Jan. 2013, doi: 10.1007/978-3-319-00284-2_3.

[48] A. Hajizadeh, "Biogas production by psychrophilic anaerobic digestion and biogas-to-hydrogen through methane reforming: experimental study and process simulation," 2021, doi: 10.48336/7S34-6Q14.

[49] K. O. Olatunji, N. A. Ahmed, and O. Ogunkunle, "Optimization of biogas yield from lignocellulosic materials with different pretreatment methods: a review," Biotechnol. Biofuels 2021 141, vol. 14, no. 1, pp. 1–34, Jul. 2021, doi: 10.1186/S13068-021-02012-X.

[50] N. Du, M. Li, Q. Zhang, M. D. Ulsido, R. Xu, and W. Huang, "Study on the biogas potential of anaerobic digestion of coffee husks wastes in Ethiopia," https://doi.org/10.1177/0734242X20939619, vol. 39, no. 2, pp. 291–301, Jul. 2020, doi: 10.1177/0734242X20939619.

[51] K. Paritosh, S. K. Kushwaha, M. Yadav, N. Pareek, A. Chawade, and V. Vivekanand, "Food Waste to Energy: An Overview of Sustainable Approaches for Food Waste Management and Nutrient Recycling," Biomed Res. Int., vol. 2017, 2017, doi: 10.1155/2017/2370927.

[52] M. Kelif Ibro, V. Ramayya Ancha, and D. Beyene Lemma, "Biogas Production Optimization in the Anaerobic Codigestion Process: A Critical Review on Process Parameters Modeling and Simulation Tools," J. Chem., vol. 2024, pp. 1–25, Apr. 2024, doi: 10.1155/2024/4599371.

[53] C. Carotenuto, G. Guarino, B. Morrone, and M. Minale, "Temperature and ph effect on methane production from buffalo manure anaerobic digestion," Int. J. Heat Technol., vol. 34, no. Special Issue 2, pp. S425–S429, 2016, doi: 10.18280/ijht.34S233.

[54] T. Zhang, C. Mao, N. Zhai, X. Wang, and G. Yang, "Influence of initial pH on thermophilic anaerobic co-digestion of swine manure and maize stalk," Waste Manag., vol. 35, pp. 119–126, Jan. 2015, doi: 10.1016/J.WASMAN.2014.09.004.

[55] R. Franqueto, J. D. da Silva, and M. Konig, "Effect of Temperature Variation on Codigestion of Animal Waste and Agricultural Residue for Biogas Production," Bioenergy Res., vol. 13, no. 2, pp. 630–642, Jun. 2020, doi: 10.1007/S12155-019-10049-Y/METRICS.

[56] V. Toutian, M. Barjenbruch, T. Unger, C. Loderer, and C. Remy, "Effect of temperature on biogas

yield increase and formation of refractory COD during thermal hydrolysis of waste activated sludge," Water Res., vol. 171, p. 115383, Mar. 2020, doi: 10.1016/J.WATRES.2019.115383.

[57] M. A. Rahman, R. Shahazi, S. N. B. Nova, M. R. Uddin, M. S. Hossain, and A. Yousuf, "Biogas production from anaerobic co-digestion using kitchen waste and poultry manure as substrate—part 1: substrate ratio and effect of temperature," Biomass Convers. Biorefinery, vol. 13, no. 8, pp. 6635–6645, Jun. 2023, doi: 10.1007/S13399-021-01604-9/TABLES/4.

[58] B. McNeil and L. M. Harvey, "Fermentation: A Practical Approach," Jan. 1990, doi: 10.1093/OSO/9780199630448.001.0001.

[59] N. Kesharwani and S. Bajpai, "Pilot scale anaerobic co-digestion at tropical ambient temperature of India: Digester performance and techno-economic assessment," Bioresour. Technol. Reports, vol. 15, p. 100715, Sep. 2021, doi: 10.1016/J.BITEB.2021.100715.

[60] S. Paudel, Y. Kang, Y. S. Yoo, and G. T. Seo, "Effect of volumetric organic loading rate (OLR) on H2 and CH4 production by two-stage anaerobic co-digestion of food waste and brown water," Waste Manag., vol. 61, pp. 484–493, Mar. 2017, doi: 10.1016/J.WASMAN.2016.12.013.

[61] L. S. Avinash and A. Mishra, "Enhancing biogas production in anaerobic digestion of MSW with addition of bio-solids and various moisture sources," Fuel, vol. 354, p. 129414, Dec. 2023, doi: 10.1016/J.FUEL.2023.129414.

[62] D. Li et al., "Effect of temperature on the anaerobic digestion of cardboard with waste yeast added: Dose-response kinetic assays, temperature coefficient and microbial co-metabolism," J. Clean. Prod., vol. 275, p. 122949, Dec. 2020, doi: 10.1016/J.JCLEPRO.2020.122949.

[63] M. Bharati, S. Shete, and N. P. Shinkar, "Anaerobic Digestion of Dairy Industry Waste Water-Biogas Evolution-A Review," Int. J. Appl. Environ. Sci., vol. 12, no. 6, pp. 1117–1130, 2017, Accessed: May 09, 2024. [Online]. Available: http://www.ripublication.com

[64] M. Wang et al., "A comparative study on Mesophilic and thermophilic anaerobic digestion of different total solid content sludges produced in a long sludge-retention-time system," Results Eng., vol. 19, p. 101228, Sep. 2023, doi: 10.1016/J.RINENG.2023.101228.

[65] M. Das Ghatak and P. Mahanta, "Effect of temperature on biogas production from lignocellulosic biomasses," Proc. 2014 1st Int. Conf. Non Conv. Energy Search Clean Safe Energy, ICONCE 2014, pp. 117–121, 2014, doi: 10.1109/ICONCE.2014.6808702.

[66] M. A. Dareioti, K. Tsigkou, A. I. Vavouraki, and M. Kornaros, "Hydrogen and Methane Production from Anaerobic Co-Digestion of Sorghum and Cow Manure: Effect of pH and Hydraulic Retention Time," Ferment. 2022, Vol. 8, Page 304, vol. 8, no. 7, p. 304, Jun. 2022, doi: 10.3390/FERMENTATION8070304.

[67] A. J. Kang, Q. Yuan, A. J. Kang, and Q. Yuan, "Enhanced Anaerobic Digestion of Organic Waste,"

Solid Waste Manag. Rural Areas, Sep. 2017, doi: 10.5772/INTECHOPEN.70148.

[68] A. Khalid, M. Arshad, M. Anjum, T. Mahmood, and L. Dawson, “The anaerobic digestion of solid organic waste,” Waste Manag., vol. 31, no. 8, pp. 1737–1744, Aug. 2011, doi: 10.1016/J.WASMAN.2011.03.021.

[69] J. Fernández-Rodríguez, M. Pérez, and L. I. Romero, “Comparison of mesophilic and thermophilic dry anaerobic digestion of OFMSW: Kinetic analysis,” Chem. Eng. J., vol. 232, pp. 59–64, Oct. 2013, doi: 10.1016/J.CEJ.2013.07.066.

