A Life-Long Learning XAI Metaheuristic-based Type-2 Fuzzy System for Solar Radiation Modelling

Majid Almaraashi, *Member, IEEE*, Mahmoud Abdulrahim, and Hani Hagras, *Fellow, IEEE*

***Abstract*—Solar photovoltaic (PV) power generation is one of the most important sources for renewable energy. However, PV power generation is entirely dependent on the amount of downward solar radiation reaching the solar cells. This is determined by uncertain and uncontrollable meteorological factors such as temperature, humidity, wind speed, and direction, as well as other factors such as topographical characteristics. Good solar radiation prediction models can increase energy output while decreasing the operation costs of photovoltaic power generation. For example, in some provinces in China, PV stations are required to upload short-term online power forecast information to power dispatching agencies. Numerous AI, statistical, and numerical weather prediction models have been used in many real-world renewable energy applications, with a focus on modeling accuracy. However, there is a need for Explainable AI (XAI) models that could be easily understood, analyzed, and augmented by the stakeholders. In this paper, we present a compact, explainable, and lifelong learning metaheuristic-based Interval Type-2 Fuzzy Logic System (IT2FLS) for Solar Radiation Modeling. The generated model will be composed of a small number of short IF-Then rules that have been optimized via simulated annealing to produce models with high prediction accuracy. These models are updated through a life-long learning approach to maximize their accuracy and maintain interpretability. In the process of lifelong learning, the proposed method transferred the model's knowledge to new geographical locations with minimal forgetting. The proposed method achieved good prediction accuracy and outperformed on new geographical locations other transparent and black-box models by 13.2% as well as maintaining excellent generalization ability. The resulting models have been evaluated and accepted by experts, and thanks to the generated transparency, the experts were able to augment the models with their expertise, which increased the models' accuracy.**

***Index Terms*—Solar Energy, type-2 fuzzy systems, XAI.**

I. INTRODUCTION

Renewable energy is the foundation of any energy transition that aims to achieve net-zero emissions and meet future energy demands [1]. According to the annual report of the International Energy Agency (IEA), renewable power capacity additions are on track to set new records in the coming years, led by solar PhotoVoltaic (PV) power generation. For example, more than half of all renewable energy expansion in 2021 came from solar PV alone, followed by wind and hydropower [1]. The global trend toward solar energy is driven by the fact that the Sun is an inexhaustible source of energy capable of meeting all of humanity's energy requirements [2].

Solar PV power generation is entirely dependent on the amount of downward solar radiation over solar cells, which is determined by uncertain and uncontrollable meteorological factors such as temperature, humidity, wind speed, and direction, as well as other factors such as topographical characteristics. Consequently, the power output of a PV system fluctuates over time due to the variable nature of environmental factors [3]. Good solar radiation prediction models can lower the operating and maintenance costs of photovoltaic power generation. For example, in some provinces in China, photovoltaic power stations are required to upload short-term online power forecast information to power dispatching agencies 0–3 hours in advance. Some stations are fined as much as $14 million each year for missing power assessments [4]. From an operational standpoint, solar radiation forecasting up to 2 hours in advance with a granularity of 30 seconds to 5 minutes is associated with ramping events and operational variability, whereas solar radiation forecasting one to six hours in advance with an hourly granularity is associated with load following forecasting [5].

Numerous AI, statistical, and numerical weather prediction models have been used in many real-world renewable energy applications, with a focus on modeling accuracy [6]. Black-box models such as neural networks [7], support vector regression [8], Gaussian process regression [9], and recently, deep learning models [10] have all been used to predict solar radiation. Comprehensive reviews of the applications of AI models in predicting solar radiation from an application perspective were presented in some studies, such as the reviews reported in [11] and [12]. On the other hand, fuzzy logic systems [13], and decision trees [9], [14] have an advantage over black-box models because domain experts can easily comprehend them and therefore evaluate, modify, and improve them based on their knowledge.

The lack of a transparent and interpretable process that yields model results is highlighted by many experts in different application domains as a barrier to the widespread adoption of AI models [15]. The presence of explainability can increase the level of trust between domain experts and AI models, allowing for more productive interactions. The significance of the model's explainability for the solar radiation prediction problem can be seen from a variety of perspectives. Firstly, ***understanding the relationship between surface solar radiation and its influencing factors is one of the most important requirements from an expert's perspective*** when

Majid Almaraashi is with the School of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia, and with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK.

Mahmoud Abdulrahim is with the Department of Meteorology, Faculty of Meteorology, Environment, and Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia. email: mahussen@kau.edu.sa.

Hani Hagras is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK.

evaluating model outputs [16]. In addition, due to the dynamic nature of solar radiation and underlying weather variables that differ from one location to another, ***in many situations, experts must modify the trained models to account for a new location characteristic to transfer the models to predict for other locations instead of training new models***. Another reason is the ***lack of representative long-term ground data sets for many locations, which hinders the trained model's ability to capture uncommon scenarios***. For instance, severe dust storms may occur once every few years but have a significant impact on the solar radiation level. It is imperative that an expert can add/delete/modify rules to an explainable model to capture these instances. The advantage of scientifically related numerical weather prediction (NWP) models over many black-box AI models in terms of explainability and traceability of a given prediction output is preferred or even required by domain experts to increase confidence in the prediction outcome and reduce potential losses. Furthermore, the lack of interpretability may limit subsequent solar operations deployments such as model predictive control (MPC) [17].

In this paper, we present a fully explainable and transferable type-2 fuzzy logic system to predict hourly solar radiation. The generated model can be easily understood, analyzed, and augmented by the experts and various stakeholders. The proposed system aims to achieve good prediction accuracy of hourly global solar radiation with a small set of short fuzzy rules so that the generated linguistic model is highly interpretable and transferable to similar locations. The model implemented a life-long learning approach, which allows it with the ability to learn and fine-tune its rules without the need to repeat the whole training process for each new location. The system was evaluated with real-world data sets for nine locations in Saudi Arabia with distinct climatic and topographical characteristics.

The paper is organized as follows: Section II reports on related background, while the proposed system is presented in Section III. The experiments and results are discussed in Section IV while the conclusions and future work are reported in Section V.

II. Related Background

A. XAI for Solar Energy Prediction

Some work in the literature has tried to produce interpretable AI models for solar prediction where [17] applied a model for predicting solar radiation based on a graph convolutional network (LSTM). The interpretability of the prediction process in terms of temporal and spatial dependencies was improved using the attention mechanism and the graph neural network. In [18], Light Gradient Boosting (LightGBM), which is based on decision trees, was used to predict global solar radiation. Then, Permutation Feature Importance (PFI) and Shapley Additive Explanations (SHAP) were used to get the explanation features in terms of feature importance. The latter method, known as feature importance, is widely used for explaining predictive machine learning algorithms. In order to estimate the global solar radiation, this approach seeks to determine the most relevant meteorological variables [18]. The work in [10] tried to build an interpretable deep learning framework in which the convolutional neural network (CNN) works as a trainable feature extractor and the support vector regressor (SVR) works as a global solar radiation predictor. This approach, which tries to explain predictions partially without clarifying the mechanism by which models function, is referred to as a "post-hoc" interpretation [19] cannot be fully understood, analyzed, and augmented by domain experts.

Mamdani fuzzy logic systems, on the other hand, are fully understandable in their structure and mechanism and can incorporate the expert's opinion into the prediction process rather than relying on post-processing and other techniques. Previous work [20], did employ TSK fuzzy logic systems for modelling photovoltaic array as well modeling solar radiation. However, they did not tackle the problems addressed in this paper related to the generation of compact Explainable AI models that can be easily analyzed and augmented by the expert and which are capable of lifelong learning and can be ported between various geographical locations since TSK models are not explainable [21].

B. A Brief Overview on Simulated Annealing

Simulated annealing (SA) is one of the most widely used meta-heuristic search methods for tackling combinatorial problems. The formulation for solving combinatorial optimization problems with simulated annealing can be illustrated as follows [22]: suppose S is a finite set of states, and C is the cost of each state in S. Combinatorial problems can be solved by searching the state space for the pair (S, C) with the lowest cost. SA creates the starting sequence of states $N = N_0$ from an initial state S_0 and an initial "temperature" T_0. The temperature is reduced after conducting a series of iterations, such as a nonstationary Markov Chain [23]. If a candidate state's cost is lower than the present state's cost, then it is accepted. There is still a threshold probability $Prob_T(S_j)$ at a specific temperature T_k at which the new candidate state S_j will be accepted if its cost is higher than the cost of the current state S_i as follows:

$$Prob_T(S_j) = \begin{cases} 1 & if\ f(s_j) \le f(s_i) \\ exp\left(\frac{f(s_i)-f(s_j)}{T_k}\right) & if\ f(s_j) > f(s_i) \end{cases} \quad (1)$$

Combinatorial minimization is the optimization of functions with several different discrete configurations and a finite number of solutions in the combinatorial neighborhood, as opposed to the continuous case, which has an infinite number of solutions. In the context of fuzzy logic systems optimization, fuzzy rules and variable selection processes are classified as combinatorial optimization, while the adaptation of continuous values of certain membership function parameters is classified as continuous optimization.

