Intelligent Energy Management and Prediction of Micro Grid Operation Based On Machine Learning Algorithms and Genetic Algorithm

Mohamed Elweddad*, Muhammet Tahir Güneşer**,

* Department of Electrical and Electronic Engineering, Karabuk University, Karabuk, Turkey

** Department of Electrical and Electronic Engineering, Karabuk University, Karabuk, Turkey

(mohmeeedmali@gmail.com, mtguneser@karabuk.edu.tr,)

‡ Corresponding Author; Mohamed Elweddad, Karabuk, Turkey, Tel: +905467199910,

Fax: 0 (370) 418 7085 / 7085, mohmeeedmali@gmail.com

Received: 09.09.2022 Accepted: 27.10.2022

Abstract- Micro grid energy management has become critically important due to inefficient power use in the residential sector. High energy consumption necessitates developing a strategy to manage the power flow efficiently. For this purpose, this work has been divided into two phases: The first is the "ON/OFF" operation, which has been executed using a genetic algorithm for the hybrid system, including diesel generator, solar photovoltaic (PV), wind turbine, and battery. Then, in the second phase, the output results were used as input in three algorithms to predict load and supply dispatch one month ahead. This study has two objectives; the first is to decide which energy source should meet the load one month ahead. The second is to compare the outcomes of machine-learning techniques, namely Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbours (KNN), to determine the one that performs the best. The results indicated that the DT technique has the best performance in the application of classification with an accuracy of 100%. The findings also show that the RF approach gives acceptable results with an accuracy of up to 98%, and the KNN algorithm was poor in terms of accuracy with a value of 28%.

Keywords Renewable energy, Power management, Load classification, Machine learning algorithms.

1. Introduction

The world has been facing a severe electricity crisis. Globally, large industrial factories, airports, residential compounds, and high-rise buildings are built to improve the economic infrastructure. But, this has increased power demand, which has posed a great challenge for countries to satisfy their energy requirements [1]. The balance between the demand and supply is one of the factors that should be considered while building any hybrid energy system. However, it is possible to solve this problem through energy management and control systems using machine learning algorithms, optimization techniques, and artificial intelligence [2]. Therefore, energy system planning and optimization must perform simultaneously to ensure the high-efficiency design and low-cost operation [3]. to overcome the drawbacks of the traditional systems, several techniques have been implemented to obtain better performance. Intelligent energy management has gained the interest of several experts because it is very important for designing reliable systems [4]. Renewable energy is popular because of cheaper operations, lower maintenance costs, and simpler deployment than traditional generation systems [5]. Improving energy generation in micro grids would significantly reduce operating costs and technical problems. Furthermore, accurate energy consumption prediction is the base for designing a reliable system for building energy demand. In this regard, energy management and forecasting have recently received significant attention. Some investigations in this topic have been conducted utilizing various techniques, as reviewed in the following section.

1.1. Literature Review

Luu Ngoc et al. implemented an operational control approach using the fuzzy logic controller and a commonly-utilized ON-OFF controller to manage a PV-diesel generator-battery hybrid energy system. This method optimizes the battery and generator performance under all operating conditions [6]. Different renewable energy systems, both stand-alone and grid-connected, have been used a lot due to the low costs of installation, generation, maintenance, and operation. In addition, it's a response to scheduling and management programs. To evaluate 24-hour operational scheduling, the authors have designed a hybrid system equipped with PV cells and a diesel generator [7]. It is difficult to realize such an interactive, resilient, and sustainable model. In reference, a novel modelling and control prototype for smart grid integration of renewable energy sources is presented [8]. the primary objective of [9] is to design a Hybrid Energy System to fulfil the electricity needs of a large residential hostel for graduate students. Specific objectives included estimating energy demand and conducting economic and technical analyses of various combinations of renewable power sources that can meet the estimated energy demand. The main goal of study [10] is to achieve balance by arranging production, transmission, and consumption in real time. Because of this, it is very important to estimate power generation accurately. Meteorological and geological data from a wind power plant were used to estimate the amount of energy that could be made from wind. Artificial neural networks and the Adaptive Network Based Fuzzy Inference System were used to predict generation because of how well they could predict linear and nonlinear models. Reference [11] proposes a simple, highly reliable, and cost-effective PV-based water pumping system. The proposed system is modelled and tested using MATLAB/Simulink under different irradiation of the PV array to increase efficiency while decreasing cost and complexity. The model is tested with and without MPPT control, and the system efficiency is successfully evaluated.

Another study suggested a robust model for energy optimization to minimize hybrid system operating costs while managing power flow while taking increasing demands, intermittent solar radiation, and battery charge levels into account [12]. Selecting the optimal combination of several energy sources results in significant cost savings. In another study, the authors proposed an ideal system design, which included many configurations of renewable energy sources. They applied PSO, GA, TLBO, and BFPSO optimization algorithms to determine the best structure [13]. Dulout and Hernández et al. suggested integrating renewable energy sources with storage systems using a scheduling technique, considering essential factors affecting battery lifespan. They also conducted a cost evaluation to find out the optimal hourly storage capacity [14].

Reference [15] designs and implements a novel hybrid generation system including a solar system, energy storage, wind turbine, electrolytic cell, hydrogen gas turbine (HGT), and hydrogen energy storage. For economic evaluation, three scenarios are studied based on significant weather parameters and generation costs. The study [16] by Kiptoo and Adewuyi et al. presented a system that optimally combines various generating units, and there has been a lot of focus on prioritizing and cost-effectively scheduling. This system contains a photovoltaic (PV) array, a wind turbine (WT), and batteries. A MATLAB and MILP model were used to determine the optimal BESS, WT, and PV sizes to satisfy the system's demand while keeping costs as low as possible. In reference [17], To ensure sustainable energy transfer between sources and loads, the authors proposed a Rule-based energy management method for designing the best system configuration for an isolated micro grid.

Furthermore, in [18], the authors provided an ideal design of a hybrid system for micro grid applications that included some waste biomass and rooftop PV. They implemented a general optimization algorithm combining with a multi-perspective evaluation process. The proposed method was evaluated with data from Beijing, China. In a paper [19], the energy demand in Cyprus was predicted using machine learning methods such as linear regression, support vector machines (SVM), and artificial neural networks (ANN). The prediction models were modified using long-term and short-term analysis to forecast accurate data representing the power consumption in 2016 and 2107. The results indicated that SVM and ANN are better than other machine learning methods. Ref. [20] established accurate classification model. in order to achieve energy sustainability, building energy modelling is essential for improving energy efficiency. The researchers used machine learning techniques to categorize buildings depending on their energy efficiency.

Based on the fact that the main factor in hybrid system design is determining how power will be generated. The authors in [21] utilized Multi-Layer Perceptron (MLP) model to predict the expected power generation in the next 24 hours from the mirrored system. The researchers in [22] predicted solar power generation from a PV plant using a Recurrent Neural Network (RNN) algorithm, and the investigators in [23] forecasted PV and wind power production using three different techniques, including SVM, Radial Basis Function Neural Network (RBFNN), and Auto-Regressive Integrated Moving Average (ARIMA). Furthermore, constraints like the variety of consumption patterns require machine learning to understand and predict residential demand. Since load management is necessary in residential load, the author discussed it in details using machine learning algorithms [24]. When focusing on the accuracy of results, several algorithms can be utilized to evaluate short-, mid-, and long-term load forecasting such as multivariate adaptive regression splines, and linear regression techniques [25].