[70] L. A. Fdez-Güelfo, C. Álvarez-Gallego, D. Sales, and L. I. Romero García, “Dry-thermophilic anaerobic digestion of organic fraction of municipal solid waste: Methane production modeling,” Waste Manag., vol. 32, no. 3, pp. 382–388, Mar. 2012, doi: 10.1016/J.WASMAN.2011.11.002.

[71] L. Jiunn-Jyi, L. Yu-You, and T. Noike, “Influences of pH and moisture content on the methane production in high-solids sludge digestion,” Water Res., vol. 31, no. 6, pp. 1518–1524, Jun. 1997, doi: 10.1016/S0043-1354(96)00413-7.

[72] M. C. Hernández-Berriel, L. Márquez-Benavides, D. J. González-Pérez, and O. Buenrostro-Delgado, “The effect of moisture regimes on the anaerobic degradation of municipal solid waste from Metepec (México),” Waste Manag., vol. 28, no. SUPPL. 1, pp. S14–S20, Jan. 2008, doi: 10.1016/J.WASMAN.2008.03.021.

[73] E. Leonidas, S. Ayvar-Soberanis, H. Laalej, S. Fitzpatrick, and J. R. Willmott, “A Comparative Review of Thermocouple and Infrared Radiation Temperature Measurement Methods during the Machining of Metals,” Sensors 2022, Vol. 22, Page 4693, vol. 22, no. 13, p. 4693, Jun. 2022, doi: 10.3390/S22134693.

[74] W. Root, T. Bechtold, and T. Pham, “Textile-Integrated Thermocouples for Temperature Measurement,” Mater. 2020, Vol. 13, Page 626, vol. 13, no. 3, p. 626, Jan. 2020, doi: 10.3390/MA13030626.

[75] L. Manjakkal, D. Szwagierczak, and R. Dahiya, “Metal oxides based electrochemical pH sensors: Current progress and future perspectives,” Prog. Mater. Sci., vol. 109, p. 100635, Apr. 2020, doi: 10.1016/J.PMATSCI.2019.100635.

[76] A. Nsair, S. O. Cinar, A. Alassali, H. A. Qdais, and K. Kuchta, “Operational Parameters of Biogas Plants: A Review and Evaluation Study,” Energies 2020, Vol. 13, Page 3761, vol. 13, no. 15, p. 3761, Jul. 2020, doi: 10.3390/EN13153761.

[77] X. Y. Li et al., “Model-based mid-infrared spectroscopy for on-line monitoring of volatile fatty acids in the anaerobic digester,” Environ. Res., vol. 206, p. 112607, Apr. 2022, doi: 10.1016/J.ENVRES.2021.112607.

[78] A. Tewari and B. B. Gupta, “Security, privacy and trust of different layers in Internet-of-Things

(IoTs) framework," Futur. Gener. Comput. Syst., vol. 108, pp. 909–920, Jul. 2020, doi: 10.1016/J.FUTURE.2018.04.027.

[79] B. Tonanzi et al., "Anaerobic digestion of mixed urban biowaste: The microbial community shift towards stability," N. Biotechnol., vol. 55, pp. 108–117, Mar. 2020, doi: 10.1016/J.NBT.2019.10.008.

[80] A. Villa-Henriksen, G. T. C. Edwards, L. A. Pesonen, O. Green, and C. A. G. Sørensen, "Internet of Things in arable farming: Implementation, applications, challenges and potential," Biosyst. Eng., vol. 191, pp. 60–84, Mar. 2020, doi: 10.1016/J.BIOSYSTEMSENG.2019.12.013.

[81] H. Mrabet, S. Belguith, A. Alhomoud, and A. Jemai, "A Survey of IoT Security Based on a Layered Architecture of Sensing and Data Analysis," Sensors 2020, Vol. 20, Page 3625, vol. 20, no. 13, p. 3625, Jun. 2020, doi: 10.3390/S20133625.

[82] C. Bayılmış, M. A. Ebleme, Ü. Çavuşoğlu, K. Küçük, and A. Sevin, "A survey on communication protocols and performance evaluations for Internet of Things," Digit. Commun. Networks, vol. 8, no. 6, pp. 1094–1104, Dec. 2022, doi: 10.1016/J.DCAN.2022.03.013.

[83] "Introduction to MQTT protocol for IOT applications. – Concurrency." Accessed: Jun. 16, 2024. [Online]. Available: https://concurrency.com/blog/introduction-to-mqtt-protocol-for-iot-applications/

[84] "CoAP Protocol: Key Features, Use Cases, and Pros/Cons | EMQ." Accessed: Jun. 16, 2024. [Online]. Available: https://www.emqx.com/en/blog/coap-protocol

[85] I. M. Insan and F. Samopa, "Implementation of Http Security Protocol for Internet of Things Based on Digital Envelope," Procedia Comput. Sci., vol. 234, pp. 1332–1339, Jan. 2024, doi: 10.1016/J.PROCS.2024.03.131.

[86] K. T. M. Tran, A. X. Pham, N. P. Nguyen, and P. T. Dang, "Analysis and Performance Comparison of IoT Message Transfer Protocols Applying in Real Photovoltaic System," Int. J. Networked Distrib. Comput., vol. 12, no. 1, pp. 131–143, Jun. 2024, doi: 10.1007/S44227-024-00021-4/TABLES/3.

[87] N. S. Chilamkurthy, O. J. Pandey, A. Ghosh, L. R. Cenkeramaddi, and H. N. Dai, "Low-Power Wide-Area Networks: A Broad Overview of Its Different Aspects," IEEE Access, vol. 10, pp. 81926–81959, 2022, doi: 10.1109/ACCESS.2022.3196182.

[88] P. Sharma and R. P. Singh, "Coverage hole identification & healing in Wireless Underground Sensor Networks," Meas. Sensors, vol. 24, p. 100540, Dec. 2022, doi: 10.1016/J.MEASEN.2022.100540.

[89] D. Sambo and A. Förster, "Wireless Underground Sensor Networks: A Comprehensive Survey and Tutorial," ACM Comput. Surv., vol. 56, no. 4, Oct. 2023, doi: 10.1145/3625388.

[90] M. Majid et al., "Applications of Wireless Sensor Networks and Internet of Things Frameworks in the Industry Revolution 4.0: A Systematic Literature Review," Sensors (Basel)., vol. 22, no. 6, Mar.