C. Lifelong learning and model transfer

In many instances, it is preferable to construct a framework for lifelong (long-term) learning that can handle continuous batch learning of data sets, particularly when model transfer is required. These systems should evolve and modify their structure dynamically to accept new information and be capable of incremental online adaptation [24]. Life-long learning can be combined with online learning for additional advantages. In addition to the ability of the model to learn incrementally, the model must maintain knowledge retention, which means being resistant to the catastrophic forgetting problem and capable of

refining prior experience when acquiring new information [25]. Another issue that might arise with lifelong learning of fuzzy logic systems is related to the expansion and redundancy of the rule base. When the rule base is derived from data, it may include duplicate or unnecessary rules [26]. In the field of solar radiation prediction, using some streams, such as online meteorological services, it is possible to increase the accuracy of data-driven forecasts with the availability of additional historical records [27]. ***The ability to transfer pretrained models to different places without the need for intensive training processes or large amounts of historical data is another motivation***. Nevertheless, additional precautions should be taken when using pretrained models because of the complexity of the solar radiation problem, which involves different complex patterns across locations. For instance, the analysis for the prediction of a pretrained model on a subset of the available data, if it already exists, can be used by weather experts to modify, or add new rules to handle specific patterns. When the system's rule base expands, it may be useful to implement rule base pruning mechanisms to remove/prune redundant or less relevant rules that contribute less to the system's performance [24].

III. The Proposed Life-Long Learning XAI Metaheuristic-based Type-2 Fuzzy System for Solar Radiation Modelling

The proposed method encompasses the following five main stages:

1- Initial rule identification using Wang-Mendel (WM) method.
2- Rule selection using Simulated Annealing (SA).
3- Linguistic variable selection in each rule using SA (rule length reduction).
4- Domain expert contribution, including evaluation, amendment, and addition of new rules.
5- Life-long learning and model transfer to adapt the model to suit new conditions and locations.

The following subsections detail these steps, while Fig 1 and Fig 3 present flowcharts of the five stages and their main steps.

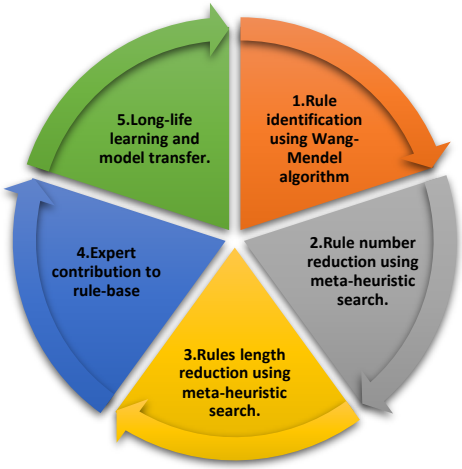


Fig. 1. The five stages of the method as part of lifelong learning.

A. Data Preparation

The used data in this study is a combined data set for nine locations in Saudi Arabia with different topographical characteristics. The data is obtained from two available sources: solar ground stations and a Numerical Weather Prediction (NWP) model. This research made use of ground station data from nine Saudi Arabian sites' stations. These stations are owned and maintained by the national renewable energy authority [28]. The use of the NWP models are needed to cover missing variables: dust, cloud cover, and modeled clear-sky Global Solar Irradiation (GHI). The first two variables are known to affect the level of radiation but are not measured by the available ground stations, so they are anticipated by some variables in the NWP model. The NWP model used is the fifth generation of atmospheric reanalysis of global climate (ERA5) published by the European Centre for Medium-Range Weather Forecasts (ECMWF) [29]. The ground stations documented the existence of uncertainties in the observed records with regard to the data's reliability, despite the equipment's routine maintenance and calibration. These uncertainties vary from ±2% to ±9% for GHI data, which makes the use of type-2 fuzzy logic systems desirable to handle the faced uncertainties.

The target is to predict the next hour's total solar radiation, known as global solar irradiation (GHI), which is an observed value from ground stations.

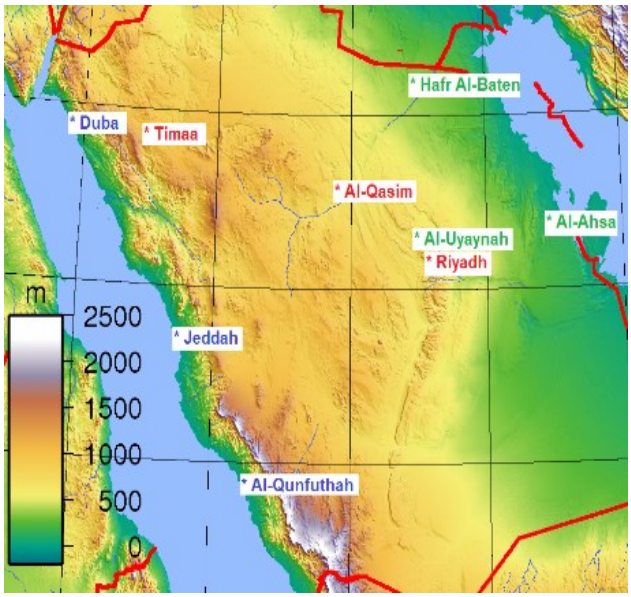


Fig. 2: Saudi Arabia map with case studies' stations (blue font =case study 1 (Coastal), red= case study 2 (Desert and High Altitude), green=case study 3 (Plain Desert). Source: amended from [35].

The inputs are ten weather variables of the current hour related to the GHI, which are:

- the temperature of the air, measured in degrees Celsius.
- day of the year as a number (from 1 to 366 (leap year))
- hour number in a 24-hour period.
- average wind direction from the north.
- average wind speed, measured in meters per second.
- relative humidity percentage.
- station pressure, measured in millibars.
- NWP modeled friction velocity refers to the velocity of air moving over a surface that generates a force and

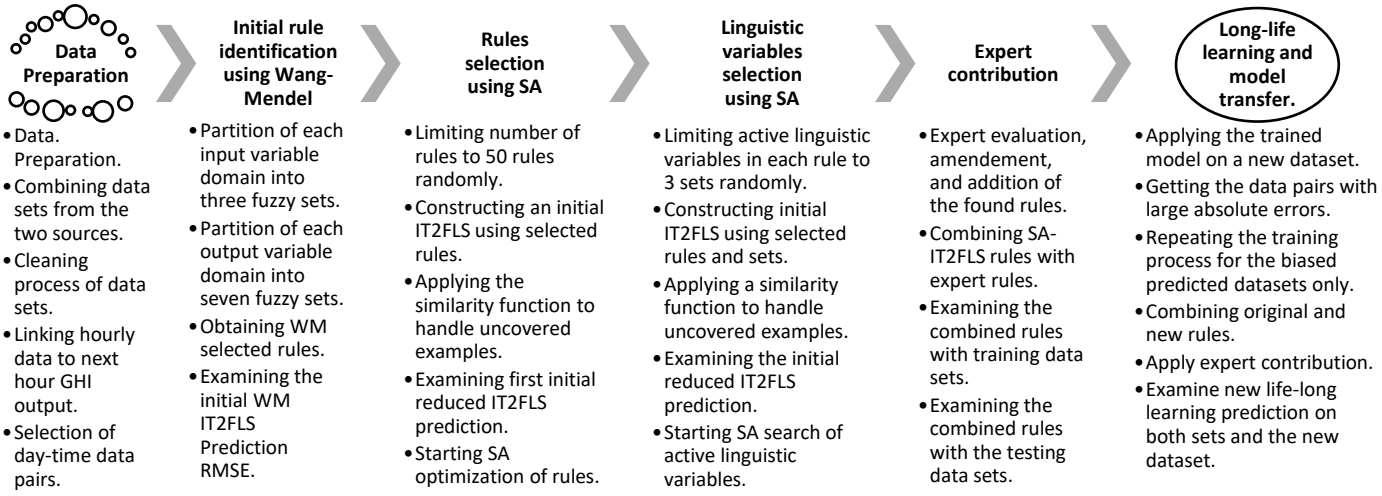


Fig. 3. The main steps in the five stages of the method.

transfers momentum to the surface, therefore slowing the wind. It is used to serve as a leading indicator for dust.

- NWP modeled total cloud cover, which ranges from 0 to 1, is the percentage of a grid box that is covered in clouds.
- mathematically modeled clear-sky GHI for the target hour, assuming clear-sky (cloudless) conditions. The modeled values are calculated using the mathematical function proposed in [30].

The first seven input variables are obtained from the ground stations, while the remaining three variables are estimated from the NWP model and a mathematical function. The output is the GHI for the next hour in Watt per square meter Wh/m^2. The predicted period is only during the extended daytime, from 6:00 a.m. to 19:00 p.m., based on the previous hour's data.

The combined dataset is then cleaned and verified by removing data pairs that have missing values as well as converting date and time values to numerical values. The availability of ground data for each station starts at a different date based on the initial start time of recording. Therefore, the training data pairs start from the first available day in 2013. Table I displays the ground station specifications as well as the training and testing periods and sample percentages. The nine data sets are divided into three **case studies** (of different topographical and weather patterns) to evaluate the proposed method for lifelong learning, as shown in Table I and Fig. 2. The **first case study** relates to coastal areas, while the **second case study** relates to Desert areas with high altitude and the third case study relates to plain desert. The **first dataset** in each case study is fully used for training the model, while the **second dataset** is partially used to improve the trained model. **The third dataset** in each case study is employed for testing (unseen data set) the final model.