Energy sources' classification is essential for many energy management applications, so it was necessary to establish a reliable energy system before utilizing machine learning by applying simulations to a virtual server for micro grid control. The authors illustrated a flexible cloud-based architecture that enables real-time energy control [26]. Furthermore, energy generation prediction also involves weather data (such as wind speed, temperature, humidity, and solar radiation). But it's hard to accurately predict a building's energy consumption when so many factors are considered, such as its structure (insulation, glazing, window-to-wall ratio, orientation, number of residents, appliances, loads, and operational hours) [27]. Utilizing soft computing techniques for energy prediction is essential in solving these issues. It's significant to forecast electricity usage before establishing new power plants. For accurate power forecasting, SVR and MLP algorithms were investigated in paper [28]. In another study, authors tested a Random Forest-based prediction model for long-term energy usage forecast in several buildings using the hourly data collected in the form of a one-year dataset.

According to the results, the model showed high prediction accuracy [29]. The idea behind our work was inspired by a study [30]. Perera et al. tested some supervised algorithms, including Naive Bayes, multi-layer perceptron, RF, Bayesian network, DT, and Naive Bayes Tree, to classify five car traffic types. Classification accuracy was the highest when RF and DT algorithms were applied.

In general, it is difficult to determine which machine learning (ML) model is the best because the literature shows that all models provide high accuracy. To select the most appropriate model, it is essential to analyse the structure of the collected data and the application in detail. Machine learning enables rapid and accurate load forecasting and classification for energy management when input factors are unclear. Comparison between different ML models is available in the literature. As previously discussed, several studies compare the suggested ML approach against the traditional or basic ML models without giving adequate structural depth. To summarize, the literature review is subdivided into two parts: The first deals with micro grid energy optimization and management, and the second predicts energy, presenting key findings of a few papers based on supervised algorithms.

1.2. Paper Contribution

This work has been performed in two stages: First, the optimal scheduling of the hybrid energy system was found, and the mentioned sources are classified based on operational time. The study initially incorporates renewable energy sources and utilizes a genetic algorithm-based energy management model. Commercial solvers easily handle the model, determining the best options throughout the schedule (720 hours). Secondly, a reliable classification model for energy management is proposed. This model aims to make an operational decision to determine when the three sources are turned on or off based on the input data from the first part (genetic algorithm). Then we demonstrated how machine learning algorithms could perfectly categorize and forecast the operation time of energy sources. Finally, the proposed work was validated by comparing the accuracy of the results. According to our knowledge, no previous study has explored this topic using this methodology.

1.3. Significance of the Study

This study is important for four reasons:

1. Determine the best scheduling of energy sources to obtain the optimal operation for all the energy sources.

2. Use the developed classification models for another building under similar weather conditions and power system components. Based on the results, the model can determine the appropriate energy source for each hour based on the input data, such as temperature, wind speed, solar radiation, humidity, and wind and solar energies.

3. Obtain a reliable and tested model to save time and effort in deciding the energy sources required for other buildings.

4. Because Libya's state has set the aim of using only renewable technology to cover most of its power consumption, the plan includes replacing some fossil fuel and gas stations with a hybrid power system. These systems will be established as soon as possible in several parts of the country. For this reason, we decided to supply these buildings using renewable energy.

1.4. Research Gap and Paper Structure

This study proposes a novel combination of genetic algorithms and machine learning techniques for energy management. In comparison with previous studies in the literature review, several authors have utilized different algorithms to control and manage micro grids' power generation. The mentioned studies have mainly focused on power resource management (production) or consumption scheduling (demand) to minimize operational costs. Unlike previous studies, our study applies supervised algorithms (KNN, DT, and RF) for smart building energy system improvement and power source classification based on measured and forecasted data. Four different algorithms were used to manage energy sources. The first algorithm (GA) produces an optimal management plan for the energy output of each source to meet the demand for every hour. The second (classification) algorithm predicts the power dispatch of three energy sources throughout the month to ensure the sustainability of the supplied energy

This paper has been designed as follows: A description of the site and micro grid components is given in Section 2, and Section 3 shows the proposed methodology. Section 4 includes the results and discussion. Finally, Section 5 shows the conclusions.

2. System Description

In this work, 5 residential buildings were selected to study their daily power profile and energy needs throughout a whole month. The load data was collected in real-time from these buildings by using measurement devices. A hybrid system is used to supply the demand load of these buildings. The idea is to use the measured data to forecast the load pattern and arrange required power supply ideally to achieve end user comfort and power sustainable. The off grid hybrid system contains two renewable sources, Diesel generator and the batteries for storage system as seen in Fig. 1. And table 1 illustrates technical system components.

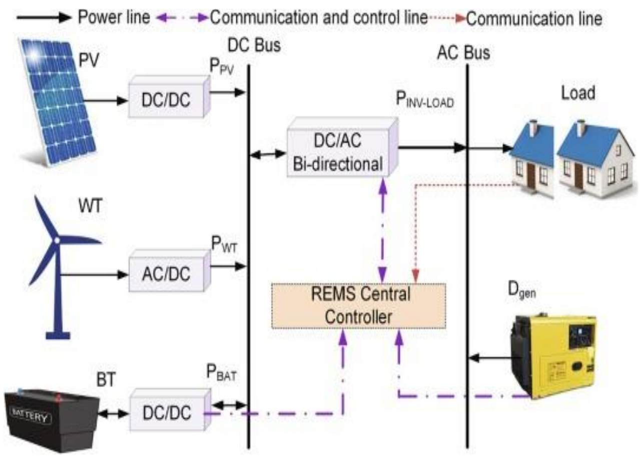

Fig. 1. System description of micro grid

Table 1. Micro grid parameters

Symbol	Description	Value	Unit
P_{pv}	Photovoltaic system-rated power	70	kW
P_{pw}	Wind system-rated power	150	kW
P_L	Load demand	85-247	kW
P_{batt}	Battery capacity	200	kWh
SOC_{max}	Maximum Battery system SOC	90	%
SOC_{min}	Minimum Battery system SOC	10	%
P_{DG}	Diesel generator-rated power	100	kW
Scheduling T	Simulation time	720	H
Classification T	Simulation time	720	H

2.1. Site Description

The study was conducted based on data obtained from five buildings located in Tripoli, Libya (32°00′17″N 11°19′51″E). A total of 260 residents belonging to 50 families live in the mentioned buildings. The selected location has a power outage every day. energy consumption data was obtained for one month as a case study to get more accurate results. Meteorological data, given in the figures, has been taken from the NASA website.

3. Methodology Description

As shown in Fig. 1, the system under investigation is an off-grid hybrid system consisting of batteries, a solar system, a diesel generator, and a wind system. It is utilized to provide power to a residential community. The highest demand is 210kW, and the lowest is 86kW. Hybrid energy sources (HES) were scheduled for more than 720 hours during the one-month period. The work is divided into two sections. The first section comprises a model for an energy source schedule that utilizes a genetic algorithm. It starts by entering the weather data and building electrical demand for one month (720 hours). The program then calculates the output power of renewable energy sources and the battery charge level. Based on the changing loads, the algorithm will select the most cost-effective energy sources to satisfy the demand for each hour, considering the available energy and the minimal cost. Then, the results will be used as the classification model's input data. Next, the model is trained on this data to help the algorithm understand the scheduling process based on the value of the load, the available power, and the weather data. Finally, based on all these inputs, the model will determine the energy sources that must supply the load per hour in other buildings in the same conditions. Fig. 2 shows the research methodology. The management and classification process is explained in detail in the following sections of the paper.