2022, doi: 10.3390/S22062087.
[91] A. Boualem, M. Ayaida, and C. De Runz, “Hybrid Model Approach for Wireless Sensor Networks Coverage Improvement,” Proc. - 2020 Int. Conf. Wirel. Networks Mob. Commun. WINCOM 2020, Oct. 2020, doi: 10.1109/WINCOM50532.2020.9272504.
[92] O. Ameri Sianaki, A. Yousefi, A. Rajabian Tabesh, and M. Mahdavi, “Internet of everything and machine learning applications: Issues and challenges,” Proc. - 32nd IEEE Int. Conf. Adv. Inf. Netw. Appl. Work. WAINA 2018, vol. 2018-January, pp. 704–708, Jul. 2018, doi: 10.1109/WAINA.2018.00171.
[93] “1 Introduction to Smart Energy Systems in Recent Trends | part of Applications of Big Data and Artificial Intelligence in Smart Energy Systems Smart Energy System: Design and its State-of-The Art Technologies : Volume 1 | River Publishers books | IEEE Xplore.” Accessed: May 15, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10137409
[94] S. Sambhi, S. Sambhi, and V. S. Bhadoria, “IoT-Based Optimized and Secured Ecosystem for Energy Internet: The State-of-the-Art,” Internet Things Bus. Transform. Dev. an Eng. Bus. Strateg. Ind. 5.0, pp. 91–125, Jan. 2021, doi: 10.1002/9781119711148.CH7.
[95] N. H. Motlagh, M. Mohammadrezaei, J. Hunt, and B. Zakeri, “Internet of things (IoT) and the energy sector,” Energies, vol. 13, no. 2. MDPI AG, 2020. doi: 10.3390/en13020494.
[96] D. DIeudonne and H. Shima, “Effectiveness of applying IoT to improve biogas digesters in Rwanda,” Int. Conf. Appl. Syst. Innov., pp. 441–444, Jun. 2018, doi: 10.1109/ICASI.2018.8394279.
[97] G. Xu, Y. Shi, X. Sun, and W. Shen, “Internet of Things in Marine Environment Monitoring: A Review,” Sensors 2019, Vol. 19, Page 1711, vol. 19, no. 7, p. 1711, Apr. 2019, doi: 10.3390/S19071711.
[98] J. Mabrouki, M. Azrour, G. Fattah, D. Dhiba, and S. El Hajjaji, “Intelligent monitoring system for biogas detection based on the Internet of Things: Mohammedia, Morocco city landfill case,” Big Data Min. Anal., vol. 4, no. 1, pp. 10–17, Mar. 2021, doi: 10.26599/BDMA.2020.9020017.
[99] M. Lowe, R. Qin, and X. Mao, “A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring,” Water 2022, Vol. 14, Page 1384, vol. 14, no. 9, p. 1384, Apr. 2022, doi: 10.3390/W14091384.
[100] E. T. de Camargo et al., “Low-Cost Water Quality Sensors for IoT: A Systematic Review,” Sensors 2023, Vol. 23, Page 4424, vol. 23, no. 9, p. 4424, Apr. 2023, doi: 10.3390/S23094424.
[101] A. Khudoyberdiev, I. Ullah, and D. Kim, “Optimization-assisted water supplement mechanism with energy efficiency in IoT based greenhouse,” J. Intell. Fuzzy Syst., vol. 40, no. 5, pp. 10163–10182, Jan. 2021, doi: 10.3233/JIFS-200618.
[102] Z. Parsa, R. Dhib, and M. Mehrvar, “Dynamic Modelling, Process Control, and Monitoring of

Selected Biological and Advanced Oxidation Processes for Wastewater Treatment: A Review of Recent Developments," Bioengineering, vol. 11, no. 2, 2024, doi: 10.3390/bioengineering11020189.

[103] Z. Shi, G. Ferrari, P. Ai, F. Marinello, and A. Pezzuolo, "Artificial Intelligence for Biomass Detection, Production and Energy Usage in Rural Areas: A review of Technologies and Applications," Sustain. Energy Technol. Assessments, vol. 60, no. November, p. 103548, 2023, doi: 10.1016/j.seta.2023.103548.

[104] "18 Cutting-Edge Artificial Intelligence Applications in 2024." Accessed: May 15, 2024. [Online]. Available: https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/artificial-intelligence-applications

[105] R. Abbasi, P. Martinez, and R. Ahmad, "The digitization of agricultural industry – a systematic literature review on agriculture 4.0," Smart Agric. Technol., vol. 2, p. 100042, Dec. 2022, doi: 10.1016/J.ATECH.2022.100042.

[106] M. Soori, B. Arezoo, and R. Dastres, "Artificial intelligence, machine learning and deep learning in advanced robotics, a review," Cogn. Robot., vol. 3, pp. 54–70, Jan. 2023, doi: 10.1016/J.COGR.2023.04.001.

[107] Z. Hao et al., "Physics-Informed Machine Learning: A Survey on Problems, Methods and Applications," Nov. 2022, Accessed: Jun. 18, 2024. [Online]. Available: https://arxiv.org/abs/2211.08064v2

[108] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Comput. Sci., vol. 2, no. 3, pp. 1–21, May 2021, doi: 10.1007/S42979-021-00592-X/FIGURES/11.

[109] J. Wang and F. Biljecki, "Unsupervised machine learning in urban studies: A systematic review of applications," Cities, vol. 129, p. 103925, Oct. 2022, doi: 10.1016/J.CITIES.2022.103925.

[110] A. K. Shakya, G. Pillai, and S. Chakrabarty, "Reinforcement learning algorithms: A brief survey," Expert Syst. Appl., vol. 231, p. 120495, Nov. 2023, doi: 10.1016/J.ESWA.2023.120495.

[111] J. M. Duarte and L. Berton, "A review of semi-supervised learning for text classification," Artif. Intell. Rev., vol. 56, no. 9, pp. 9401–9469, Sep. 2023, doi: 10.1007/S10462-023-10393-8/TABLES/13.

[112] D. K. Barupal and O. Fiehn, "Generating the blood exposome database using a comprehensive text mining and database fusion approach," Environ. Health Perspect., vol. 127, no. 9, pp. 2825–2830, 2019, doi: 10.1289/EHP4713.

[113] F. Kartal and U. Özveren, "Investigation of the chemical exergy of torrefied biomass from raw biomass by means of machine learning," Biomass and Bioenergy, vol. 159, p. 106383, Apr. 2022, doi: 10.1016/J.BIOMBIOE.2022.106383.

[114] Y. Liu, S. Liu, J. Li, X. Guo, S. Wang, and J. Lu, "Estimating biomass of winter oilseed rape using vegetation indices and texture metrics derived from UAV multispectral images," Comput. Electron. Agric., vol. 166, p. 105026, Nov. 2019, doi: 10.1016/J.COMPAG.2019.105026.

[115] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems," ACM Comput. Surv., vol. 55, no. 4, pp. 1–35, 2022, doi: 10.1145/3514228.

[116] F. ; Zhou et al., "Hybrid Model of Machine Learning Method and Empirical Method for Rate of Penetration Prediction Based on Data Similarity," Appl. Sci. 2023, Vol. 13, Page 5870, vol. 13, no. 10, p. 5870, May 2023, doi: 10.3390/APP13105870.

[117] J. Miao, P. L. Polak, C. Homescu, G. Kazantsev, A. Mullhaupt, and S. Uryasev, "Online Ensemble of Models for Optimal Predictive Performance with Applications to Sector Rotation Strategy," Mar. 2023, Accessed: Jun. 18, 2024. [Online]. Available: https://arxiv.org/abs/2304.09947v1

[118] [118] L. Wen and M. Hughes, "Coastal Wetland Mapping Using Ensemble Learning Algorithms: A Comparative Study of Bagging, Boosting and Stacking Techniques," Remote Sens. 2020, Vol. 12, Page 1683, vol. 12, no. 10, p. 1683, May 2020, doi: 10.3390/RS12101683.

[119] M. A. Yaman, F. Rattay, and A. Subasi, "Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Face Recognition," Procedia Comput. Sci., vol. 194, pp. 202–209, Jan. 2021, doi: 10.1016/J.PROCS.2021.10.074.