B. First Stage: Initial rule identification using Wang-Mendel

The initial rule base identification is obtained using the WM algorithm. Firstly, each input variable $(x_1, x_2, \dots x_n)$ domain is partitioned into three gaussian type-1 fuzzy sets denoted by linguistic labels $(Low, Midium, High)$. The domain for each variable is bound by the minimum and maximum values of each variable in the training pairs. Since the gaussian fuzzy set has nonzero membership, there will be one vertex in the middle of the region with a membership value of 1, and the other two vertices will have membership values approaching 0 at the middle of the two adjacent regions. The standard deviation for each gaussian fuzzy set is chosen heuristically to achieve this condition. Each output variable domain y is partitioned into seven type-1 gaussian fuzzy sets denoted by linguistic labels $(VeryLow, Low, LowerMidium, Midium, HigherMidium, High, VeryHigh)$.

After the input/output domain partitioning, we employ the second step to generate a number of type-1 fuzzy rules from the numerical data pairs following the WM approach described in [31]. The membership function of each fuzzy set is converted as an interval type-2 gaussian fuzzy set with uncertain means and a unified standard deviation determined by the following formula:

$$\tilde{f}(x) = exp^{-\left\{\frac{x-m}{2\sigma}\right\}^2}, m \in [m_1, m_2] \quad (2)$$

TABLE I
CASE STUDIES AND SITES DATA SETS DETAILS.

Case Study	Purpose	Site	Latitude	Longitude	Elevation (meter)	Training dates / samples	Testing dates / (samples)
1	Training (Stages 1-4)	Jeddah	21.49	39.24	75	29/5/2013-28/5/2014 (5109)	29/5/2014-31/12/2014 (3038)
	Partial training (Stage 5)	Al-Qunfuthah	19.15	41.08	20	05/09/2013-04/09/2014 (5025) (partially)	05/09/2014-31/12/2014 (1652)
	Independent Testing	Duba	27.34	35.72	45	-	26/07/2013-31/12/2014 (7335)
2	Training (Stages 1-4)	Timaa	27.61	38.52	844	24/7/2013-23/7/2014 (5109)	23/7/2014-31/12/2014 (2254)
	Partial training (Stage 5)	Al-Qassim	26.35	43.77	688	03/06/2013-02/06/2014 (4526) (partially)	03/06/2014-31/12/2014 (2944)
	Independent Testing	Riyadh	24.71	46.68	668	-	16/01/2014-31/12/2014 (9897)
3	Training (Stages 1-4)	Al-Ahsa	25.34	49.59	170	29/5/2013-28/5/2014 (4965)	29/5/2014-31/12/2014 (3038)
	Partial training (Stage 5)	Al-Uyanah	24.9	46.4	779	14/01/2013- 13/01/2014 (5109) (partially)	14/01/2014-31/12/2014 (4928)
	Independent Testing	Hafr Al-baten	28.33	45.96	383	-	06/10/2014-31/12/2014 (1204)

Consequently, the upper $\bar{\mu}_{\tilde{A}(x)}$ and lower membership functions $\underline{\mu}_{\tilde{A}(x)}$ are defined by the following two formulas [32]:

$$\bar{\mu}_{\tilde{A}(x)} = \begin{cases} exp^{-\left\{\frac{x-m_1}{2\sigma}\right\}^2} & if\ x < m_1, \\ 1 & if\ m_1 \le x \le m_2, \\ exp^{-\left\{\frac{x-m_2}{2\sigma}\right\}^2} & if\ x > m_2 \end{cases} \tag{3}$$

$$\underline{\mu}_{\tilde{A}(x)} = \begin{cases} exp^{-\left\{\frac{x-m_2}{2\sigma}\right\}^2} & if\ x \le \left(\frac{m_1+m_2}{2}\right), \\ exp^{-\left\{\frac{x-m_1}{2\sigma}\right\}^2} & if\ x > \left(\frac{m_1+m_2}{2}\right) \end{cases} \tag{4}$$

where , $m \in [m_1, m_2]$ are the two means of the type-2 fuzzy set, and σ is the standard deviation. The area between the upper $\bar{\mu}_{\tilde{A}(x)}$ and lower membership functions $\underline{\mu}_{\tilde{A}(x)}$ is known as the "footprint of uncertainty (FOU)," which allows many types of uncertainty to be modeled within.

The fuzzy logic system processes begin with the fuzzification of the input values into their interval type-2 fuzzy sets using the membership function defined above. Then, the minimum t-norm function is used to combine all the membership grades for all the inputs to get firing level values for each rule. As mentioned above, the rules at this stage are the ones obtained from the WM algorithm. The defuzzification process is carried out using the Nie-Tan defuzzification algorithm, which computes the defuzzified output value $y(x')$ directly using the fired upper and lower final antecedent values $\overline{f}^n(x'), and\ \underline{f}^n(x')$ based on the following formula [33]:

$$y(x') = \frac{\sum_{n=1}^{N} y^n \left[\underline{f}^n(x') + \overline{f}^n(x')\right]}{\sum_{n=1}^{N} \left[\underline{f}^n(x') + \overline{f}^n(x')\right]} \tag{5}$$

The IT2FLS incorporates the similarity function proposed by [34]. Similarity functions can be beneficial to handle uncovered data pair examples created as a result of when a subset of rules is chosen from the full rule base. In this paper, we will employ the similarity metrics employed in [34] for determining the rules in the rule base that are most relevant to classify newly uncovered examples. The proposed method can be used to solve regression-based problems by using the proposed function to find the rules that are most similar to the missed rules. The approach first generates a set of fuzzy rules for the revealed example by considering positive membership degrees for each variable. Then, all the possible fuzzy rules are extracted by running all the possible combinations of the previously matched membership functions in the antecedent part only. Then, the following similarity function formula is used to determine the similarity of fuzzy rules in the rule-base to the set of fuzzy rules for the uncovered example $\mathrm{SFRU}(x_p)$:

$$Similarity\left(R'_j, R_j\right) = \prod_{k=1}^{n} \left| 1 - \frac{abs\left(V\left(R'_{jK}\right) - V\left(R_{jK}\right)\right)}{NL_K} \right| \tag{6}$$

where $j' = \{1, \ldots . \left|\mathrm{SFRU}(x_p)\right|\}$, and $\left|\mathrm{SFRU}(x_p)\right|$ is the number of rules in the set of fuzzy rules for the uncovered example $\mathrm{SFRU}(x_p)$. R'_j is the jth rule in $\mathrm{SFRU}(x_p)$. R_j is the jth rule in the rule-base where $j = \{1, \ldots, L\}$. n is the number of variables, while NL_K is the number of fuzzy sets (linguistic labels) for the *kth* variable. The function V refers to the integer-coded location of the *kth* fuzzy set of the examined rule. This function will assign a similarity degree to each rule in the rule base for every rule in $\mathrm{SFRU}(x_p)$. The rule in the rule-base with the highest degree of similarity is then assigned to each rule in $\mathrm{SFRU}(x_p)$. In this work, the presented similarity function is modified to suit regression-based problems and to handle the reduction of the full combination rules to WM selected rules. The first amendment is that the chosen gaussian membership functions do not return zero membership grades for any value within the data range. Hence, a threshold is applied to determine when to use the similarity function. The threshold is set to 0.1 in this study, meaning that any data pair with maximum firing level values of the upper memberships less than or equal to the threshold will be passed to the similarity function. Human experts can also adjust the level of this threshold to achieve the desired level of data pair coverage for the system. However, threshold values should be chosen with care to avoid redundancy in automatically generated rule bases. In the second amendment, the similarity function selects the most similar rule among WM rule-base as follows:

1. The data pair is considered one of the uncovered examples in the set of fuzzy rules for the uncovered example $\mathrm{SFRU}(x_p)$ if the maximum firing level of upper memberships is less than or equal to the similarity threshold of 0.1. In this situation, the similarity function should be applied.
2. For each uncovered data pair, find the most similar (missing) rule (R') among the original WM rule base (R). Therefore, R_j is the jth rule in the WM rule-base where $j = \{1, \ldots, L_{WM}\}\ in\ Eq\ (6)$.
3. Using the similarity function formula, determine the degree of similarity between R' and the (50) subset of selected rules (50 is the desired number of rules to be present in the rule base to maximize the system interpretability, the best rules will be found via simulated annealing as discussed in the next subsection). The calculation for similarity is based on the mean values of the upper and lower memberships in this experiment.
4. The rule with the highest degree of similarity among the 50 selected rules is chosen for each uncovered data pair.
5. The uncovered data pair is assumed to have full firing level values $(1/1)$ for both upper and lower firing levels in the most similar rule. As a result, the consequent parts of the most similar rule are calculated with a full firing of upper and lower bounds to their consequent output fuzzy sets.
6. The type-reduction of each uncovered data pair is carried out using $50 + 1 = 51$ rules. Therefore, the output values for these data pairs are calculated using the new set of combined rules. The root mean squared error (RMSE) was selected as the prediction error metric during the search as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y})^2} \tag{11}$$

where $N$is the number of samples, y_i is the system predicted value, and $\hat{y}$ is the observed (target) value.