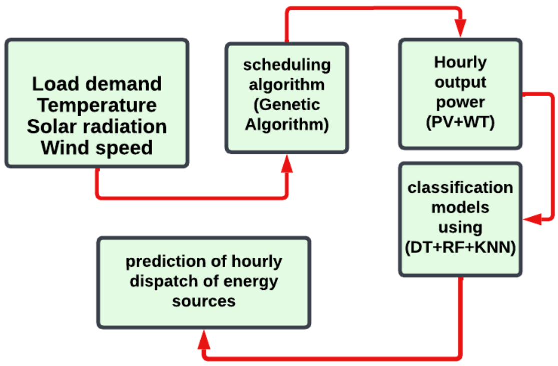

Fig. 2. The research methodology

3.1 Energy Sources Management

The hybrid system consists of a 150kW rated wind turbine (WT), 70kW PV, a 100kW diesel generator, and an energy storage system. An Energy Management System (EMS) is proposed to manage power-sharing among the specified sources based on hourly weather conditions and load demand. A Genetic Algorithm (GA) handles the EMS, making the decisions based on the system's running costs while prioritizing inputs from renewable sources. The following steps explain the scheduling methodology:

- Photovoltaic and wind energy are mostly utilized to supply loads.
- When renewable energy output exceeds demand, BESS charges with extra energy.
- BESS discharges when the PV system and WTs are insufficient.
- BESS injects a portion of power when a high consumption occurs or when the renewable resources' output not stable.
- Diesel generator provides power when PV, WT, and BESS systems are inadequate.
- When production exceeds demand and BESS is fully charged, extra renewable energy will be delivered to the grid.
- This method always prioritizes a low-cost energy.

The following are the limitations of the management system: As illustrated in Equation 1, the energy flow in the microgrid system must be continuously balanced.

$$P_L(t) = P_{PV}(t) + P_{Wt}(t) + P_{Bat}(t) + P_{DG}(t) \quad (1)$$

Equations (2) and (3) represent surplus and deficit power, respectively

$$P_{EX}(t) = \left[P_{pv}(t) + P_{wt}(t)\right] - P_L(t) \quad (2)$$

$$P_{Def}(t) = P_L(t) - [P_{pv}(t) + P_{wt}(t)] \quad (3)$$

$$P_{DG}(t) \leq P_{DG\,max} \quad (4)$$

Where $P_{PV}(t)$ is the PV power, $P_{Wt}(t)$ is the wind system power, $P_L(t)$ represents the demand power, $P_{Bat}(t)$ is the energy obtained from the battery, $P_{DG}(t)$ is the power the diesel generator delivers, $P_{EX}(t)$ is the surplus power, and $P_{Def}(t)$ is the deficit power.

3.1.1.The Proposed Scheduling Algorithm

The Genetic Algorithm is classified as a search heuristic because it is a computational search strategy to find approximate optimization solutions, specifically when a search is involved. It starts by creating a population of possible solutions, which is randomly done to ensure that the whole search space is covered. A chromosome is a possible solution defined by a collection of parameters known as genes. Each individual in the population is assigned a fitness value, and the population is evaluated. The selected individuals are subjected to crossover and mutation. It is the stage during which new people (children) are generated for the next generation. The process continues until the optimal option is found.

The methodology of the proposed GA-based energy scheduling model is explained in some steps below, as given in Fig. 3. The merit of the proposed algorithm is that it can integrate multiple renewable energy supply sources.

Step 1: Parameters initialization
Step 2: Input is forecasted for 720-hour energy demand
Step 3: Hourly input forecasts of solar irradiance and wind speed .
Step 4: Prepare an hourly output power generation report for RE sources.
Step 5: Input generator size and fuel price.
Step 6: Schedule power production share between various sources.
Step 7: Generate an optimal hourly solution.

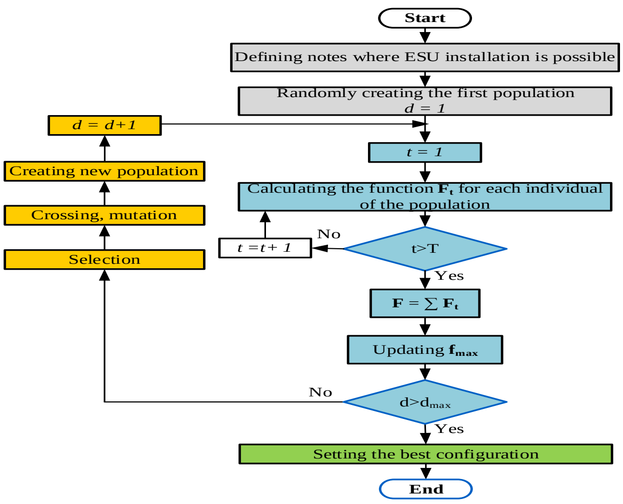

Fig. 3. The Genetic Algorithm flowchart

3.2. Energy Sources Classification

In the second stage of the study, three ML algorithms were implemented to classify the time of using energy sources to schedule their operations. The dataset was divided into two sets: 70% of measurements were devoted to training, and at the same time, the remaining 30% were kept for testing. The dataset includes 720 samples, with four inputs, seven attributes, and a single output. In this case, inputs include hourly consumed power, power generation through WT and PV, wind speed, ambient temperature, and solar data.

In the field of building energy analysis, supervised machine learning techniques have received significant interest. It is used to classify unlabeled data based on hidden patterns and feature similarities. This approach is beneficial for energy prediction applications, which essentially require energy requirement estimations for energy demand and supply in similar buildings; however, classification algorithms provide more precise tools for forecasting energy consumption than traditional methods. Fig. 4 depicts a flowchart of the complete classification method. The following section briefly explains the three algorithms used in this work.

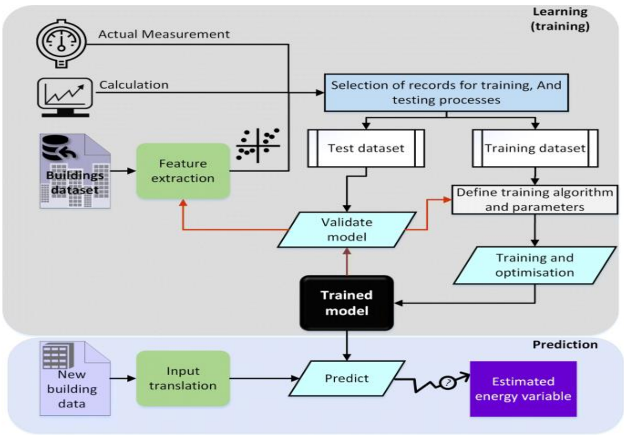

Fig. 4. Flowchart of the classification process

4.1. Random Forest (RF) Algorithm

RF algorithm is a type of ensemble learning method that can be used for classification, regression and other tasks. It works by creating numerous decision trees throughout the training process. Individual tree prediction is returned as the average predictions in regression tasks. When used for classification problems, the output of the RF is the class that has the greatest number of trees. RF is a classifier including a set of classifiers i.e., {h(x, ϴ_k), k = 1,...}, where k is an independent distributed random vector. At input x, each tree votes with one unit for the most popular class. The flowchart of RF algorithm is presented in Fig. 3. [31].

The random forest can be generated by following the instructions below:

Step 1: The algorithm selects random samples (k) from the specified dataset.

Step 2: For each chosen sample, the algorithm will generate a N number of decision tree. Then, each decision tree will obtain a prediction result.

Step 3: Voting will then be taken for each predicted output. It will employ scalp mode for a classification issue.

Step 4: Finally, the algorithm will choose the result with the most votes as the final prediction.