[120] E. Lin, C. H. Lin, and H. Y. Lane, "A bagging ensemble machine learning framework to predict overall cognitive function of schizophrenia patients with cognitive domains and tests," Asian J. Psychiatr., vol. 69, p. 103008, Mar. 2022, doi: 10.1016/J.AJP.2022.103008.

[121] C. Zhao et al., "BoostTree and BoostForest for Ensemble Learning," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 7, pp. 8110–8126, Jul. 2023, doi: 10.1109/TPAMI.2022.3227370.

[122] M. Koopialipoor, P. G. Asteris, A. Salih Mohammed, D. E. Alexakis, A. Mamou, and D. J. Armaghani, "Introducing stacking machine learning approaches for the prediction of rock deformation," Transp. Geotech., vol. 34, p. 100756, May 2022, doi: 10.1016/J.TRGEO.2022.100756.

[123] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, "The role of renewable energy in the global energy transformation," Energy Strategy Reviews, vol. 24, pp. 38–50, Apr. 2019, doi: 10.1016/J.ESR.2019.01.006.

[124] S. A. Sarker, S. Wang, K. M. M. Adnan, and M. N. Sattar, "Economic feasibility and determinants of biogas technology adoption: Evidence from Bangladesh," Renewable and Sustainable Energy Reviews, vol. 123, p. 109766, May 2020, doi: 10.1016/J.RSER.2020.109766.

[125] R. Inglesi-Lotz and E. Dogan, "The role of renewable versus non-renewable energy to the level of CO2 emissions a panel analysis of sub-Saharan Africa's Big 10 electricity generators," Renew Energy,

vol. 123, pp. 36–43, Aug. 2018, doi: 10.1016/J.RENENE.2018.02.041.

[126] R. Khoie, K. Ugale, and J. Benefield, “Renewable resources of the northern half of the United States: potential for 100% renewable electricity,” Clean Technol Environ Policy, vol. 21, no. 9, pp. 1809–1827, Nov. 2019, doi: 10.1007/S10098-019-01751-8/FIGURES/15.

[127] “Africa Energy Outlook 2019 – Analysis - IEA.” https://www.iea.org/reports/africa-energy-outlook-2019 (accessed Apr. 13, 2023).

[128] E. Press et al., NDCs in 2020: Advancing renewables in the power sector and beyond. 2019. [Online]. Available: www.irena.org

[129] B. Bharathiraja, T. Sudharsana, J. Jayamuthunagai, R. Praveenkumar, S. Chozhavendhan, and J. Iyyappan, “RETRACTED: Biogas production–A review on composition, fuel properties, feedstock and principles of anaerobic digestion,” Renewable and Sustainable Energy Reviews, vol. 90, pp. 570–582, Jul. 2018, doi: 10.1016/J.RSER.2018.03.093.

[130] “Biogas composition.” http://www.biogas-renewable-energy.info/biogas_composition.html (accessed Feb. 23, 2019).

[131] “Special Programs.” https://rab.gov.rw/index.php?id=134 (accessed Apr. 27, 2023).

[132] C. O. Onwosi et al., “Cattle manure as a sustainable bioenergy source: Prospects and environmental impacts of its utilization as a major feedstock in Nigeria,” Bioresource Technology Reports, vol. 19. Elsevier Ltd, Sep. 01, 2022. doi: 10.1016/j.biteb.2022.101151.

[133] J. P. Namahoro, Q. Wu, H. Xiao, and N. Zhou, “The asymmetric nexus of renewable energy consumption and economic growth: New evidence from Rwanda,” Renew Energy, vol. 174, pp. 336–346, Aug. 2021, doi: 10.1016/j.renene.2021.04.017.

[134] R. G. Hamid and R. E. Blanchard, “An assessment of biogas as a domestic energy source in rural Kenya: Developing a sustainable business model,” Renew Energy, vol. 121, pp. 368–376, Jun. 2018, doi: 10.1016/J.RENENE.2018.01.032.

[135] M. Silaen et al., “Lessons from Bali for small-scale biogas development in Indonesia,” Environ Innov Soc Transit, vol. 35, pp. 445–459, Jun. 2020, doi: 10.1016/J.EIST.2019.09.003.

[136] Y. Cai et al., “The absolute concentration and bioavailability of trace elements: Two vital parameters affecting anaerobic digestion performance of chicken manure leachate,” Bioresour Technol, vol. 350, p. 126909, Apr. 2022, doi: 10.1016/J.BIORTECH.2022.126909.

[137] I. Rocamora, S. T. Wagland, R. Villa, E. W. Simpson, O. Fernández, and Y. Bajón-Fernández, “Dry anaerobic digestion of organic waste: A review of operational parameters and their impact on process performance,” Bioresour Technol, vol. 299, p. 122681, Mar. 2020, doi: 10.1016/J.BIORTECH.2019.122681.

[138] S. Panigrahi and B. K. Dubey, “A critical review on operating parameters and strategies to improve

the biogas yield from anaerobic digestion of organic fraction of municipal solid waste," Renew Energy, vol. 143, pp. 779–797, Dec. 2019, doi: 10.1016/J.RENENE.2019.05.040.

[139] V. Toutian, M. Barjenbruch, T. Unger, C. Loderer, and C. Remy, "Effect of temperature on biogas yield increase and formation of refractory COD during thermal hydrolysis of waste activated sludge," Water Res, vol. 171, p. 115383, Mar. 2020, doi: 10.1016/J.WATRES.2019.115383.

[140] Q. Lin et al., "Temperature regulates deterministic processes and the succession of microbial interactions in anaerobic digestion process," Water Res, vol. 123, pp. 134–143, Oct. 2017, doi: 10.1016/J.WATRES.2017.06.051.

[141] Z. N. Abudi, Z. Hu, A. R. Abood, D. Liu, and A. Gao, "Effects of Alkali Pre-treatment, Total Solid Content, Substrate to Inoculum Ratio, and pH on Biogas Production from Anaerobic Digestion of Mango Leaves," Waste Biomass Valorization, vol. 11, no. 3, pp. 887–897, Mar. 2020, doi: 10.1007/S12649-018-0437-0/METRICS.

[142] C. Mao, Y. Feng, X. Wang, and G. Ren, "Review on research achievements of biogas from anaerobic digestion," Renewable and Sustainable Energy Reviews, vol. 45, pp. 540–555, May 2015, doi: 10.1016/J.RSER.2015.02.032.

[143] W. Zhang, Y. Xu, B. Dong, and X. Dai, "Characterizing the sludge moisture distribution during anaerobic digestion process through various approaches," Science of The Total Environment, vol. 675, pp. 184–191, Jul. 2019, doi: 10.1016/J.SCITOTENV.2019.04.095.

[144] "Artificial Intelligence (AI) in Renewable Energy Market Report 2022-2030." https://www.precedenceresearch.com/artificial-intelligence-in-renewable-energy-market (accessed Apr. 13, 2023).

[145] M. Barhamgi, D. A. Meedeniya, L. Mainetti, M. Aprile, E. Mele, and R. Vergallo, "A Sustainable Approach to Delivering Programmable Peer-to-Peer Offline Payments," Sensors 2023, Vol. 23, Page 1336, vol. 23, no. 3, p. 1336, Jan. 2023, doi: 10.3390/S23031336.