C. Rule selection using SA.

The second stage involves the optimization of the discovered

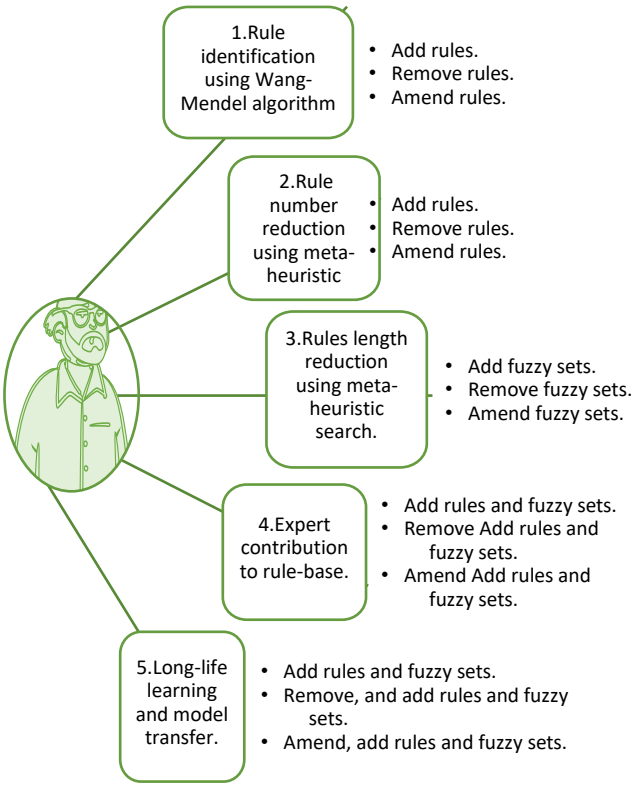


Fig. 4. Domain experts' potential contribution in each stage.

WM rules by searching for a small subset of rules within the WM rule base of fixed length (to maximize the model interpretability) while maximizing the model accuracy. First, an adequate number of rules from the WM rule base are selected at random. After that, the best set of rules with the smallest prediction error are found iteratively through a combinatorial search of SA. In our experiments, the fixed number of rules (suggested by the domain experts) is set at 50, though any other limited number that enables interaction with experts can be chosen. The neighborhood representation of the search processes in this stage is formulated as combinatorial search optimization. The SA searches for the best set of rules by swapping one of the current ones with another one from the non-selected WM rule base. The initial temperature of the SA is chosen heuristically to be 2, which allows enough opportunities to avoid being stuck in local minima. To allow an exploration of the search space in the first iterations, the search swaps three values at a time during the first three Markov chains and one thereafter. To control the annealing process, Markov chains are used to reduce the temperature once after 120 iterations for each chain. The maximum number of Markov chains is 50, after which the search terminates. The cooling process of the temperature parameter is based on the static cooling function of the current Markov chain and the initial temperature as follows:

$$T_{mc} = T_0 * C \tag{7}$$

Where T_{mc} is the temperature value at the current Markov chain, T_0 is the initial temperature, and C is a cooling constant set to 0.85. Therefore, the IT2FLS is updated with the new rule set, and the output of the training data set is evaluated at each iteration. The SA acceptance mechanism is used, which always accepts better states while only accepting worse states randomly with a controlled probability. As described in the previous section, the similarity function is utilized in the IT2FLS calculation process. Therefore, as the search continues, better states should be discovered. It is important to note that when the search begins with a random set from the WM rule base, the objective function may initially appear to worsen. However, an adequate search should quickly overcome this issue and locate states that are better compared to WM's complete sets of rules. The search terminates after 6,000 iterations encapsulated in 50 Markov chains.

D. Third Stage: Rule length reduction using SA.

The third stage involves optimizing the 50 best-discovered rules by searching for a small subset of linguistic labels (fuzzy sets) in each rule. Initially, a few but sufficient linguistic labels are chosen at random for each rule. The optimal set of linguistic labels for each rule is then determined iteratively via combinatorial search utilizing SA. In this experiment, the number of rule antecedents is fixed at three but any other small number that enables an adequate interaction with experts is acceptable. In this stage, the neighborhood representation of search processes is expressed as combinatorial search optimization. The SA searches for the optimal set of linguistic labels by exchanging one of the current linguistic labels with another from the same rule. The SA and IT2FLS configurations are identical to those described in the previous stage. At each iteration, the IT2FLS is updated with a new set of linguistic labels for the 50 rules, and the output of the training data set is measured. Within the IT2FLS procedure, the similarity function is still used. When the search begins with a random set from the previous stage's rule base, the objective function may initially appear to deteriorate. Nevertheless, a sufficient search should quickly overcome this issue and identify states that are preferable to the sets of rules from the previous stage in terms of the objective function.

E. Fourth Stage: Domain expert contribution

The goal of this stage is to incorporate domain experts' knowledge in various ways. First, experts can examine the best-found rules identified by the first three stages and assess their consistency with their knowledge and scientific observations. Then experts can remove and modify existing rules, as well as add new ones. Experts can analyze uncovered examples separately and provide additional rules as needed. The expert should be involved from the beginning, recommending the input variables and their level of granularity. Combining a data-driven knowledge base with expert knowledge can increase the level of trust in a continuously evolving system. Following an expert contribution to the current set of rules, the dataset prediction is examined using the new combined set of rules. APPENDIX B presents a sample of generated data-driven and expert rules, while Fig. 4 describes the potential contribution of domain experts in each stage.

F. Fifth Stage: Long-life learning and model transfer

In this stage, the previously trained model is transferred to a new dataset for another location with partial training. Transferring the model to a location with similar topographical characteristics and weather patterns is recommended for better results, especially when these characteristics are not included in

the first data set. The following are the proposed steps to complete this stage:

- **S5-A**: Evaluate the new dataset using a pretrained fuzzy logic system that has previously been trained by the previous four stages for another similar or close site.

- **S5-B**: Perform Stages 1, 2, and 3 to find the best limited set of rules to handle a selected subset of bad predicted data pairs and therefore add them to the rule-base. The data pairs chosen should have the greatest prediction errors in both the first and second data sets. This can be accomplished by filtering prediction errors using a predetermined absolute error threshold. The chosen threshold is problem-dependent and related to the allowable level of error by experts. The number of new rules chosen in this experiment should be small compared to the first rule-base to preserve the performance and consistency of the original rule base where possible. The added number of rules is limited in our system to only 12 rules. The training process is expected to be quick at this stage because the number of rules and data sets is limited. Domain experts may modify the newly discovered rules or add new site-specific rules either before or after this step. We then combine the new subset-related rules with the pretrained set of rules to form a combined rule-base that should represent an adequate modeling of the two datasets (for the first two sites). Then, using the combined set of rules, evaluate all data pairs for the second site as well as the first site.
- **S5-C**: Apply Stage 4, in which domain specialists may analyze the prediction of the new added or combined rule-base and contribute as necessary. Experts may revise previous rules or add general rules applicable to both sites. This step may marginally increase the cumulative prediction error when a site-dependent rule or a rule to handle the minority of cases is added (rare scenarios).
- **S5-D**: Apply a pruning process to remove rules that may degrade prediction after subsequent rule-combining processes. The process starts iteratively by removing a rule each time from the last version of the rule-base and reevaluating the combined training sets (the first set and the second minimal set). Then, a rule is eliminated if its removal improves the prediction. The simple pruning process is repeated until there is no further improvement. Using the new combined set of rules, testing data pairs from the first and second sites as well as any new data sets can be evaluated.

At this point, an IT2FLS should reach a general state acceptable for predicting new similar data examples. The fifth stage can be repeated whenever a new batch of different data sets arrives. The model can be moved to new locations by repeating the Stage 5 steps. Table II presents the rules and the relevant reductions as detailed in the next section.

IV. Experiments and Results

The system was evaluated with real-world data sets from nine locations in Saudi Arabia with distinct climatic and topographical characteristics. The experiments are divided into three case studies, each with 3 sites' datasets, as shown in Table I. In each case study, the first site data set is used to train the proposed system in the first four stages, followed by the training of a small subset of the second site dataset to build lifelong learning capabilities, and finally testing on a totally independent third site dataset. To show the impact of each stage, an evaluation of both training and testing data sets is carried out after each stage.

TABLE II
THE NUMBER OF RULES, RULE REDUCTIONS, AND LINGUISTIC VARIABLES DURING THE FIVE STAGES (LV=LINGUISTIC VARIABLE).

Case Study 1								
Stage	**S1**	**S2**	**S3**	**S4**	**S5A**	**S5B**	**S5C**	**S5D**
Selected Rules	579	50	50	52	52	64	65	63
WM Rules + Expert Rules	597	597	597	581	581	891	892	892
All Combinations Rules	59049	59049	59049	59049	118098	118098	118098	118098
Selected LVs in Rules	5790	500	150	156	156	192	195	189
All Combinations LVs	590490	590490	590490	590490	1180980	1180980	1180980	1180980
Rule Reduction %	99.02%	99.92%	99.92%	99.91%	99.96%	99.95%	99.94%	99.95%
Case Study 2								
Stage	**S1**	**S2**	**S3**	**S4**	**S5A**	**S5B**	**S5C**	**S5D**
Selected Rules	753	50	50	52	52	64	66	65
WM Rules + Expert Rules	753	753	753	755	755	1304	1306	1306
All Combinations Rules	59049	59049	59049	59049	118098	118098	118098	118098
Selected LVs in Rules	7530	500	150	156	156	192	198	195
All Combinations LVs	590490	590490	590490	590490	1180980	1180980	1180980	1180980
Rule Reduction %	98.72%	99.92%	99.92%	99.91%	99.96%	99.95%	99.94%	99.94%
Case Study 3								
Stage	**S1**	**S2**	**S3**	**S4**	**S5A**	**S5B**	**S5C**	**S5D**
Selected Rules	867	50	50	53	53	65	67	66
WM Rules + Expert Rules	867	867	867	870	870	1569	1571	1571
All Combinations Rules	59049	59049	59049	59049	118098	118098	118098	118098
Selected LVs in Rules	8670	500	150	159	159	195	201	198
All Combinations LVs	590490	590490	590490	590490	1180980	1180980	1180980	1180980
Rule Reduction %	98.53%	99.92%	99.92%	99.91%	99.96%	99.94%	99.94%	99.94%

TABLE III
CASE STUDY 1 PREDICTION ERROR RESULTS FOR FIRST SITE (STAGES S1-S4).