In the next paragraph, the mathematical equations that used in RF classification model will be explained. At first, a given a group of classifiers $h_1(x)$, $h_2(x)$, . . ., h_K (x), with set of randomly trained distribution vector (X,Y), determine the margin function as follows:

$$mg(X,Y) = av_k I(h_k(X) = Y) - max_{j\neq Y} av_k I(h_k(X) = j) \quad (5)$$

where I is defined as indicator function; the margin indicates how much the average number of votes for the right class at (X, Y) out numbers the average value for any class. When greater margin is obtained, it leads to more confidence in classification. The generalization error can be expressed as:

$$PE^* = P_{X,Y}(mg(X,Y) < 0) \quad (6)$$

The subscripts X, Y denote that the probability is over the space (X, Y). In random forests, $h_k(X) = h(X, \Theta_k)$. The Strong Law of tree structure is used for a larger number of trees.

The function of margin for an RF is,

$$mr(X,Y) = P_\Theta(h(X,\Theta) = Y) - max_{j\neq Y} P_\Theta(h(X,\Theta) = \mathrm{j}) \quad (7)$$

The strength of classifiers $\left(h(X,\Theta)\right)$ is given as:

$$s = E_{X.Y} mr(X,Y) \quad (8)$$

The raw of margin function is expressed as:

$$rmg(\Theta, X, Y) = I(h(X,\Theta) = Y) - I\left(h(X,\Theta) = j(X,Y)\right) \quad (9)$$

if the set of possible values of y is denoted by φ, minimizing E_{XY} (L (y, f (x))) for zero-one loss gives

$$f(x) = argmax_{y\in\varphi} P(Y = y.\ X = x \quad (10)$$

Finally, f (x) is the predicted class (“voting”) can be calculated by the following equation.

$$f(x) = argmax_{y\in\varphi} \sum_{j=1}^{J} I(y = h_j(x)) \quad (11)$$

4.2. Decision Tree (DT) algorithm

It is a supervised learning method in statistics, data mining, and machine learning applications. A predictive model such as a classification or regression decision tree is used to solve many complicated problems. The DT has many nodes, which further create a rooted and directed tree with no incoming edges. A node is called a "root." Each node has an incoming edge. A test/internal node has outgoing edges. The additional nodes are called leaves (also called decision or terminal nodes). In the DT, each internal node divides the instance space into two or more sub-spaces. It happens based on the input attribute values' discrete function. Usually, the goal is to discover the best DT by lowering the generalization error. Other objective functions include minimizing the average depth and reducing the number of nodes [32].

In the following paragraph, the processes involved in DT classification will be explained.

1- *Entropy*

A decision tree is constructed from the top downward, starting with a root node to include dividing data into subsets. Entropy using the frequency table of one attribute is given as:

$$E(S) = \sum_{i=1}^{c} -P_i \, log_2 P_i \quad (12)$$

where P_i is the probability of picking a data point from class i. The following equation represents the entropy using the frequency table of two attributes.

$$E(T,X) = \sum_{c\in X} P(c)\, E(c) \quad (13)$$

2- *Information Gain*

By splitting the examples according to the attributes, information gain calculates the expected reduction in entropy

$$Gain(S,A) Entropy(S) \sum_{V\in Values(A)} \left(\frac{|S_V|}{|S|}\right) Entropy(S_v) \quad (14)$$

where S is a collection of examples, A is an attribute, *Values (A)* is possible values of attribute A, S_v is the subset of S for which the attribute A has a value V.

The dataset is then segmented based on various criteria. Each branch's entropy is calculated. The total entropy for the split is then added proportionally. Before the split, the resulting entropy is deducted from the entropy. The information gain or decrease in entropy can be obtained from:

$$Gain(T,X) = Entropy(T) - Entropy(T,X) \quad (15)$$

3- *Gini*

Gini is a misclassification metric that is applied when data has several classifications. Gini is comparable to entropy; however, it is significantly faster to calculate. It is calculated as:

$$Gini = 1 - \sum_i P_i^2 \quad (16)$$

where i represents the class number

4-*Gini Impurity*

Gini Impurity is defined as the measure of variance of different classes

$$G(node) = \sum_{k=1}^{c} Pk(1 - Pk) \quad (17)$$

where P_k is the probability of picking a data point from class k

4.3. K-Nearest Neighbors (KNN) Algorithm

It is a common ML algorithm because it is simple and versatile. Since KNN uses all the training data, it needs significant time to read and save it. In the references section, benefits and flipsides of KNN have been provided [29]. KNN is useful for regression and classification. The input data is added as a set with the closest training samples (k) for both purposes. The KNN classification outcome is a class membership. An object falls into a category when it gets its neighbors' majority vote; for instance, an object may be allocated among its k closest neighbors in the most common class (k is a small and positive integer) [31].

Calculating distance:

Calculating the distance between each training point and the new point is the first step. There are several ways to determine this distance, the most popular is the Euclidian. Euclidean distance can be calculated as the square root of the sum of the squared differences between an existing point (y) and a new point (x) as can be seen in following equation

$$\text{Distance function}=\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (18)$$

4.4. The Algorithms Evaluation

Performance evaluation of a machine learning model is a significant step toward its development. Various measurements and comparison metrics can be used for this purpose. The following four steps validate the model's performance and reliability.

Accuracy: It is a popular evaluation method for classification tasks, and it is determined as follows [33]:

$$\text{Accuracy}= \frac{\mathbf{TP+TN}}{\mathbf{To}} \quad (19)$$

In this formula, TP implies true positive, T_O represents the total number of classifications, and TN means true negative. If the predicted value and actual output are "yes," TP is applied, and if both are "no," TN is used. Generally, total classification accuracy is unacceptable for evaluating the model performance when a dataset contains imbalanced data [38]. Furthermore, the algorithm may only identify one or two classes at a time, which means that a single class may be preferred over the others. The most effective model performance evaluation measures for imbalanced input data are mentioned below:

Precision: It is actually a proportion of the relevant TP samples, which are predicted to correspond to a specific class. Equation 13 states that the precision metric is determined by dividing TPs by the sum of TPs and False Positives (FPs) [33].

$$\text{Precision}=\frac{\mathbf{TP}}{\mathbf{TP+FP}} \quad (20)$$

Recall: It implies the ratio between correctly-predicted samples (TP) (which belong to a specific class of actual examples) and the total TPs and FNs [32]

$$\text{Recall}=\frac{\mathbf{TP}}{\mathbf{TP+FN}} \quad (21)$$

F1 Score: This metric shows a model's precision, resilience, and balance between recall and precision. It can be mathematically stated as [33]

$$\text{F1-score} = \frac{\mathbf{2}}{\frac{\mathbf{1}}{\mathbf{precision}}+\frac{\mathbf{1}}{\mathbf{recall}}} \quad (22)$$

5. Data Analysis

The scheduling dataset has 720 instances, seven features, a single output variable, and four input parameters. The four mentioned inputs are solar power, wind power, and hourly demand, while the output is hybrid energy source scheduling. The class type contains six attributes. Table 2 shows how these elements are encoded.