[146] E. Hitimana, G. Bajpai, R. Musabe, L. Sibomana, and J. Kayalvizhi, "Implementation of iot framework with data analysis using deep learning methods for occupancy prediction in a building," Future Internet, vol. 13, no. 3, pp. 1–20, Mar. 2021, doi: 10.3390/fi13030067.

[147] F. Valenzuela, A. García, E. Ruiz, M. Vázquez, J. Cortez, and A. Espinoza, "An IoT-based glucose monitoring algorithm to prevent diabetes complications," Applied Sciences (Switzerland), vol. 10, no. 3, Feb. 2020, doi: 10.3390/app10030921.

[148] R. Chaganti, V. Varadarajan, V. S. Gorantla, T. R. Gadekallu, and V. Ravi, "Blockchain-Based Cloud-Enabled Security Monitoring Using Internet of Things in Smart Agriculture," Future Internet, vol. 14, no. 9, Sep. 2022, doi: 10.3390/fi14090250.

[149] W. Lyu and J. Liu, "Artificial Intelligence and emerging digital technologies in the energy sector,"

Appl Energy, vol. 303, p. 117615, Dec. 2021, doi: 10.1016/J.APENERGY.2021.117615.

[150] R. Mishra, B. K. R. Naik, R. D. Raut, and M. Kumar, "Internet of Things (IoT) adoption challenges in renewable energy: A case study from a developing economy," J Clean Prod, vol. 371, p. 133595, Oct. 2022, doi: 10.1016/J.JCLEPRO.2022.133595.

[151] M. Logan, M. Safi, P. Lens, and C. Visvanathan, "Investigating the performance of internet of things based anaerobic digestion of food waste," Process Safety and Environmental Protection, vol. 127, pp. 277–287, Jul. 2019, doi: 10.1016/J.PSEP.2019.05.025.

[152] S. Cinar, S. O. Cinar, N. Wieczorek, I. Sohoo, and K. Kuchta, "Integration of artificial intelligence into biogas plant operation," Processes, vol. 9, no. 1. MDPI AG, pp. 1–18, 2021. doi: 10.3390/pr9010085.

[153] A. H. Abdurrahman, M. R. Kirom, and A. Suhendi, "Biogas Production Volume Measurement and Internet of Things based Monitoring System," 2020 IEEE International Conference on Communication, Networks and Satellite, Comnetsat 2020 - Proceedings, pp. 213–217, Dec. 2020, doi: 10.1109/COMNETSAT50391.2020.9328948.

[154] M. A. Aguida, S. Ouchani, and M. Benmalek, "An IoT-based Framework for an Optimal Monitoring and Control of Cyber-Physical Systems: Application on Biogas Production System," ACM International Conference Proceeding Series, pp. 143–149, Nov. 2021, doi: 10.1145/3494322.3494341.

[155] P. Ilangovan, M. Sharmila Begum, and P. K. Srividhya, "Development of online monitoring device and performance evaluation of biogas plants using enhanced methane prediction algorithm (EMPA)," Sustainable Energy Technologies and Assessments, vol. 56, p. 103041, Mar. 2023, doi: 10.1016/J.SETA.2023.103041.

[156] I. Pandian, S. Begum, and S. P. Kumaravel, "An integrated IoT and fuzzy logic controller system for biogas digester to predict methane generation," Environ Dev Sustain, pp. 1–13, Nov. 2021, doi: 10.1007/S10668-021-01943-7/METRICS.

[157] K. Sha, T. A. Yang, W. Wei, and S. Davari, "A survey of edge computing-based designs for IoT security," Digital Communications and Networks, vol. 6, no. 2, pp. 195–202, May 2020, doi: 10.1016/J.DCAN.2019.08.006.

[158] J. Shestovskaya, "MSc thesis Edge cloud architectures : a survey," 2020.

[159] M. Goudarzi, S. Ilager, and R. Buyya, "Cloud Computing and Internet of Things: Recent Trends and Directions," Internet of Things, pp. 3–29, 2022, doi: 10.1007/978-3-031-05528-7_1.

[160] S. Ahmad, I. Shakeel, S. Mehfuz, and J. Ahmad, "Deep learning models for cloud, edge, fog, and IoT computing paradigms: Survey, recent advances, and future directions," Comput. Sci. Rev., vol. 49, p. 100568, Aug. 2023, doi: 10.1016/J.COSREV.2023.100568.

[161] "The Republic of Rwanda AGRICULTURAL HOUSEHOLD SURVEY 2020 REPORT," 2021.

[162] “Rwamagana, Rwanda - Climate & Monthly weather forecast.” https://www.weather-atlas.com/en/rwanda/rwamagana-climate#temperature (accessed Mar. 29, 2023).

[163] C. A. Győrödi, D. V. Dumşe-Burescu, D. R. Zmaranda, and R. Győrödi, “A Comparative Study of MongoDB and Document-Based MySQL for Big Data Application Data Management,” Big Data and Cognitive Computing 2022, Vol. 6, Page 49, vol. 6, no. 2, p. 49, May 2022, doi: 10.3390/BDCC6020049.

[164] “MongoDB Documentation.” https://www.mongodb.com/docs/ (accessed Mar. 30, 2023).

[165] “The Easy-to-understand Guide to MQTT Protocol | EMQ.” https://www.emqx.com/en/mqtt (accessed Mar. 30, 2023).

[166] B. Dammak, M. Turki, S. Cheikhrouhou, M. Baklouti, R. Mars, and A. Dhahbi, “LoRaChainCare: An IoT Architecture Integrating Blockchain and LoRa Network for Personal Health Care Data Monitoring,” Sensors 2022, Vol. 22, Page 1497, vol. 22, no. 4, p. 1497, Feb. 2022, doi: 10.3390/S22041497.

[167] K. Alexopoulos, N. Nikolakis, and E. Xanthakis, “Digital Transformation of Production Planning and Control in Manufacturing SMEs-The Mold Shop Case,” Applied Sciences 2022, Vol. 12, Page 10788, vol. 12, no. 21, p. 10788, Oct. 2022, doi: 10.3390/APP122110788.

[168] A. Manowska, A. Wycisk, A. Nowrot, and J. Pielot, “The Use of the MQTT Protocol in Measurement, Monitoring and Control Systems as Part of the Implementation of Energy Management Systems,” Electronics 2023, Vol. 12, Page 17, vol. 12, no. 1, p. 17, Dec. 2022, doi: 10.3390/ELECTRONICS12010017.

[169] R. Chaganti, V. Varadarajan, V. S. Gorantla, T. R. Gadekallu, and V. Ravi, “Blockchain-Based Cloud-Enabled Security Monitoring Using Internet of Things in Smart Agriculture,” Future Internet 2022, Vol. 14, Page 250, vol. 14, no. 9, p. 250, Aug. 2022, doi: 10.3390/FI14090250.

[170] A. Hue, G. Sharma, and J. M. Dricot, “Privacy-Enhanced MQTT Protocol for Massive IoT,” Electronics 2022, Vol. 11, Page 70, vol. 11, no. 1, p. 70, Dec. 2021, doi: 10.3390/ELECTRONICS11010070.