Stage		S1		S2		S3		S4	
Metric	Metric Stats	Wang-Mendel		Rule-base selection		Linguistic Variables selection		Expert contribution	
	Site - Sample	1st - Train	1st - Test	1st - Train	1st - Test	1st - Train	1st - Test	1st - Train	1st - Test
RMSE	Mean	178.8	182.10	77.94	77.15	70.55	63.33	71.10	62.80
	Min	178.8	182.1	76.92	76.04	69.17	60.98	69.87	60.90
	Max	178.8	182.1	78.83	78.51	71.82	65.10	72.42	64.55
rRMSE	Mean	38.27	37.91	17.58	16.82	15.95	13.88	16.08	13.76
	Min	38.27	37.91	17.42	16.46	15.64	13.36	15.80	13.35
	Max	38.27	37.91	17.75	17.14	16.21	14.28	16.35	14.16
MBE	Mean	19.32	9.46	-0.25	-12.12	1.32	-7.72	6.22	-2.78
	Min	19.32	9.46	-2.96	-14.93	0.64	-12.06	5.32	-7.11
	Max	19.32	9.46	3.58	-6.27	3.08	-3.91	7.65	0.44

TABLE V
"CASE STUDY 2 PREDICTION ERROR RESULTS FOR FIRST SITE (STAGES S1-S4).

Stage		S1		S2		S3		S4	
Metric	-	1st - Train	1st - Test	1st - Train	1st - Test	1st - Train	1st - Test	1st - Train	1st - Test
RMSE	Mean	206.7	196.7	83.56	71.26	73.43	61.77	73.42	61.61
	Min	206.7	196.7	81.83	68.86	69.48	56.56	69.47	56.45
	Max	206.7	196.7	84.73	74.87	75.13	64.59	75.13	64.39
rRMSE	Mean	40.49	38.80	17.44	15.46	15.24	13.26	15.24	13.23
	Min	40.49	38.80	17.12	15.04	14.45	12.13	14.45	12.12
	Max	40.49	38.80	17.72	16.33	15.56	13.82	15.57	13.79
MBE	Mean	15.59	37.37	-0.37	7.10	-0.66	8.81	-1.76	8.03
	Min	15.59	37.37	-4.58	4.13	-2.70	4.96	-3.87	4.23
	Max	15.59	37.37	6.85	16.10	1.51	11.08	0.49	10.28

TABLE IV
CASE STUDY 1 PREDICTION ERROR RESULTS FOR FIFTH STAGE (ALL SITES).

Stage		5A	5B		5C				5D		
Metric	Metric Stats	First Pre-trained Model	Combined Rule-bases		Expert Contribution				Final Pruned Combined Model		
	Site - Sample	2nd - Train	2nd - Train	2nd - Test	2nd - Train	2nd - Test	1st - Train	1st - Test	1st - Test	2nd - Test	3rd - Test
RMSE	Mean	79.89	80.98	80.95	80.00	75.83	78.75	68.01	65.53	75.38	78.88
	Min	75.68	77.59	76.41	76.55	72.41	76.08	65.24	63.41	70.48	72.37
	Max	88.81	86.75	85.65	85.45	79.74	82.09	72.95	66.80	78.95	83.38
rRMSE	Mean	18.31	18.40	19.55	18.08	18.03	17.53	14.59	14.26	18.29	17.08
	Min	17.34	17.67	18.52	17.35	17.29	16.97	14.06	13.79	17.04	15.66
	Max	20.39	19.70	20.41	19.32	18.63	18.12	15.35	14.49	19.20	18.17
MBE	Mean	-14.73	-8.89	-19.4	6.87	-0.47	20.98	10.39	-0.30	-11.7	-22.1
	Min	-21.73	-11.24	-21.1	3.98	-4.04	14.89	0.16	-11.7	-21.3	-34.2
	Max	-9.28	-6.71	-17.68	9.46	2.43	26.63	17.55	9.22	-0.80	-14.2

Different metrics recommended for the solar radiation prediction problem are used to analyze the results to provide a clearer understanding of the results' accuracy. In addition to Root Mean Square Error (RMSE), relative Root Mean Squared Error (rRMSE) and Mean Bias Error (MBE) are also employed as metrics. Although RMSE is advantageous for evaluating the system accuracy, MBE is used to highlight the average model's bias and thus can reveal the under-estimation or over-estimation characteristics of the prediction models. The rRMSE illustrates the prediction errors relative to the values of GHI. In addition to RMSE, which was defined above, the following metrics have corresponding mathematical formulas:

$$rRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y})^2}}{\frac{1}{N}\sum_{i=1}^{N}\hat{y}} \cdot 100 \quad (8)$$

$$\mathrm{MBE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}) \quad (9)$$

where N is the number of samples, y_i is the system predicted value, and $\hat{y}$ is the observed (target) value. The results are detailed in Tables III, IV, V, VI, VII, and VIII, where each case's results are separated into two tables, with the first table displaying the impact of each stage from Stages 1–4 and the second table displaying the impact of the fifth stage in each site's dataset. To show the structure of the proposed system, Table II shows the reduction levels of the rule-base during the five stages and Appendix A shows a sample of the final rules. Fig. 5 depicts an example of a contour map illustrating the coverage of the final rule base of the model's ten inputs. An example of generated input and output fuzzy sets is drawn in APPENDIX B.

The results analysis aims to answer the following questions:

i. Has the training in Stages 1-4 resulted in an explainable model with an acceptable level of prediction accuracy for the first site? And to what extent has each stage contributed to the system's performance?
ii. Did the Stage 5 training improve the prediction of a different second site over the pretrained first site model? And to what extent has each step in Stage 5 contributed to the resulting system's performance?
iii. Considering the explainability and design restrictions of the final model, is the final model capable of predicting the third site dataset, which is an entirely new dataset, with an acceptable level of accuracy relative to benchmark models?

The first question*: Has the training in Stages 1-4 resulted in an explainable model with an acceptable level of prediction accuracy for the first site? And to what extent has each stage contributed to the system's performance?*

a. First Stage: Initial rule identification using Wang-Mendel

The first stage of generating an initial set of rules for the first site using WM resulted in a relatively large set of rules compared to the subsequent stages (Table II), where (579, 753, and 867) rules were generated for the three case studies respectively. However, these figures represent a very high reduction in the rule base from 59049 potential rules (minimum 98.53% reduction). The generated rules introduced relatively high prediction errors (Tables III, V, VII) compared to subsequent stages, with an average RMSE of (178.8, 206.7, and 194.7) for the training sets and similar results for the testing results.

TABLE VI
CASE 2 PREDICTION ERROR RESULTS FOR FIFTH STAGE (ALL SITES).

Stage		5A	5B		5C				5D		
Metric	-	2nd - Train	2nd - Train	2nd - Test	2nd - Train	2nd - Test	1st - Train	1st - Test	1st - Test	2nd - Test	3rd - Test
RMSE	Mean	95.22	90.81	73.29	90.62	73.06	77.83	63.90	63.57	70.65	92.68
	Min	90.01	87.23	69.64	87.02	69.39	73.95	59.71	60.21	66.34	87.25
	Max	97.79	93.15	76.60	93.03	76.51	80.35	66.44	65.82	75.00	95.66
rRMSE	Mean	20.64	19.74	15.04	19.70	15.00	16.15	13.72	13.68	14.49	19.86
	Min	19.59	18.99	14.26	18.95	14.21	15.29	12.63	12.84	13.60	18.76
	Max	21.18	20.20	15.71	20.18	15.69	16.74	14.39	14.33	15.42	20.45
MBE	Mean	15.40	6.71	-2.77	5.67	-3.50	-11.9	-2.26	6.66	5.24	19.90
	Min	12.38	1.89	-8.49	0.83	-9.22	-16	-6.33	-6.91	0.89	14.67
	Max	18.56	13.50	5.92	12.44	5.18	-8.06	2.46	11.19	13.67	26.90

TABLE VII
CASE STUDY 3 PREDICTION ERROR RESULTS FOR FIRST SITE (STAGES S1-S4).

Stage		S1		S2		S3		S4	
Metric	-	1st - Train	1st - Test	1st - Train	1st - Test	1st - Train	1st - Test	1st - Train	1st - Test
RMSE	Mean	194.7	199.95	84.65	76.49	76.15	67.85	77.18	67.13
	Min	194.7	199.95	82.31	73.73	74.60	65.20	75.52	64.66
	Max	194.7	199.95	87.66	81.03	77.50	70.75	78.63	69.91
rRMSE	Mean	40.45	39.80	18.59	16.02	16.58	14.02	16.79	13.84
	Min	40.45	39.80	18.12	15.55	16.21	13.48	16.39	13.34
	Max	40.45	39.80	19.19	16.86	16.89	14.64	17.14	14.44
MBE	Mean	22.72	8.90	-3.84	-13.34	2.52	-7.51	3.83	-6.38
	Min	22.72	8.90	-6.19	-19.10	1.35	-10.04	2.73	-8.88
	Max	22.72	8.90	-0.63	-9.69	4.20	-4.20	5.43	-3.17

TABLE VIII
CASE 3 PREDICTION ERROR RESULTS FOR FIFTH STAGE (ALL SITES).