Table 2. Encoded values for different classes

Class	Class Encoding
Wind power + batteries	**1**
Wind power +Diesel generator	**2**
Wind power+Solar power+Battries	**3**
Solar power +Wind power	**4**
Solar +wind+Battries+Diesel generator	**5**
Wind power+Battries+Diesel generator	**6**
Diesel generator	**7**

5. Results and Discussion

In this section, only 24-hour data are analyzed to provide enough insight into the process of managing and scheduling energy sources. A hybrid PV-WT-DG-battery system and energy management system, utilizing the genetic algorithm, were proposed to supply loads of various residential buildings. This study used a typical load profile with a 210kW peak load. The generation systems are designed to supply peak loads effectively, but the intermittent nature of RES and consumer behavior may impact power sustainability and system stability. Fig. 5 depicts a building's load profile and renewable energy generation graphs. As can be noticed, there are peaks and drops in the load throughout the day. As a result, the output of power plants must be raised to satisfy demand during the peak hours (afternoon, early morning, and evening, i.e., from 7 pm to 12 pm and from 6 am to 8 am). When energy consumption is above average, the operators require additional production capacity to overcome the unexpected peaks. Energy costs will decrease due to less dependency on batteries and diesel generators. Utilizing renewable energy often lowers the cost of energy production. Customers can benefit from less energy costs during the off-peak hours of 9 am to 7 pm when energy prices are lower.

most of the time, load demand surpasses the output power of wind and solar systems together. Consequently, the generator and battery storage will be coupled to the load during periods of peak demand. On the other hand, there are periods in which renewable energy output exceeds load demand. In this situation, the excess energy will be used to charge the batteries. In order to lower the total cost of system operating and make power generation sustainable, the generator remains on standby during these periods. The ideal mix of these sources was then determined for each hour by solving a multi-objective function of energy sharing between the energy sources and demand in a single scenario.

The generated power of each energy source is shown in Fig. 5. Conforming to the solar radiation profile, the solar PV generator starts producing power at 08:00 and continues throughout the daytime. It is also clear that the PV system delivers almost its rated power 68 kw at 14:00, which corresponds to the maximum solar irradiance. It also applies to wind generation, and the total generated power from the hybrid system at a particular hour satisfies the load demand during that specific hour. For example, during the first three hours (01:00 am – 04:00 am), the wind turbine and the battery completely supply the demanded electricity, while the generator and PV do not contribute any power. The reason is that the PV system has no output power during that time.

Moreover, the peak demand of the residential building, as shown in Fig. 5, occurred at 17:00, which is perfectly fulfilled by the PV (42 kW), WT with 100kW, diesel generator (31 kW), and the battery storage (37 kW) and the lowest load measured at 3:00 am (86 kw) which supplied by (WT+BSS). Generally, energy production and consumption are always balanced, and all constraints are fulfilled. The power to be shared is mainly governed by the availability of the energy source and energy cost. If the load demand exceeds the aggregate energy generated from renewable sources, the generator becomes the viable option regardless of the energy price during that hour. Fig. 6 depicts the GA fitness function during the iteration process.

We mentioned that the maximum battery charging is up to 90% SoC, and the minimum discharge level is 20%; however, the generator is needed when the SoC drops to 20% and the renewable energy not enough; therefore, from 04:00 am to 07:00 am, the batteries are charged up solely from the generator because the power produced from WT is less than the load requirement, and insufficient to charge the batteries. Thus, between 08:00 am and 10:00 am, the storage system worked as a complementary energy source for PV and wind power systems and dilever 59 kw/h ; however, from 11:00 to 15:00, the hybrid PV-wind could meet the required energy demand without any support from the battery or the conventional generator.

From 11:00 am to 3:00 pm, the BSS restored its nominal capacity from the available renewable energy generation. Interestingly, the microgrid sold the surplus RE generation to the main grid during this period. The BSS remained in the discharge mode from 4:00 pm to 7:00 pm before charging up again at 20:00 for two hours. The maximum amount of electricity received from the generator was from 21:00 to 10:00 pm(100 kw), during which the RE was mainly received from the wind resource(110 kw), and the BSS reached complete exhaustion. To reduce generator dependency, the BSS is utilized again with the WT to cover approximately 67% of the load from 9:00 pm to 11:00 pm, as shown in Fig. 5.

From 3 am until 7 am, demand exceeds generation by 648 kilowatts, so the generator starts to fill the gap while the batteries recharge. After sundown, the BESS and the WT feed the load to their minimum SOC limit, decreasing the load placed on the generator, increasing efficiency, and saving money. With the help of the GA algorithm, we can observe that the generator shuts down for 10 hours a day, between the hours of 8 am and 4 pm, and that the total amount of electricity delivered by the generator drops to 618 KW.Due to the unreliable nature of RESs, it is essential to charge the battery to prevent load shedding or power outages in the next hours. Considering that the battery works at lower DOD values, extending its lifespan. It is assumed that the microgrid's battery storage initially would be charged to 90% SOC.

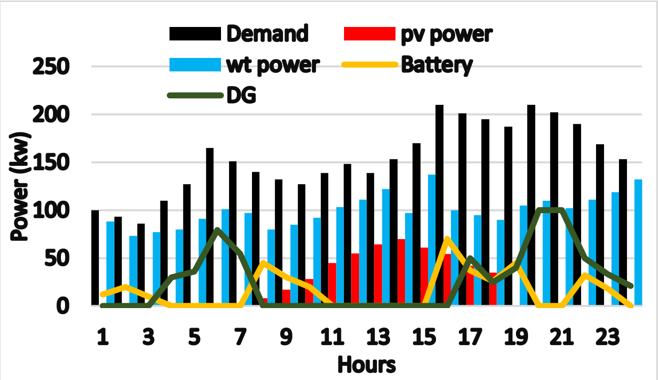

Fig. 5. Generated power from each energy source

During the day, battery charging and discharging have an impact on the overall price of energy. In this scenario, the batteries are charged between 3 am and 7 am, 11 am and 3 pm, and 8 pm and 10 pm so that their power may contribute to the generation of energy during peak hours, reducing reliance on the generator and wind system during those times. After integrating the storage systems (BESS), the EMS can optimize power sharing between the various energy sources and load cost-efficiently, thereby decreasing the need to purchase power from the generator. Additionally, the EMS with these storage systems can decrease unexpected energy fluctuations caused by changes in the weather.

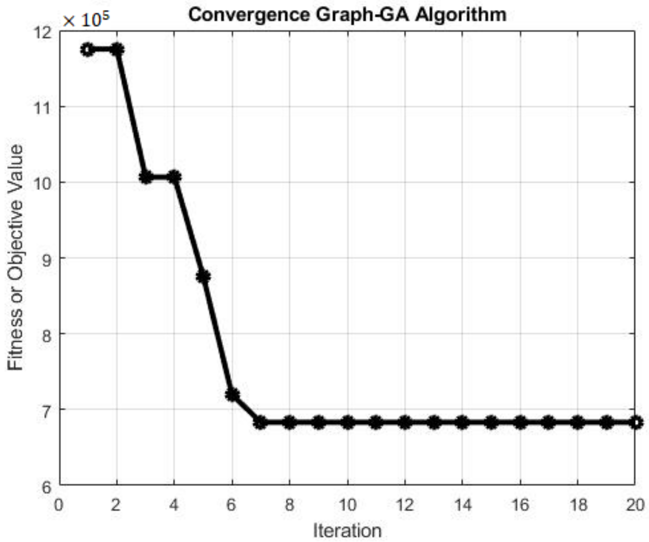

Fig. 6. Convergence of GA Algorithm fitness function

Fig. 7 illustrates the energy supply schedule based on dispatch hours throughout the month. It is evident that the WT-generator pair is the most utilized energy source (180 hours), followed by the PV-WT pair (150 hours), PV-WT-Bat combination (120 hours), WT-Bat combination (60 hours), and DG-Bat-WT combination (90 hours). The four available energy sources switched on concurrently for 60 hours only, making this combination the least utilized hybrid system. Finally, the diesel generator is used for about 60 hours. The GA algorithm decides the optimal power dispatch by prioritizing renewable energy to obtain the minimum electricity import from the microgrid.