[171] A. Kaushal and A. Shankar, “House Price Prediction Using Multiple Linear Regression,” SSRN Electronic Journal, Apr. 2021, doi: 10.2139/SSRN.3833734.

[172] “Ordinary Least Squares Method: Concepts & Examples - Data Analytics.” https://vitalflux.com/ordinary-least-squares-method-concepts-examples/ (accessed Jun. 09, 2023).

[173] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” PeerJ Comput Sci, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[174] H. Atmanspacher and M. Martin, “Correlations and How to Interpret Them,” Information 2019,

Vol. 10, Page 272, vol. 10, no. 9, p. 272, Aug. 2019, doi: 10.3390/INFO10090272.

[175] S. Atalla et al., “IoT-Enabled Precision Agriculture: Developing an Ecosystem for Optimized Crop Management,” Information 2023, Vol. 14, Page 205, vol. 14, no. 4, p. 205, Mar. 2023, doi: 10.3390/INFO14040205.

[176] C. Yang, W. Shen, and X. Wang, “Applications of Internet of Things in manufacturing,” in Proceedings of the 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2016, Institute of Electrical and Electronics Engineers Inc., Sep. 2016, pp. 670–675. doi: 10.1109/CSCWD.2016.7566069.

[177] Taryudi, D. B. Adriano, and W. A. Ciptoning Budi, “Iot-based Integrated Home Security and Monitoring System,” in Journal of Physics: Conference Series, Institute of Physics Publishing, Dec. 2018. doi: 10.1088/1742-6596/1140/1/012006.

[178] Y. Liu, C. Yang, L. Jiang, S. Xie, and Y. Zhang, “Intelligent Edge Computing for IoT-Based Energy Management in Smart Cities,” IEEE Netw, vol. 33, no. 2, pp. 111–117, Mar. 2019, doi: 10.1109/MNET.2019.1800254.

[179] M. J. B. Kabeyi and O. A. Olanrewaju, “Biogas Production and Applications in the Sustainable Energy Transition,” Journal of Energy, vol. 2022, pp. 1–43, Jul. 2022, doi: 10.1155/2022/8750221.

[180] “How IoT Works - 4 Main Components of IoT System - DataFlair.” Accessed: Dec. 06, 2023. [Online]. Available: https://data-flair.training/blogs/how-iot-works/

[181] P. V. Malaji, S. F. Ali, and G. Litak, “Energy harvesting: materials, structures and methods,” The European Physical Journal Special Topics 2022 231:8, vol. 231, no. 8, pp. 1355–1358, May 2022, doi: 10.1140/EPJS/S11734-022-00578-7.

[182] “Emerging Energy Harvesting Techniques for No-Battery Solutions.” Accessed: Jan. 22, 2024. [Online]. Available: https://www.electronicsforu.com/electronics-projects/electronics-design-guides/emerging-energy-harvesting-techniques-low-power-no-battery-solutions

[183] [9] S. Rabah, A. Zaier, J. Lloret, and H. Dahman, “Efficiency Enhancement of a Hybrid Sustainable Energy Harvesting System Using HHHOPSO-MPPT for IoT Devices,” Sustainability 2023, Vol. 15, Page 10252, vol. 15, no. 13, p. 10252, Jun. 2023, doi: 10.3390/SU151310252.

[184] [10] R. Gill and P. Chawla, “Energy Harvesting Sensors based Internet of Things System for Precision Agriculture,” Proceedings of 2nd International Conference on Innovative Practices in Technology and Management, ICIPTM 2022, pp. 270–273, 2022, doi: 10.1109/ICIPTM54933.2022.9754203.

[185] H. Elahi, K. Munir, M. Eugeni, S. Atek, and P. Gaudenzi, “Energy Harvesting towards Self-Powered IoT Devices,” Energies 2020, Vol. 13, Page 5528, vol. 13, no. 21, p. 5528, Oct. 2020, doi: 10.3390/EN13215528. H. Elahi, K. Munir, M. Eugeni, S. Atek, and P. Gaudenzi, “Energy Harvesting

towards Self-Powered IoT Devices," Energies 2020, Vol. 13, Page 5528, vol. 13, no. 21, p. 5528, Oct. 2020, doi: 10.3390/EN13215528.

[186] H. Sharma, A. Haque, and Z. A. Jaffery, "Maximization of wireless sensor network lifetime using solar energy harvesting for smart agriculture monitoring," Ad Hoc Networks, vol. 94, p. 101966, Nov. 2019, doi: 10.1016/J.ADHOC.2019.101966.

[187] A. Satharasinghe, T. Hughes-Riley, and T. Dias, "A review of solar energy harvesting electronic textiles," Sensors (Switzerland), vol. 20, no. 20. MDPI AG, pp. 1–39, Oct. 02, 2020. doi: 10.3390/s20205938.

[188] "5 Methods of Solar Energy Harvesting - Energy Theory." Accessed: Jan. 22, 2024. [Online]. Available: https://energytheory.com/5-methods-of-solar-energy-harvesting/

[189] [16] S. M. Antony, S. Indu, and R. Pandey, "An efficient solar energy harvesting system for wireless sensor network nodes," Journal of Information and Optimization Sciences, vol. 41, no. 1, pp. 39–50, Jan. 2020, doi: 10.1080/02522667.2020.1714182.

[190] O. B. Akan, O. Cetinkaya, C. Koca, and M. Ozger, "Internet of Hybrid Energy Harvesting Things," IEEE Internet Things J., vol. 5, no. 2, pp. 736–746, 2018, doi: 10.1109/JIOT.2017.2742663.

[191] H. Elahi, K. Munir, M. Eugeni, S. Atek, and P. Gaudenzi, "Energy Harvesting towards Self-Powered IoT Devices," Energies 2020, Vol. 13, Page 5528, vol. 13, no. 21, p. 5528, Oct. 2020, doi: 10.3390/EN13215528.

[192] H.-C. Yang, M.-S. Alouini, and M.-S. Alouini, "Characterizing Energy Efficiency of Wireless Transmission for Green Internet of Things: A Data-Oriented Approach," May 2018, Accessed: Jun. 19, 2024. [Online]. Available: https://arxiv.org/abs/1805.11725v2

[193] S. K. Ram, B. B. Das, A. K. Swain, and K. K. Mahapatra, "Ultra-low power solar energy harvester for IoT edge node devices," Proc. - 2019 IEEE Int. Symp. Smart Electron. Syst. iSES 2019, pp. 205–208, Dec. 2019, doi: 10.1109/ISES47678.2019.00053.

[194] [21] "What is Depth of Discharge and why is it so important? | Federal Batteries | Leading Battery Brands | The Best Battery Solutions." Accessed: Jan. 24, 2024. [Online]. Available: https://federalbatteries.com.au/news/what-depth-discharge-and-why-it-so-important.

[195] "What is a Solar Charge Controller altE." Accessed: Jan. 24, 2024. [Online]. Available: https://www.altestore.com/store/info/solar-charge-controller/

[196] "Three performance factors you must consider when sizing batteries | EEP." Accessed: Jan. 24, 2024 [Online]. Available:https://electrical-engineering-portal.com/performance-factors-sizing-batteries

[197] National Grid Group. What Is Biogas? Available online: https://www.nationalgrid.com/stories/energy-explained/what-is-biogas (accessed on 19 October

2023).