Stage		5A	5B		5C				5D		
Metric	-	2nd - Train	2nd - Train	2nd - Test	2nd - Train	2nd - Test	1st - Train	1st - Test	1st - Test	2nd - Test	3rd - Test
RMSE	Mean	93.05	92.46	93.62	93.49	91.95	82.22	72.50	70.23	91.27	96.80
	Min	90.28	86.18	86.64	88.72	86.01	78.78	68.01	66.73	85.63	89.18
	Max	98.03	97.99	98.75	98.24	96.16	86.91	78.51	74.86	97.16	103.7
rRMSE	Mean	20.30	20.01	19.40	20.15	18.94	17.61	14.61	14.29	18.90	22.21
	Min	19.61	18.71	17.99	19.18	17.76	16.95	13.77	13.70	17.78	20.60
	Max	21.47	21.13	20.39	21.08	19.73	18.47	15.68	15.04	20.05	23.56
MBE	Mean	-5.37	0.99	-12.4	16.22	2.16	19.71	12.05	-0.73	-8.78	1.72
	Min	-21.3	-5.61	-18.2	9.68	-3.67	13.66	3.51	-9.11	-14.1	-2.09
	Max	7.15	6.62	-7.06	20.14	5.87	25.84	20.40	13.98	-6.03	9.00

TABLE IX
PERCENTAGE OF IMPROVEMENT IN PREDICTION ERRORS FOR THE *FIRST* SITE TESTING DATA SET OVER WM RULES.

Site		Case Study 1 - Site 1				Overall
Stage		S2	S3	S4	S5	S2-S5
RMSE	Mean	57.63	17.92	0.84	-4.36	63.34
	Min	58.24	19.80	0.14	-4.12	64.53
	Max	56.89	17.09	0.85	-3.49	62.64
rRMSE	Mean	55.63	17.48	0.86	-3.65	62.73
	Min	56.58	18.80	0.14	-3.32	63.97
	Max	54.79	16.66	0.88	-2.34	62.14
Site		Case Study 2 - Site 1				Overall
Stage		S2	S3	S4	S5	S2-S5
RMSE	Mean	63.78	13.33	0.26	-3.19	69.25
	Min	65.00	17.86	0.19	-6.65	70.88
	Max	61.95	13.73	0.30	-2.22	68.16
rRMSE	Mean	60.17	14.22	0.23	-3.41	66.22
	Min	61.24	19.32	0.15	-5.97	68.29
	Max	57.91	15.36	0.27	-3.93	64.62
Site		Case Study 3 - Site 1				Overall
Stage		S2	S3	S4	S5	S2-S5
RMSE	Mean	61.74	11.30	1.06	-4.62	63.93
	Min	63.13	11.57	0.83	-3.21	65.73
	Max	59.48	12.68	1.19	-7.08	61.56
rRMSE	Mean	59.76	12.47	1.28	-3.23	64.68
	Min	60.93	13.32	1.04	-2.72	66.13
	Max	57.64	13.16	1.40	-4.21	62.80
Average rRMSE Improvement	**Mean**	**58.52**	**14.72**	**0.79**	**-3.43**	**64.54**

TABLE X
PERCENTAGE OF IMPROVEMENT IN PREDICTION ERRORS FOR THE *SECOND* SITE TESTING DATA SETS AFTER PERFORMING STAGE 5'S STEPS.

Site		Case Study 1 - Site 2			Overall S5B-S5D
Stage		S5B	S5C	S5D	
RMSE	Mean	-1.33	6.33	0.60	5.65
	Min	-0.96	5.24	2.67	6.87
	Max	3.56	6.89	1.00	11.10
rRMSE	Mean	-6.76	7.78	-1.43	0.14
	Min	-6.82	6.65	1.43	1.71
	Max	-0.08	8.71	-3.06	5.84
Site		Case Study 2 - Site 2			Overall S5B-S5D
Stage		S5B	S5C	S5D	
RMSE	Mean	23.03	0.32	3.30	25.81
	Min	22.63	0.36	4.39	26.29
	Max	21.67	0.11	1.98	23.31
rRMSE	Mean	27.14	0.29	3.37	29.80
	Min	27.22	0.32	4.29	30.57
	Max	25.83	0.07	1.76	27.19
Site		Case Study 3 - Site 2			Overall S5B-S5D
Stage		S5B	S5C	S5D	
RMSE	Mean	-0.61	1.78	0.74	1.91
	Min	4.03	0.73	0.44	5.16
	Max	-0.73	2.62	-1.04	0.89
rRMSE	Mean	4.43	2.35	0.24	6.90
	Min	8.26	1.30	-0.10	9.36
	Max	5.02	3.25	-1.62	6.62
Average rRMSE Improvement	**Mean**	**8.27**	**3.47**	**0.73**	**12.28**

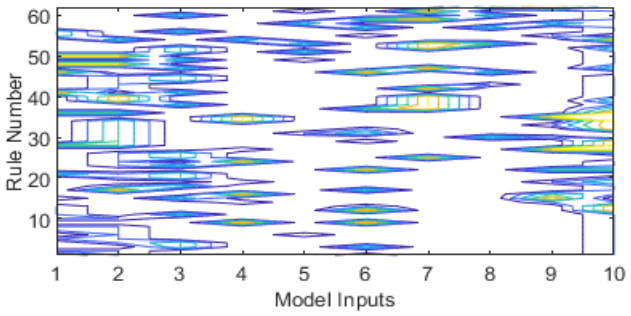


Fig. 5. A contour map of the final rule base coverage of the model's ten inputs for Al-Ahsa site.

To answer this question, the first four stages' contributions are analyzed separately as follows:

b. Second Stage: Rule selection using SA.

The second stage of rule reduction using SA resulted in rule bases of only 50 rules (Table II), with 99.92% reduction (comparing the 50 rules with the All-Combinations Rules). The average improvement in the relative RMSE (rRMSE) prediction errors compared to the WM rule bases were high at 58.52% (Table IX).

However, up to this stage, it has been very difficult for an expert to understand these rules since each rule is composed of 10 linguistic variables, which is the purpose of the next stage.

c. Third Stage: Rule length reduction using SA.

The third stage of rule length reduction (linguistic variables selection) utilizing SA resulted in rule bases consisting of 50 rules with only three linguistic variables per rule. The average improvement in relative RMSE prediction errors (rRMSE) was 14.72 percent beyond the second stage (Table IX). At this stage, the rule base is intended to be comprehended by an expert, as it consists of a small number of rules with maximum three antecedents per rule. In addition, the contour map in Fig. 5 further illustrates the effects of the variable selection process by showing that all ten model inputs were covered with varying grades, indicating the validity of the variable selection process.

d. Fourth Stage: Domain expert contribution.

In the fourth stage, the expert contribution includes the addition of new rules (the expert might choose to have more antecedents per rules) after analyzing the prediction behavior of the automated rules. Because of their high levels of explainability, the models were easily understood, analyzed, and augmented by the relevant meteorology experts. For instance, the expert discovered in Case Study 3 that the automated rules did not account for some fog and dust circumstances, resulting in an underestimation of their effects. Consequently, the expert proposed adding new rules to cover these circumstances (samples of experts' rules are shown in Appendix B). An example of an expert added rule to handle a dust storm scenario is:

IF (Day Number is 1st Quadrimester of Year) and (Avg Wind Speed is High) and (Relative Humidity is Low) and (Avg Wind Direct is High) THEN (Next Hour GHI is Low)

Besides this, he described the generated rules from the models as "acceptable" and "simple" in general. The first case study sites are coastal sites, where he suggested adding rules that represent some extreme high and low solar conditions. Intriguingly, a small number of expert rules (2 or 3) resulted in a marginal improvement of 0.79 rRMSE on average beyond the third stage for the three testing datasets (Table IX). The significance of the addition of this type of rule can also be viewed through the lens of minority case modeling. Due to the rareness of these situations, the automated optimization process tends to disregard them in favor of rules that represent the majority of cases. This behavior is known from previous studies in which researchers proposed solutions to handle modeling of minorities [34]. In addition, captured data will not cover all the needed patterns, **and our approach enables to fuse two sources of knowledge which are data-driven knowledge and expert based knowledge. This approach allows human experts knowledge to bridge any gaps from the data driven models.**

Previous analysis of the first four stages revealed that each stage added extra improvement. In conclusion, a positive answer could be given to the question of whether stages 1-4 provided a reliable improvement to the XAI model designed with assistance from a domain expert.

The second question: *Did the Stage 5 training improve the prediction of the second site over the pretrained first site model? And to what extent has each step in Stage 5 contributed to the resulting system's performance?*

Following the completion of Stage 5 steps, Table X displays the improvement of the prediction rRMSE for the second site data sets in the three case studies during the Fifth Stage. These steps include the following: evaluating these data sets using the pretrained model (5-A), partially training a small number of chosen examples from the first and second datasets (5-B), the expert's participation (5-C), and the rule pruning process (5-D). The overall average results demonstrate that each step produced an incremental improvement, and the average improvement over the pretrained first site rRMSE prediction is 12.28% (Table X). The partial training process in step 5-B contributed the most (an average of 8.27% improvement (Table X). Interestingly, when small deteriorations appeared in step 5-B for case studies 1 and 3, they were recovered in the following step by expert-contributed rules. Thus, the Stage 5 training improves the Stages 1-4 pretrained model in terms of improving the prediction of the second site data set.