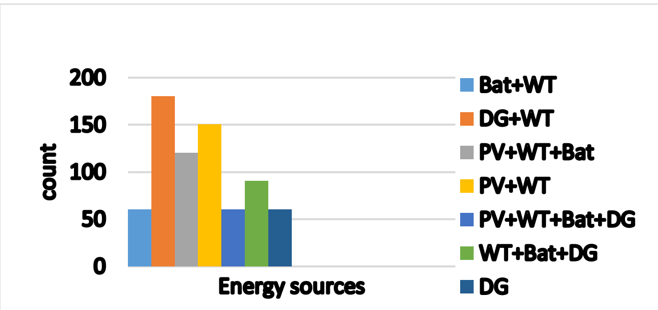

Fig. 7. Operating hours for each class

5.2 Classification Results

Classification algorithms are implemented for energy supply forecasting. Scheduling datasets are divided into two groups: 70% of the dataset helps training and 30% supports testing. After training the model on the output data from the study's first phase, the algorithm gets ready to manage any micro grid generation units that supply power to other buildings based on variable inputs, such as solar and wind energy, residential loads, battery capacity, and diesel generator power. The data was used to train three classification algorithms, after which the model can classify and predict the sources used to connect to the loads. Finally, the model was updated with new building data (available hourly energy and demand) for the next month using DT algorithm. The results demonstrate that the algorithm effectively classified the loads and forecasted the energy sources that must be operated to meet the demand.

Fig. 8 illustrates how regularly energy sources are switched on over a month. The figure shows that the algorithm selects the optimal combination of energy sources for every hour to ensure that loads are supplied and that there is never a power deficit. It has been noted that the Bat-WT was most frequently used (160 hours) throughout the whole month. Normally, energy is in excess in the middle of the day, mostly because of high PV energy output. The wind system is used most of the day and night, while the batteries and diesel generator make up for the energy deficit during the night. The figure also shows switching on/off (WT+DG) that takes place during 150 hours over a month. During some hours, the combination of Bat+WT cannot meet the load. In this case, a diesel generator only operates for 45 hours to compensate for the generation-consumption imbalance. The results also show that the PV system was switched on for 330 hours per month; this helped to reduce the total energy cost and CO2 emission. The results indicated that the classification algorithm managed energy sources in residential buildings with high efficiency and reliability. It is possible to utilize this model in any building with the same power sources. This process will be much easier if a small micro grid is constructed, which includes renewable energy sources and a generator is added to supply residential loads.

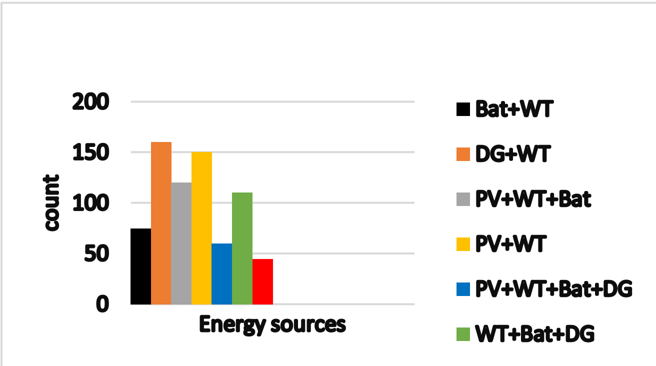

Fig. 8. Optimal combination of energy sources for every hour

5.3. Algorithm Evaluation

This section evaluates and analyzes the model's performance, and the results are compared for accuracy. Figs 9-11 provide the precision, F1-score, and recall measures to evaluate the ML algorithms' performances. Results show that DT is the best algorithm with 100% accuracy. KNN algorithm, on the contrary, showed the least accuracy. Figs show that the algorithms DT and RF have higher overall performance for all classes and bad performance for the KNN algorithm. It should be remembered that for a successful classifier, the recall should be as high as possible, or preferably, it should be one. That is only possible when the numerator and the denominator of Equation 20 are identical, which means that FN is zero. As the factor FN increases, the recall value decreases; therefore, a good classifier should ideally have a precision of 1. The precision value is equal to one only when FP equals zero. As shown in Figs 9, 10, and 11, the F1-score yields the same performance indicators as the precision and recall metrics. The attained results proved that the KNN algorithm does not efficiently classify.

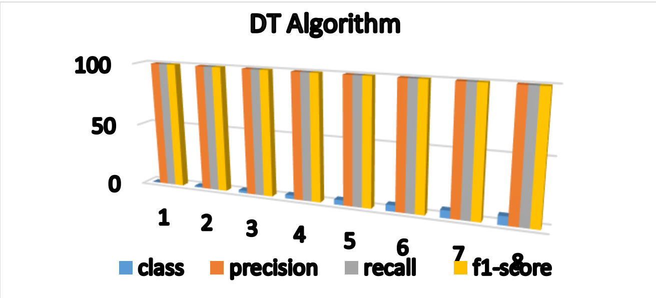

Fig. 9. DT algorithm evaluation measures

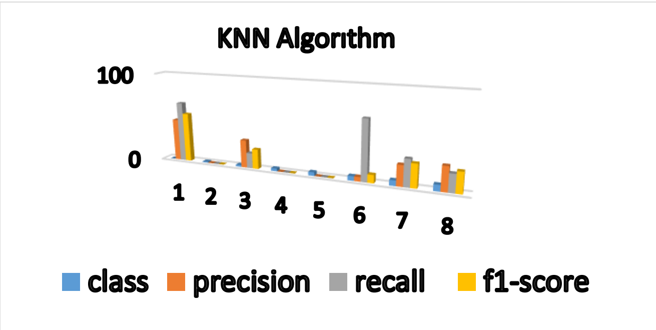

Fig. 10. KNN algorithm evaluation measures

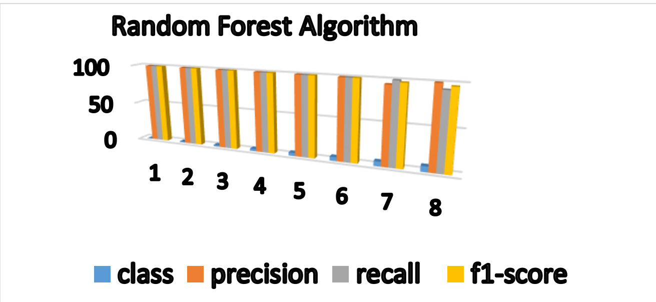

Fig. 11. RF algorithm evaluation measures

- *RF algorithm*

To forecast accurate classes, the RF algorithm first generates DTs, which have already been trained using data flows, and then various outputs are combined. For the best outcomes, we used entropy as a variable parameter. Fig. 10 shows that overall, RF was 98% accurate during classifying data. The evaluation results were as follows: Precision is 1.00 for six classes and 0.95 for class number 7, recall is 1.00 for all, F1-score is 1.00 for six classes, and 0.98 for class number 7, which is obvious in Table 4.

- *DT algorithm*

To forecast classes, the DT algorithm has several nodes and parameters before reaching its leaves. For better speed, we used the entropy metric. As a result, the DT method has the best accuracy of all the classification algorithms, which is 100%. Table 3 lists the accuracy, recall, and F1-score for each class acquired, while high measured values for most classes are around 1.00.