[198] Afotey, B.; Sarpong, G.T. Estimation of biogas production potential and greenhouse gas emissions reduction for sustainable energy management using intelligent computing technique. Meas. Sens. 2023, 25, 100650. https://doi.org/10.1016/j.measen.2022.100650.

[199] Kang, S.; Kim, G.; Jeon, E.-C. Ammonia Emission Estimation of Biogas Production Facilities in South Korea: Consideration of the Emission Factor Development. Appl. Sci. 2023, 13, 6203. https://doi.org/10.3390/app13106203.

[200] Saraswat, M.; Garg, M.; Bhardwaj, M.; Mehrotra, M.; Singhal, R. Impact of variables affecting biogas production from biomass. IOP Conf. Series: Mater. Sci. Eng. 2019, 691, 012043. https://doi.org/10.1088/1757-899x/691/1/012043.

[201] Malet, N.; Pellerin, S.; Nesme, T. Agricultural biomethane production in France: A spatially-explicit estimate. Renew. Sustain. Energy Rev. 2023, 185, 113603. https://doi.org/10.1016/j.rser.2023.113603.

[202] Bumharter, C.; Bolonio, D.; Amez, I.; Martínez, M.J.G.; Ortega, M.F. New opportunities for the European Biogas industry: A review on current installation development, production potentials and yield improvements for manure and agricultural waste mixtures. J. Clean. Prod. 2023, 388, 135867. https://doi.org/10.1016/j.jclepro.2023.135867.

[203] Sudiartha, G.A.W.; Imai, T.; Mamimin, C.; Reungsang, A. Effects of Temperature Shifts on Microbial Communities and Biogas Production: An In-Depth Comparison. Fermentation 2023, 9, 642. https://doi.org/10.3390/fermentation9070642.

[204] Møller, H.B.; Sørensen, P.; Olesen, J.E.; Petersen, S.O.; Nyord, T.; Sommer, S.G. Agricultural Biogas Production—Climate and Environmental Impacts. Sustainability 2022, 14, 1849. https://doi.org/10.3390/su14031849.

[205] Gopal, L.C.; Govindarajan, M.; Kavipriya, M.; Mahboob, S.; Al-Ghanim, K.A.; Virik, P.; Ahmed, Z.; Al-Mulhm, N.; Senthilkumaran, V.; Shankar, V. Optimization strategies for improved biogas production by recycling of waste through response surface methodology and artificial neural network: Sustainable energy perspective research. J. King Saud Univ.-Sci. 2021, 33, 101241. https://doi.org/10.1016/j.jksus.2020.101241.

[206] Induchoodan, T.G.; Haq, I.; Kalamdhad, A.S. Factors affecting anaerobic digestion for biogas production: A review. Adv. Org. Waste Manag. Sustain. Pract. Approaches 2022, 223–233. https://doi.org/10.1016/B978-0-323-85792-5.00020-4.

[207] Kunatsa, T.; Zhang, L.; Xia, X. Biogas potential determination and production optimisation through optimal substrate ratio feeding in co-digestion of water hyacinth, municipal solid waste and cow dung. Biofuels 2022, 13, 631–641. https://doi.org/10.1080/17597269.2020.1835452.

[208] Artificial Intelligence in Renewable Energy Market Size, Share 2023 to 2032. Available online: https://www.precedenceresearch.com/artificial-intelligence-in-renewable-energy-market (accessed on 20 October 2023).

[209] Shaw, R.N.; Ghosh, A.; Mekhilef, S.; Balas, V.E. Applications of AI and IOT in Renewable Energy; Elsevier BV: Amsterdam, The Netherlands, 2022; ISBN 9780323916998.

[210] Lyu, W.; Liu, J. Artificial Intelligence and emerging digital technologies in the energy sector. Appl. Energy 2021, 303, 117615. https://doi.org/10.1016/j.apenergy.2021.117615.

[211] Onu, P.; Mbohwa, C.; Pradhan, A. Artificial intelligence-based IoT-enabled biogas production. In Proceedings of the 2023 International Conference on Control, Automation and Diagnosis, ICCAD 2023, Rome, Italy, 10–12 May 2023. https://doi.org/10.1109/ICCAD57653.2023.10152349.

[212] Yang, Y.; Zheng, S.; Ai, Z.; Jafari, M.M.M. On the Prediction of Biogas Production from Vegetables, Fruits, and Food Wastes by ANFIS- and LSSVM-Based Models. BioMed Res. Int. 2021, 2021, 9202127. https://doi.org/10.1155/2021/9202127.

[213] Kour, V.P.; Arora, S. Particle Swarm Optimization Based Support Vector Machine (P-SVM) for the Segmentation and Classification of Plants. IEEE Access 2019, 7, 29374–29385. https://doi.org/10.1109/access.2019.2901900.

[214] Meza, J.K.S.; Yepes, D.O.; Rodrigo-Ilarri, J.; Rodrigo-Clavero, M.-E. Comparative Analysis of the Implementation of Support Vector Machines and Long Short-Term Memory Artificial Neural Networks in Municipal Solid Waste Management Models in Megacities. Int. J. Environ. Res. Public Health 2023, 20, 4256. https://doi.org/10.3390/ijerph20054256.

[215] Chen, W.-Y.; Chan, Y.J.; Lim, J.W.; Liew, C.S.; Mohamad, M.; Ho, C.-D.; Usman, A.; Lisak, G.; Hara, H.; Tan, W.-N. Artificial Neural Network (ANN) Modelling for Biogas Production in Pre-Commercialized Integrated Anaerobic-Aerobic Bioreactors (IAAB). Water 2022, 14, 1410. https://doi.org/10.3390/w14091410.

[216] Chiu, M.-C.; Wen, C.-Y.; Hsu, H.-W.; Wang, W.-C. Key wastes selection and prediction improvement for biogas production through hybrid machine learning methods. Sustain. Energy Technol. Assess. 2022, 52, 102223. https://doi.org/10.1016/j.seta.2022.102223.

[217] Renard, B.; Kavetski, D.; Kuczera, G.; Thyer, M.; Franks, S.W. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resour. Res. 2010, 46. https://doi.org/10.1029/2009wr008328.

[218] Amina, M.K.; Chithra, N.R. Predictive uncertainty assessment in flood forecasting using quantile regression. H2Open J. 2023, 6, 477–492. https://doi.org/10.2166/h2oj.2023.040.

[219] A. Nik-Khorasani, A. Mehrizi, and H. Sadoghi-Yazdi, “Robust hybrid learning approach

for adaptive neuro-fuzzy inference systems," *Fuzzy Sets Syst.*, vol. 481, p. 108890, Apr. 2024, doi: 10.1016/J.FSS.2024.108890.

[220] Gupta, A.; Jain, V.; Singh, A. Stacking Ensemble-Based Intelligent Machine Learning Model for Predicting Post-COVID-19 Complications. New Gener. Comput. 2022, 40, 987–1007. https://doi.org/10.1007/s00354-021-00144-0.