The third question: *Considering the explainability and design restrictions of the final model, is the final model capable of predicting the third site dataset, which is an entirely new dataset, with an acceptable level of accuracy relative to benchmark models?*

To measure the prediction efficiency over the third site data set, which is an entirely new dataset, we compared our results against the three main machine learning techniques used widely in this area which are Decision Trees, Bilayered Neural Networks, Gaussian Process Regression and two deep learning models: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). The same sample data sets used in Stages 1-4 and Stage 5-B were used to train these models. All these models, apart from decision trees, are black-box models that cannot be explained by nature. As shown in Table XI, the proposed system outperformed the reference models by an

TABLE XI

COMPARISON WITH SOME BENCHMARK MODELS RMSE RESULTS.

Case Study	Case Study 1					Case Study 2					Case Study 3					Average
Stage	Training S1-S5				Testing	Training S1-S5				Testing	Training S1-S5				Testing	Testing
Site	First		Second		Third	First		Second		Third	First		Second		Third	Third
Sample	Train	Test	Train	Test	Test	Train	Test	Train	Test	Test	Train	Test	Train	Test	Test	Test
Decision Tree (Minimum leaf size of 4).	37.64	64.39	37.72	93.55	**82.44**	34.59	65.41	39.19	55.08	**91.96**	36.59	62.32	36.58	80.07	92.63	**89.01**
Neural Networks (Bilayered).	53.74	55.91	57.53	61.10	**68.65**	49.28	54.78	52.74	47.33	**76.16**	50.23	49.91	56.46	67.95	185.23	**110.02**
Gaussian Process Regression (Squared Exponential Kernel).	46.53	51.16	51.59	59.90	**76.50**	27.64	52.42	33.94	50.54	**85.94**	40.12	46.19	49.25	66.92	86.90	**83.11**
Long short-term memory (LSTM) with 128 hidden units.	86.16	71.33	89.94	96.46	**98.25**	97.67	84.67	98.39	84.42	**99.43**	92.90	85.55	94.31	100.78	**104.14**	**100.61**
Gated recurrent unit (GRU) with 128 hidden units.	91.65	77.02	97.27	105.28	**106.87**	74.54	59.16	77.83	60.69	**81.52**	86.52	80.65	87.24	92.78	**95.82**	**94.74**
Benchmark Mean	63.14	63.96	66.81	83.26	**86.54**	56.75	63.29	60.42	59.61	**87.00**	61.27	64.92	64.77	81.70	**112.95**	**95.50**
Proposed Model Best RMSE	76.08	63.41	76.55	70.48	**72.37**	73.95	60.21	87.02	66.34	**87.25**	78.78	66.73	88.72	85.63	**89.18**	**82.93**
Difference (%)	**-17.0%**	**0.9%**	**-12.7%**	**15.3%**	**16.4%**	**-23.3%**	**4.9%**	**-30.6%**	**-10.1%**	**-0.3%**	**-22.2%**	**-2.7%**	**-27.0%**	**-4.6%**	**21%**	**13.2%**

average of 13.2% across all three test samples for best obtained RMSEs.

In addition, the proposed system outperformed the benchmark models in two of the three test samples, outperforming them by 16.4% and 21%, respectively, with the exception of the second case study, in which it underperformed by only -0.3%. The overall accuracy of the three testing samples in Stages 1-4 was very good, taking into account that the proposed system has restrictions for the sake of explainability and simplicity. The decline in the average prediction accuracy of the benchmark models when predicting test samples relative to the accuracy levels obtained during training can be attributed to either to overfitting or the inability to be tuned with expert opinion for the unseen site which are not issues with the proposed system. As can be seen from Table XI, our proposed system outperforms the benchmarks over the unseen testing sites by 13.2% in average while being 100% transparent and explainable. This is due to the proposed system ability to be adapted using expert opinions to unseen locations while the other techniques cannot be tuned via expert rules as they are data driven (even for decision trees, it is difficult for the expert to make modifications to their structure) and they cannot operate in the absence of data as opposed to our system.

The mean MBE results in Tables IV, VI, and VIII of the three independent testing samples were -22.1, 19.9 and 1.72. These values are very small compared to the scale of GHI values and demonstrate the absence of both underestimation and overestimation behaviors, which increases confidence in the model's generalization capability. Another issue with lifelong learning is how much Stage 5 training has resulted in catastrophic forgetting of the first site prediction. Despite the additional 20–25% of rules added during the fifth stage, the average loss in rRMSE is only -3.4% (Table IX), indicating very little catastrophic forgetting. This refers to the model's excellent generalization ability relative to the benchmark problems. For instance, benchmark results demonstrated the evolution of errors from very small errors for the first and second testing samples to larger errors for the independent testing samples (Table XI), indicating that the proposed model has better generalization.

According to the answers to the three questions, the proposed model can be used to create models that are:

I. Simple, interpretable, and interactive with domain experts; consequently, enhanced as necessary.

II. Capable of handling uncertainty using type-2 fuzzy sets with relatively accurate modeling. and, Generalizable and transferable to new situations through incremental and lifelong learning processes **which is an important functionality in the solar radiation modelling domain which cannot be fully achieved via other existing technique in the literature.**

V. Conclusions and Future Work

The presence of explainability and transferability when modeling solar energy can increase the level of trust between domain experts and AI data-driven models, allowing for more productive interactions and better model utilization. In addition, it is essential in this domain to transfer explainable models between similar locations with the ability to tune such models in the absence of any historical data. This capability did not exist in the literature prior to our work making it the first work to address this major opportunity in the domain of modeling solar energy. This paper presents a five-stage transparent process to design a compact, explainable, interactive, and lifelong learning solar prediction model based on the interval type-2 fuzzy logic system (IT2FLS). In the first stage, the Wang-Mendel algorithm (WM) is utilized, followed by

optimization processes to select a limited number of rules and linguistic variables. The inclusion of domain expert knowledge is then employed to evaluate, modify, and add new relevant rules to the best-found set of fuzzy rules. In the fifth and final stage, the model is converted into a lifelong learning model and transferred to new locations. The objective of the system is to predict hourly solar energy linked to many solar and weather variables. The findings showed that all five stages enhanced the modeling accuracy incrementally, resulting in a straightforward, explainable, and interactable model that could then be improved as necessary. In addition, the model can handle uncertainty using type-2 fuzzy sets with reasonably accurate modeling (13.2% better than benchmark models for unseen geographical locations) and is generalizable and transferrable to new situations through incremental and lifelong learning processes.

Future research may concentrate more on the pruning and consolidation processes to ensure that the model's core remains stable throughout lifelong learning, as well as examining the various combinations and the order of the lifelong learning processes. It may also investigate methods for making data-driven fuzzy rules and expert-driven fuzzy rules more consistent and avoiding duplication and conflicts.

Source code: https://drive.matlab.com/sharing/9cdf1248-317c-47ec-85d4-54263674242d

REFERENCES

[1] IAE, "Renewables 2021 - Analysis and forecasts to 2026." OECD, Paris, 2021. [Online]. Available: www.iea.org

[2] M. B. Hayat, D. Ali, K. C. Monyake, L. Alagha, and N. Ahmed, "Solar energy—A look into power generation, challenges, and a solar-powered future," *Int. J. Energy Res.*, vol. 43, no. 3, pp. 1049–1067, 2019, doi: 10.1002/er.4252.

[3] U. K. Das *et al.*, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, vol. 81, no. June 2017. Elsevier Ltd, pp. 912–928, 2018. doi: 10.1016/j.rser.2017.08.017.

[4] L. Yang, X. Gao, J. Hua, and L. Wang, "Intra-day global horizontal irradiance forecast using FY-4A clear sky index," *Sustain. Energy Technol. Assessments*, vol. 50, no. November 2021, p. 101816, 2022, doi: 10.1016/j.seta.2021.101816.

[5] V. Kostylev and A. Pavlovski, "Solar Power Forecasting Performance - Towards Industry Standards," *1st Int. Work. Integr. Sol. Power into Power Syst.*, 2011.

[6] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy Forecasting: A Review and Outlook," *IEEE Open Access J. Power Energy*, vol. 7, no. August, pp. 376–388, 2020, doi: 10.1109/OAJPE.2020.3029979.

[7] G. Capizzi, C. Napoli, and F. Bonanno, "Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, no. 11, pp. 1805–1815, 2012, doi: 10.1109/TNNLS.2012.2216546.

[8] M. Alrashidi, M. Alrashidi, and S. Rahman, "Global solar radiation prediction: Application of novel hybrid data-driven model," *Appl. Soft Comput.*, vol. 112, no. November 2020, p. 107768, 2021, doi: 10.1016/j.asoc.2021.107768.

[9] A. R. Hedar, M. Almaraashi, A. E. Abdel-Hakim, and M. Abdulrahim, "Hybrid machine learning for solar radiation prediction in reduced feature spaces," *Energies*, vol. 14, no. 23, pp. 1–30, 2021, doi: 10.3390/en14237970.

[10] S. Ghimire, B. Bhandari, D. Casillas-Pérez, R. C. Deo, and S. Salcedo-Sanz, "Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia," *Eng. Appl. Artif. Intell.*, vol. 112, no. November 2021, p. 104860, 2022, doi: 10.1016/j.engappai.2022.104860.

[11] B. Yang *et al.*, "Classification and summarization of solar irradiance and power forecasting methods: A thorough review," *CSEE J. Power Energy Syst.*, 2021, doi: 10.17775/cseejpes.2020.04930.

[12] Y. Zhou, Y. Liu, D. Wang, X. Liu, and Y. Wang, "A review on global solar radiation prediction with machine learning models in a comprehensive perspective," *Energy Convers. Manag.*, vol. 235, 2021.