- *KNN algorithm*

In the KNN classifier, we implement the optimal distance for performance as well as accuracy. This approach classifies data by determining the nearest neighbors (k). Then, the prediction of the data classes is based on how far the neighbors are from each other. The efficiency of the KNN algorithm is determined by the kind of distance and the value of k. As indicated in Figure 5, we find the optimal value of k based on the accuracy, which is achieved for values of k ranging from 1 to 20. The results show that accuracy decreased as the number of neighbors increased. The algorithm achieved 28% average accuracy, which is lower than other techniques. Table 5 lists the recall, F1 score, and accuracy per class determined using the KNN. The evaluation results were relatively poor as follows: Precision between 0.05 to 0.47 and zero in Category (1), recall ranges from 0 to 0.67 in four classes (1, 3, 6, 7), and the rest of the three classes have (0) values. The F1-score has the best value in class one (0.55), and the rest of the classes are weak, including three classes(2, 4, 5) with value zero (0), which is similar to recall results. Table 2 lists the classification report of the KNN algorithm.

Table 2. Testing classification report of KNN algorithm

Class	Precision	Recall	F1-score	Support	Accuracy
1	0.47	0.67	0.55	39	28
2	0.00	0.00	0.00	12	
3	0.31	0.17	0.22	24	
4	0.00	0.00	0.00	10	
5	0.00	0.00	0.00	26	
6	0.05	0.67	0.09	3	
7	0.23	0.30	0.26	20	

Table 3. Testing classification report of DT algorithm

Class	Precision	Recall	F1-score	Support	Accuracy
1	1.00	1.00	1.00	39	100
2	1.00	1.00	1.00	12	
3	1.00	1.00	1.00	24	
4	1.00	1.00	1.00	10	
5	1.00	1.00	1.00	26	
6	1.00	1.00	1.00	3	
7	1.00	1.00	1.00	20	

Table 4. Testing classification report of RF algorithm

Class	Precision	Recall	F1-score	Support	Accuracy
1	1.00	1.00	1.00	39	98
2	1.00	1.00	1.00	12	
3	1.00	1.00	1.00	24	
4	1.00	1.00	1.00	10	
5	1.00	1.00	1.00	26	
6	0.95	1.00	0.98	20	
7	1.00	0.93	0.97	15	

Fig. 12 represents the heat map; it uses the correlation between inputs and outputs. The map displays the correlation between variables on a scale of 1 to -1, where 1 is the strongest and -1 is the lowest correlation. If the correlation coefficient is zero, there is no correlation between the variables, indicating a low dependency between model inputs and predicted outcomes. The evaluated parameters in the map include demand load, temperature, wind speed, solar irradiation, wind power, and solar power. According to correlation analysis, the model obtained high output power from both the PV and WT systems when both irradiation and wind speed were high.

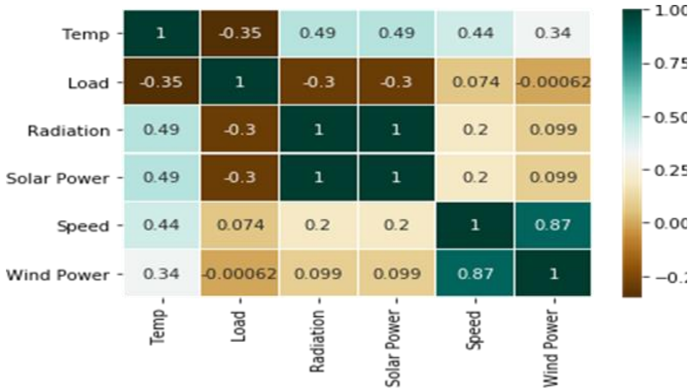

Fig. 12. Heat map

The graphic representation shows positive and negative correlations in some classes, indicating that the variables are directly and inversely related to each other. For example, the map demonstrates a strong correlation between photovoltaic output power and solar radiation, which reaches 1, and it is also good in the case of temperature, which is about 0.49. Additionally, the relationship between power generation and wind speed for wind energy is excellent, reaching 0.87. Generally, an improved model can be obtained by increasing the correlation between all the available variables to predict energy sources better. Finally, as shown in the correlation matrix analysis, all input and output variables have positive and negative correlation coefficients, which are not always strongly correlated.

For more details, to explain the algorithms' efficiency, confusion matrices were utilized, as depicted in Figure 13. The mentioned matrices were created to summarize the overall performance of the model. The columns represent actual results while predictions are represented by rows on the graph below, and correct predictions have been shown in red color. A confusion matrix summarizes the number of correct and deviated predictions generated by the classifier. A good model contains large values diagonally and smaller ones in the rest of the matrix. The matrices show that the diagonal elements are large while the rest of the cells are zero. It indicates that the model is accurate and has successfully read all the classifications, with up to 100% accuracy in the DT algorithm and 98% in the RF algorithm. In the case of the KNN algorithm, the numbers are the same between the diagonal and the rest of the matrix. For example, the first, sixth, and seventh columns contain greater-than-zero values. It indicates that the model is inaccurate and does not recognize all the classifications.

Class	**3**	**4**	**6**	**8**	**10**	**12**	**14**	**16**
3	2	0	6	0	0	4	0	0
4	11	0	4	0	0	9	0	0
6	1	0	0	0	0	5	4	0
8	11	0	0	10	0	8	7	0
10	0	0	0	0	26	2	0	1
12	3	0	0	0	0	4	6	7
14	1	0	0	0	0	2	9	3

A. Confusion matrix of KNN

class	**3**	**4**	**6**	**8**	**10**	**12**	**14**	**16**
3	39	0	0	0	0	0	0	0
4	0	12	0	0	0	0	0	0
6	0	0	24	0	0	0	0	0
8	0	0	0	10	0	0	0	0
10	0	0	0	0	26	0	0	0
12	0	0	0	0	0	3	0	0
14	0	0	0	0	0	0	20	0
16	0	0	0	0	0	0	0	14

B. Confusion matrix of RF

class	**3**	**4**	**6**	**8**	**10**	**12**	**14**	**16**
3	39	0	0	0	0	0	0	0
4	0	12	0	0	0	0	0	0
6	0	0	24	0	0	0	0	0
8	0	0	0	10	0	0	0	0
10	0	0	0	0	26	0	0	0
12	0	0	0	0	0	3	0	0
14	0	0	0	0	0	0	20	0
16	0	0	0	0	0	0	0	15

C. Confusion matrix of DT

Fig. 13. Confusion matrix of three algorithms A, B, and C

6. Conclusion

Energy generation and management are serious challenges for any country. Renewable energy sources are increasingly utilized as a source of energy. In such cases, micro grids operating as fully operational energy systems in various locations are becoming an attractive choice. The accurate prediction of power consumption by large consumer groups, such as residential, is necessary to implement a microgrid design successfully.This study aimed to develop a standard forecasting model for energy sources that should be turned on and off.

The first section of this paper discusses how the algorithm uses energy management data obtained from the genetic algorithm to determine energy consumption and the best power supply source to meet the loads. This strategy gives end users forecasting profiles for energy supply and demand, enabling them to enter the market for smart grids and alternative energy sources. This study's subject was the energy management system used in a standalone microgrid with variable parameters (demand loads, solar and wind power generation, a battery storage system, and a diesel generator). In the first phase, a rule-based methodology was established to choose the best use of energy sources based on forecasts for renewable resources. It is taken into consideration to decrease operational costs, maximize the use of renewable energy sources, and exchange power to the grid.

The power flow process in this work between loads and different sources shows how well the algorithm manages and optimizes the use of renewable resources. The second part of the study developed a novel methodology for forecasting the following month's energy requirement. Load profiles, scheduled energy data, and PV + WT output power were classified using machine learning techniques such as Decision Tree (DT), K-Nearest Neighbors (KNN), and Random Forest (RF) to predict which sources should be connected to the demand side throughout the month. The findings were confirmed by comparing the three models to determine the best. After analyzing the classification algorithms' output, it was discovered that the DT method performed remarkably better than the RF technique, while the KNN algorithm performed poorly. Furthermore, in such studies, the RF and DT algorithms proved to be accurate and robust. To obtain an improved model, the parameters of the algorithms were modified to achieve optimal accuracy, which improves system reliability and reduces training time.