[221] Li, J.; Zhang, L.; Li, C.; Tian, H.; Ning, J.; Zhang, J.; Tong, Y.W.; Wang, X. Data-Driven Based In-Depth Interpretation and Inverse Design of Anaerobic Digestion for CH4-Rich Biogas Production.

[222] Meharie, M.G.; Mengesha, W.J.; Gariy, Z.A.; Mutuku, R.N. Application of stacking ensemble machine learning algorithm in predicting the cost of highway construction projects. Eng. Constr. Archit. Manag. 2022, 29, 2836–2853. https://doi.org/10.1108/ECAM-02-2020-0128/FULL/PDF.

[223] M. Lu et al., "A Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting," Water 2023, Vol. 15, Page 1265, vol. 15, no. 7, p. 1265, Mar. 2023, doi: 10.3390/W15071265.

[224] [134] A. Chatzimparmpas, R. M. Martins, K. Kucher, and A. Kerren, "Empirical Study: Visual Analytics for Comparing Stacking to Blending Ensemble Learning," Proc. - 2021 23rd Int. Conf. Control Syst. Comput. Sci. Technol. CSCS 2021, pp. 1–8, May 2021, doi: 10.1109/CSCS52396.2021.00008.

[225] H. Liao, X. Zhang, C. Zhao, Y. Chen, X. Zeng, and H. Li, "LightGBM: an efficient and accurate method for predicting pregnancy diseases," J. Obstet. Gynaecol. (Lahore)., vol. 42, no. 4, pp. 620–629, 2022, doi: 10.1080/01443615.2021.1945006.

[226] K. Kounlavong, L. Sadik, S. Keawsawasvong, and P. Jamsawang, "Optimized CatBoost-Based Soft-Computing Models for Prediction of the Ultimate Bearing Capacity of T-Shaped Footings Subjected to Eccentric Load," Arab. J. Sci. Eng. 2024, pp. 1–23, Aug. 2024, doi: 10.1007/S13369-024-09379-7.

[227]]Z. Li, X. Lin, Q. Zhang, and H. Liu, "Evolution strategies for continuous optimization: A survey of the state-of-the-art," *Swarm Evol. Comput.*, vol. 56, p. 100694, Aug. 2020, doi: 10.1016/J.SWEVO.2020.100694.

[228] [138] R. Selvam *et al.*, "Metaheuristic Algorithms for Optimization: A Brief Review," *Eng. Proc. 2023, Vol. 59, Page 238*, vol. 59, no. 1, p. 238, Mar. 2024, doi: 10.3390/ENGPROC2023059238.

[229] Li, J.; Zhang, L.; Li, C.; Tian, H.; Ning, J.; Zhang, J.; Tong, Y.W.; Wang, X. Data-Driven Based In-Depth Interpretation and Inverse Design of Anaerobic Digestion for CH4-Rich Biogas Production. ACS ES&T Eng. 2022, 2, 642–652. https://doi.org/10.1021/acsestengg.1c00316.

[230] Mukasine, A.; Sibomana, L.; Jayavel, K.; Nkurikiyeyezu, K.; Hitimana, E. Correlation Analysis Model of Environment Parameters Using IoT Framework in a Biogas Energy Generation Context. Futur. Internet 2023, 15, 265. https://doi.org/10.3390/fi15080265.

[231] Mapundu, M.T.; Kabudula, C.W.; Musenge, E.; Olago, V.; Celik, T. Explainable Stacked Ensemble Deep Learning (SEDL) Framework to Determine Cause of Death from Verbal Autopsies. Mach. Learn. Knowl. Extr. 2023, 5, 1570–1588. https://doi.org/10.3390/make5040079.

[232] Bansal, M.; Goyal, A.; Choudhary, A. A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. Decis. Anal. J. 2022, 3, 100071. https://doi.org/10.1016/j.dajour.2022.100071.

[233] KNN Algorithm | Latest Guide to K-Nearest Neighbors. Available online: https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/ (accessed on 29 November 2023).

[234] Atmanspacher, H.; Martin, M. Correlations and How to Interpret Them. Information 2019, 10, 272. https://doi.org/10.3390/info10090272.

[235] Decision Tree Algorithm—A Complete Guide—Analytics Vidhya. Available online: https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/ (accessed on 29 November 2023).

[236] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Adv. Neural Inf. Process Syst. 2017, 30. Available online: https://github.com/Microsoft/LightGBM (accessed on 28 October 2023).

[237] Wang, D.-N.; Li, L.; Zhao, D. Corporate finance risk prediction based on LightGBM. Inf. Sci. 2022, 602, 259–268. https://doi.org/10.1016/j.ins.2022.04.058.

[238] Li, Z.; Wang, W.; Ji, X.; Wu, P.; Zhuo, L. Machine learning modeling of water footprint in crop production distinguishing water supply and irrigation method scenarios. J. Hydrol. 2023, 625, 130171. https://doi.org/10.1016/j.jhydrol.2023.130171.

[239] Zhou, Y.; Wang, W.; Wang, K.; Song, J. Application of LightGBM Algorithm in the Initial Design of a Library in the Cold Area of China Based on Comprehensive Performance. Buildings 2022, 12, 1309. https://doi.org/10.3390/buildings12091309.

[240] How LightGBM Algorithm Works—ArcGIS Pro | Documentation. Available online: https://pro.arcgis.com/en/pro-app/latest/tool-reference/geoai/how-lightgbm-works.htm (accessed on 23 October 2023).

[241] How CatBoost Algorithm Works—ArcGIS Pro | Documentation. Available online: https://pro.arcgis.com/en/pro-app/latest/tool-reference/geoai/how-catboost-works.htm (accessed on 23 October 2023).

[242] Xiang, W.; Xu, P.; Fang, J.; Zhao, Q.; Gu, Z.; Zhang, Q. Multi-dimensional data-based medium- and long-term power-load forecasting using double-layer CatBoost. Energy Rep. 2022, 8, 8511–8522. https://doi.org/10.1016/j.egyr.2022.06.063.

[243] Wang, D.; Qian, H. CatBoost-Based Automatic Classification Study of River Network. ISPRS Int. J. Geo-Inf. 2023, 12, 416. https://doi.org/10.3390/ijgi12100416.

[244] Beyer, H.-G.; Schwefel, H.-P. Evolution strategies—A comprehensive introduction. Nat. Comput. 2002, 1, 3–52. https://doi.org/10.1023/a:1015059928466.

[245] Wang, Y.; Li, T.; Liu, X.; Yao, J. An adaptive clonal selection algorithm with multiple differential evolution strategies. Inf. Sci. 2022, 604, 142–169. https://doi.org/10.1016/j.ins.2022.04.043.

[246] Lange, R.T. evosax : JAX-Based Evolution Strategies. In Proceedings of the GECCO 2023 Companion—2023 Genetic and Evolutionary Computation Conference Companion, Lisbon Portugal, 15–19 July 2023; pp. 659–662. https://doi.org/10.1145/3583133.3590733.

[247] Performance Metrics in Machine Learning [Complete Guide]—Neptune.Ai. Available online: https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide (accessed on 29 November 2023).

[248] Metrics to Evaluate your Machine Learning Algorithm | by Aditya Mishra | Towards Data Science. Available online: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234 (accessed on 29 November 2023).