[13] M. Almaraashi, "Short-term prediction of solar energy in Saudi Arabia using automated-design fuzzy logic systems," *PLoS One*, vol. 12, no. 8, p. e0182429, 2017, doi: 10.1371/journal.pone.0182429.

[14] A. Torres-Barrán, Á. Alonso, and J. R. Dorronsoro, "Regression tree ensembles for wind energy and solar radiation prediction," *Neurocomputing*, vol. 326–327, pp. 151–160, 2019.

[15] H. Hagras, "Toward Human Understandable, Explainable AI," *IEEE Comput. Mag.*, vol. September, pp. 32–40, 2018.

[16] Z. Li, H. Shi, X. Yang, and H. Tang, "Investigating the nonlinear relationship between surface solar radiation and its influencing factors in North China Plain using interpretable machine learning," *Atmos. Res.*, vol. 280, no. July, p. 106406, 2022, doi: 10.1016/j.atmosres.2022.106406.

[17] Y. Gao, S. Miyata, and Y. Akashi, "Interpretable deep learning models for hourly solar radiation prediction based on graph neural network and attention," *Appl. Energy*, vol. 321, p. 119288, 2022.

[18] M. Chaibi, E. L. M. Benghoulam, L. Tarik, M. Berrada, and A. El Hmaidi, "An interpretable machine learning model for daily global solar radiation prediction," *Energies*, vol. 14, no. 21, 2021, doi: 10.3390/en14217367.

[19] Z. C. Lipton, "The mythos of model interpretability," *Commun. ACM*, vol. 61, no. 10, pp. 35–43, 2018, doi: 10.1145/3233231.

[20] S. Jafarzadeh, M. Sami Fadali, and C. Y. Evrenosoglu, "Solar power prediction using interval type-2 TSK modeling," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 333–339, 2013, doi: 0.1109/TSTE.2012.2224893.

[21] P. P. Mendel, Jerry M and Bonissone, "Critical Thinking About Explainable AI (XAI) for Rule-Based Fuzzy Systems," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 12, pp. 3579–3593, 2021.

[22] A. J. Monticelli, R. Romero, and E. N. Asada, "Fundamentals of Simulated Annealing," *Mod. Heuristic Optim. Tech. Theory Appl. to Power Syst.*, no. 1, pp. 123–146, 2007, doi: 10.1002/9780470225868.ch7.

[23] S. Anily and A. Federgruen, "Ergodicity in Parametric Nonstationary Markov Chains: an Application To Simulated Annealing Methods.," *Oper. Res.*, vol. 35, no. 6, pp. 867–874, 1987, doi: 10.1287/opre.35.6.867.

[24] A. Baraka and G. Panoutsos, "Long-term learning for type-2 neural-fuzzy systems," *Fuzzy Sets Syst.*, vol. 368, pp. 59–81, 2019, doi: 10.1016/j.fss.2018.12.014.

[25] C. Febrinanto, Falih Gozi and Xia, Feng and Moore, Kristen and Thapa, Chandra and Aggarwal, "Graph lifelong learning: A survey," *IEEE Comput. Intell. Mag.*, vol. 18, no. 1, pp. 32--51, 2023.

[26] M. Setnes, R. Babuška, U. Kaymak, and H. R. Van Nauta Lemke, "Similarity measures in fuzzy rule base simplification," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 28, no. 3, pp. 376–386, 1998, doi: 10.1109/3477.678632.

[27] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Sol. Energy*, vol. 85, no. 11, pp. 2856–2870, 2011, doi: 10.1016/j.solener.2011.08.027.

[28] KACARE, "Renewable resource atlas - King Abdullah City for Atomic and Renewable Energy." http://rratlas.energy.gov.sa (accessed Oct. 01, 2015).

[29] J. Muñoz Sabater, "ERA5-Land hourly data from 1981 to present," *Copernicus Climate Change Service (C3S) Climate Data Store*, 2019.

[30] R. T. Haltiner, George J and Williams, *Numerical prediction and dynamic meteorology*. 1980.

[31] L.-X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst. Man. Cybern.*, vol. 22, no. 6, pp. 1414–1427, 1992, doi: 10.1109/21.199466.

[32] J. M. Mendel, *Uncertain rule-based fuzzy logic systems: introduction and new directions*. Prentice Hall, 2001.

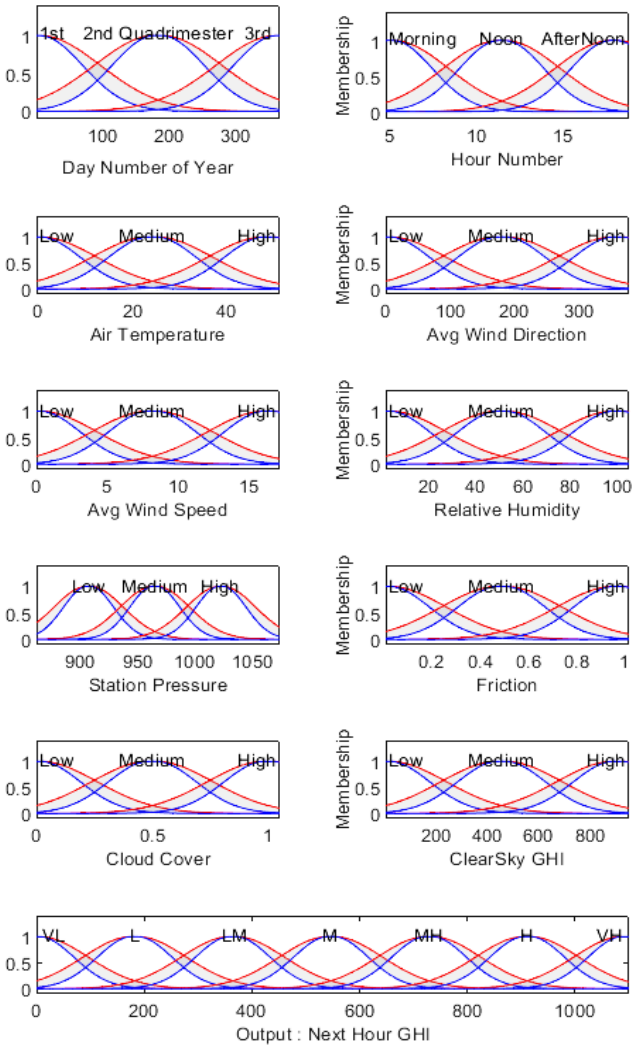
[33] M. Nie and W. W. Tan, "Towards an efficient type-reduction method for interval type-2 fuzzy logic systems," in *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, Jun. 2008, pp. 1425–1432. doi: 10.1109/FUZZY.2008.4630559.

[34] J. A. Sanz, D. Bernardo, F. Herrera, H. Bustince, and H. Hagras, "A Compact Evolutionary Interval-Valued Fuzzy Rule-Based Classification System for the Modeling and Prediction of Real-World Financial Applications with Imbalanced Data," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 973–990, 2015, doi: 10.1109/TFUZZ.2014.2336263.

[35] Sadalmelik, "Topographic maps of Saudi Arabia," *Wikimedia.org*, 2007. https://commons.wikimedia.org/wiki/User:Sadalmelik#/media/File:Saudi_Arabia_Topography.png (accessed Feb. 01, 2023).

APPENDIX A

AN EXAMPLE OF GENERATED INPUT AND OUTPUT FUZZY SETS.



APPENDIX B

SAMPLE OF GENERATED RULES FROM THE OPTIMIZATION PROCESS AND ADDED EXPERT RULES (A= Automated, E=Expert).

Rule	Source
If (Day Number is 1st Quadrimester of Year) and (Hour Number is Morning) and (Clear Sky GHI is Low) then (Next Hour GHI is VERY LOW)	A
If (Day Number is 1st Quadrimester of Year) and (Air Temperature is Low) and (Cloud Cover is High) then (Next Hour GHI is VERY LOW)	A
If (Day Number is 1st Quadrimester of Year) and (Avg Wind Direct is Low) and (Relative Humidity is High) then (Next Hour GHI is VERY LOW)	A
If (Hour Number is Noon) and (Friction is Medium) and (Cloud Cover is Low) then (Next Hour GHI is High)	A
If (Hour Number is Noon) and (Avg Wind Speed is Low) and (Cloud Cover is Low) then (Next Hour GHI is High)	A
If (Relative Humidity is Medium) and (Cloud Cover is Medium) and (Clear Sky GHI is High) then (Next Hour GHI is Medium High)	A
If (Relative Humidity is Low) and (Cloud Cover is Low) and (Clear Sky GHI is Medium) then (Next Hour GHI is Medium)	A
If (Day Number is 1st Quadrimester of Year) and (Hour Number is Morning) and (Avg Wind Speed is Low) and (Relative Humidity is High) and (Station Pressure is High) then (Next Hour GHI is LOW) / Fog Cases	E
If (Day Number is 3rd Quadrimester of Year) and (Hour Number is Morning) and (Avg Wind Speed is Low) and (Relative Humidity is High) and (Station Pressure is High) then (Next Hour GHI is LOW) / Fog Cases	E
If (Day Number is 1st Quadrimester of Year) and (Avg Wind Speed is High) and (Relative Humidity is Low) and (Avg Wind Direct is High) then (Next Hour GHI is Low) / Dust Cases	E
If (Day Number is 3rd Quadrimester of Year) and (Avg Wind Speed is High) and (Relative Humidity is Low) and (Avg Wind Direct is High) then (Next Hour GHI is Low) / Dust Cases	E