Overall, among the three algorithms, the suggested model of the DT algorithm obtains the best accuracy (100%), followed by the RF algorithm (98%), while the KNN method gives the least accuracy (28%). Additionally, when it comes to the evaluations of precision and recall, the DT algorithm performs better than other algorithms. Finally, the results show how well the model predicted and classified energy sources. This study enables more optimized use of hybrid systems in micro grids; the operator can thus control the sources and time of use to achieve the best technical and economic results. Furthermore, unlike more complex hybrid algorithms, the proposed classification methods are simple, easy-to-use and implement

.

References

[1] M .Talha, M. S .Saeed., G. Mohiuddin, "optimization in home energy management system using artificial fish swarm and genetic algorithms". International conference on intelligent networking and collaborative systems. pp. 203-213.August, 2017.

[2] H.Wang, N.Zhang, Du, E., Yan, J., Y. "A comprehensive review for wind, solar, and electrical load forecasting methods". Global Energy Interconnection, 5(1), 9-30, May 2022.

[3] H. Aly,"A proposed intelligent short-term load forecasting hybrid models of ANN, WNN, and KF based on clustering techniques for smart grid". Electric Power Systems Research, 182, 106191,May (2020).

[4] H. Aly, "An intelligent hybrid model of neuro Wavelet, time series and Recurrent Kalman Filter for wind speed forecasting. Sustainable Energy Technologies and Assessments", 41,100802. October 2020.

[5] A.Kaabeche,R.Ibtiouen,"Techno-economic optimizationofhybridphotovoltaic/wind/diesel/battery generation in a stand-alone power system"", Solar Energy 103, pp. 171–182, May 2014.

[6] L.,N.An, T. Tuan, "Dynamic programming for optimal energy management of hybrid wind–PV–diesel–battery. Energies", 11(11), 3039, Otober 2018.

[7] H. Tazvinga, X .Xia, J. Zhang, "Minimum cost solution of photovoltaic–diesel–battery hybrid power systems for remote consumers". Solar Energy, 96, 292-299. Otober 2013.

[8] A. Shahid, "Smart Grid Integration of Renewable Energy Systems," 7th International Conference on Renewable Energy Research and Applications (ICRERA), pp. 944-948, October 2018

[9] I.Opedare, T.Adekoya, A.E. Longe, "Optimal sizing of hybrid renewable energy system for off-grid electrification: A case study of University of Ibadan Abdusalam Abubakar Post Graduate Hall of Residence". Int. J. Smart Grid-ijSmartGrid, 4(4), 176-189. December, 2020.

[10] A. Geçmez and Ç. Gençer, "Wind Energy Production Estimation with ANN and ANFIS," 9th International Conference on Smart Grid (icSmartGrid), , pp. 167-173, July 2021.

[11] H. Alrajoubi, S. Oncu and S. Kivrak, "An MPPT Controlled BLDC Motor Driven Water Pumping System," 10th International Conference on Renewable Energy Research and Application (ICRERA), pp. 116-119, September 2021

[12] M.Ouassaid,M. Maaroufi, "Smart home appliances modeling and simulation for energy consumption profile development: Application to Moroccan real environment case study". In 2016 International Renewable and Sustainable Energy Conference (IRSEC) (pp. 1050 1055). IEEE. November 2016.

[13] Y.Sawle, S.Gupta, A. Bohre, "Socio-techno-economic design of hybrid renewable energy system using optimization techniques". Renewable energy, 119, 459-472,April 2014.

[14] J.Dulout, A.Hernández., A.Anvari-A. "Optimal scheduling of a battery-based energy storage system for a microgrid with high penetration of renewable sources". In ELECTRIMACS (pp. 1-6). IMACS. Energy, 2013, 96, pp. 292–299.Julay 2017.

[15] M.Guo, W.Wang, R. Chen,. "Renewable hybrid energy system scheduling strategy considering demand response". Sustainable Energy Technologies and Assessments, 52, 102247, August 2022.

[16] M.Kiptoo, O. Adewuyi, M. Lotfy,"Harnessing demand-side management benefit toward achieving a 100% renewable energy microgrid". Energy Reports, 6, 680-685. February 2020.

[17] N.Attou, S.Zidi., M.Khatir, &S. Hadjeri, Energy Management System for Hybrid Microgrids. Electrotehnica, Electronica, Automatica, 69(2), 21-30, October 2021.

[18] S.Zheng, G. Huang, A. Lai, "Techno-economic performance analysis of synergistic energy sharing strategies for grid-connected prosumers with distributed battery storages". Renewable Energy, 178, 1261-1278. November 2021.

[19] D.Solyali, "A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus". Sustainability, 12(9), 3612. April 2020.

[20] C.Benavente-Peces, & N. Ibadah, "Buildings energy efficiency analysis and classification using various machine learning technique classifiers. Energies, 13(13), 3497. July (2020).

[21] J.Wasilewski, D. Baczynski, "Short-term electric energy production forecasting at wind power plants in Pareto-optimality context". Renewable and Sustainable Energy Reviews, 69, 177-187. March2017.

[22] T.Burianek, J.Stuchly, S. Misak, "Solar power production forecasting based on recurrent neural network". In Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA (pp. 195-204). January 2015.

[23] Y.Zhang, M. Beaudin, H. Zareipour, D. Wood," Forecasting solar photovoltaic power production at the aggregated system level". In 2014 North American Power Symposium (NAPS) (pp. 1-6). IEEE. November 2014.

[24] M.Sunny, M. Kabir, I.T. Naheen, "Residential energy management": A machine learning perspective. 2020 IEEE Green Technologies Conference (GreenTech) (pp. 229-234). IEEE. December 2020.

[25] G.Nalcaci, A. Özmen, G. Weber, "Long-term load forecasting: models based on MARS, ANN and LR methods". Central European Journal of Operations Research, 27(4), 1033-104, March 2018.

[26] D.G.Rosero, N.Díaz, C. Trujillo, "Cloud and machine learning experiments applied to the energy management in a microgrid cluster". Applied Energy, 304, 117770. December 2021.

[27] A.Ahmad, M. Hassan, A. Abdullah,. "A review on applications of ANN and SVM for building electrical energy consumption forecasting". Renewable and Sustainable Energy Reviews, 33, 102-109 , May 2014.

[28] P.Khan, Y.Byun, S. Lee, "learning-based approach to predict the energy consumption of renewable and nonrenewable power sources". Energies, 13(18), 4870, September 2020.

[29] A.Pham, N. Ngo, T. Truong, "Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability". Journal of Cleaner Production, 260, 121082. July 2020.

[30] J.Jupin, T.Sutikno, M. Ismail,"Review of the machine learning methods in the classification of a phishing attack". Bulletin of Electrical Engineering and Informatics, 8(4), 1545-1555. December 2019.

[31] L. Dong, P. Kang,"Prediction of rockburst classification using Random Forest" Transactions of Nonferrous Metals Society of China, 23(2), 472-477,February 2013.

[32] B.Charbuty, A.Abdulazeez, "Classification based on decision tree algorithm for machine learning". Journal of Applied Science and Technology Trends, 2(01), 20-28,April 2021.

[33] Hossin, Mohammad, and Md Nasir Sulaiman. et al. (2015) A review on evaluation metrics for data classification evaluations." International journal of data mining & knowledge management process (5).2 March 2015.