Article

Hybrid Machine Learning for Solar Radiation Prediction in Reduced Feature Spaces

Abdel-Rahman Hedar [1,2], Majid Almaraashi [3], Alaa E. Abdel-Hakim [4,5*], and Mahmoud Abdulrahim [6]

[1] Department of Computer Science in Jamoum, Umm Al-Qura University, Makkah 25371, Saudi Arabia; ahahmed@uqu.edu.sa
[2] Department of Computer Science, Assiut University, Assiut 71526, Egypt‡; hedar@aun.edu.eg
[3] Department of Computer Sciences, College of Computing and Information Technology, University of Jeddah, Jeddah 23218, Saudi Arabia; malmaraashi@uj.edu.sa
[4] Department of Computer Science in Jamoum, Umm Al-Qura University, Makkah 25371, Saudi Arabia; adali@uqu.edu.sa
[5] Electrical Engineering Department, Assiut University, Assiut 71516, Egypt; alaa.aly@eng.au.edu.eg
[6] Department of Meteorology, Faculty of Meteorology, Environment and Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia; mahussen@kau.edu.sa
* Correspondence: alaa.aly@eng.au.edu.eg
‡ Current address

Abstract: Solar radiation prediction is an important process in ensuring optimal exploitation of solar energy power. Numerous models have been applied to this problem, such as numerical weather prediction models and artificial intelligence models. However, well-designed hybridization approaches that combine numerical models with artificial intelligence models to yield a more powerful model can provide a significant improvement in prediction accuracy. In this paper, we propose novel hybrid machine learning approaches that exploit auxiliary numerical data. The proposed hybrid methods invoke different machine learning paradigms, including feature selection, classification, and regression. Additionally, numerical weather prediction (NWP) models are used in the proposed hybrid models. Feature selection is used for feature space dimension reduction to reduce the large number of recorded parameters that affect estimation and prediction processes. The rough set theory is applied for attribute reduction and the dependency degree is used as a fitness function. We investigate the effect of the attribute reduction process with thirty different classification and prediction models in addition to the proposed hybrid model. Then, different machine learning models are constructed based on classification and regression techniques to predict solar radiation. Moreover, other hybrid prediction models are formulated to use the output of the numerical model of Weather Research and Forecasting (WRF) as learning elements in order to improve the prediction accuracy. The proposed methodologies are evaluated using a data set that is collected from different regions in Saudi Arabia.

Keywords: solar energy; solar radiation prediction; hybrid machine learning; feature selection; feature extraction; classification algorithms; regression analysis; weather research and forecasting (WRF)

1. Introduction

Solar energy is considered as a major source for future renewable energy [1]. As the dependence on renewable energy increases, more attention to solar energy is paid. Solar radiation data is the main ingredient of optimum design and operations of solar power systems [2]. It is necessary to ensure the stability of the energy supplied by solar stations. Therefore, accurate prediction of the amount of solar radiation at a specific location is critical from an operational perspective. For parties such as governments, enterprises, and energy operators, solar radiation prediction is a key for optimal strategic plans, particularly when hybridized with different energy sources. However, such an objective

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

is associated with practical difficulties. Particularly, the potential of solar energy is limited by inaccuracy of solar radiation levels prediction when compared with certain alternative resources. In order to handle this problem, several prediction models have been proposed in the literature to predict solar radiation, including numerical weather prediction (NWP) and artificial intelligence models, e.g. [3–8]. However, the large number of parameters associated with the prediction process, including weather and topography variables, significantly affects the underlying prediction models. Therefore, it is crucial to obtain a good representative set of these parameters, or features as termed in machine learning, to improve the predictor performance as well as reducing the computational cost of the real-time prediction systems.

NWP models can provide forecasts of solar radiation several days ahead along with other weather parameters, such as temperature, air pressure, relative humidity, or wind speed [9]. Such information can be useful for optimizing solar plant operating strategies. These models rely on atmospheric reanalysis to obtain initial and boundary conditions for the model run before it is realistically downscaled to a finer physical resolution using few physical equations. An NWP model that downscales reanalysis data is called a mesoscale model. As mesoscale models run within a smaller area compared with global-scale models, they include additional details. Therefore, these models can provide forecasts of solar irradiance with a high temporal spatial resolution over a wide area but with high levels of computing power. The Weather Research and Forecasting (WRF) model [10] is the most commonly-used mesoscale model, and it has been extensively applied and assessed. In this study, a nonhydrostatic WRF v3.7.1 model has been applied to simulate dust storm events over Saudi Arabia to evaluate the reliability of global horizontal irradiance (GHI) forecasts.

Regarding artificial intelligence (AI) models, A large number of AI models for predicting solar radiation or solar power have been proposed. For example, AI models have been applied to predict solar radiation using fuzzy logic sets and systems [11] [12], neuro-fuzzy systems [13], neural networks [14–16], machine/deep learning [17–24], and LSTM [25–29].

Abdel-Nasser et al. [30] developed an LSTM-based method for solar irradiance forecasting. They used LSTM models with an aggregation function based on Choquet integral. Combining Choquet integral with LSTM aimed at achieving more accurate predictions due to the memory units and the recurrent architecture which can model the temporal changes in solar irradiance. The interaction between aggregated inputs are modeled by the Choquet integral through a fuzzy measure.

Almaraashi [31] applied fuzzy logic systems that are designed and optimized using fuzzy c-means clustering (FCM) and simulated annealing (SA) algorithms to forecast global horizontal irradiance (GHI) in eight stations in Saudi Arabia. In addition, Almaraashi predicted daily solar radiation in the same eight stations in Saudi Arabia using multi-layer neural networks (NNs). This was done after applying four-feature selection methods to discover the most important variables [32]. The used four-feature selection methods are the Relief algorithm, Random-Frog algorithm, Monte Carlo Uninformative Variable Elimination algorithm (MCUVE), and Laplacian Score algorithm (LS). A hybrid model presented by Voyant et al. [33] applied the NWP model combined with a hybrid auto-regressive moving average (ARMA) and neural networks to forecast hourly global radiation for five locations in the Mediterranean area.

Boubaker et al. [34] have investigated one-day prediction of GHI using various DNN models at Hail city, Saudi Arabia. They used six different DNN models: LSTM, BiLSTM, GRU, Bi-GRU, onde dimensional CNN, and other hybrid configurations such as CNN-LSTM and CNN-BiLSTM. The used DNN models depend only on historical daily values of GHI. However, These models did not take into consideration crucial weather parameters that may affect GHI, e.g. air temperature, humidity, wind speed, wind direction, and atmospheric pressure.

 doi:10.20944/preprints202111.0202.v1

The intuitive parameter selection by experts when predicting solar radiation can result in different sets of possible input parameters in which some might appear to be redundant or irrelevant. In addition, the manual selection of the most relevant features for this problem is affected by the large dimensionality of the input feature space. Given such a case, the automatic dimension reduction of the input feature space can be a valuable solution.

Solar energy prediction needs large amounts of data, which require a large number of measuring devices and equipment. Moreover, the calculations of weather data needed for the solar energy prediction process are often computationally expensive. Therefore, one of the most important motivations for this research is to reduce the data reading and calculation processes required for solar energy prediction and to reduce the cost of this process. This helps to expand prediction operations in a broader and more comprehensive way, even beyond the scope of traditional measurement stations. Another major motivation for this study is to use the power of smart and hybrid systems in predicting short-term solar energy levels. Therefore, in this paper, we present a modified version of the tabu search attribute reduction (TSAR) [35] as a feature selection method along with different prediction models for the estimation of solar radiation levels. The main modifications of that method are adding more local search and extending some other search operations. Consequently, various classification and regression models are designed to predict solar radiation based on reduced features. Moreover, other hybrid predictive models are formulated in order to utilize the outputs of the WRF numerical model as learning elements to increase prediction accuracy. In addition to the proposed prediction models, we investigate the impact of the attribute reduction mechanism on different classification and regression models. Therefore, this research focuses on finding solutions to the following research questions and gaps:

- What is the main solar and weather feature set that should be efficiently used to predict the short-term solar radiation?
- Which machine learning models can be likely more efficient for the considered prediction problem?
- To what extent can hybrid models of machine learning and numerical weather prediction succeed in the prediction?
- Does the feature selection process still work with hybrid models?

The remainder of the paper is organized in the following manner: the proposed models and methodologies are designed in Section 2. The setup and evaluation of the experiments are discussed in Section 3. Then, the details of output results and technical discussion are presented in Section 4. Finally, the conclusion to the paper is presented in Section 5.

2. Methodology

The proposed methodology comprises different design elements, including feature selection, classification, regression, and numerical models. In this section, these design elements are discussed and their integration in creating our prediction models are illustrated. The main layout of the proposed methodology is presented in Figure 1. Four possible methods can be implemented using this layout based on:

- invoking reduced solar feature data or not; and
- applying only machine learning prediction models or hybrid models with the numerical WRF models.

2.1. Feature Selection

The proposed feature selection (FS) method is designed on the basis of a modified version of the Tabu Search Attribute Reduction (TSAR) method [35]. The proposed method adds more local search and extends some other search operations. The modified feature selection is denoted by *m*TSAR. The FS method selects the best features in order to use them in building classifiers and prediction models. The main steps in the work-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

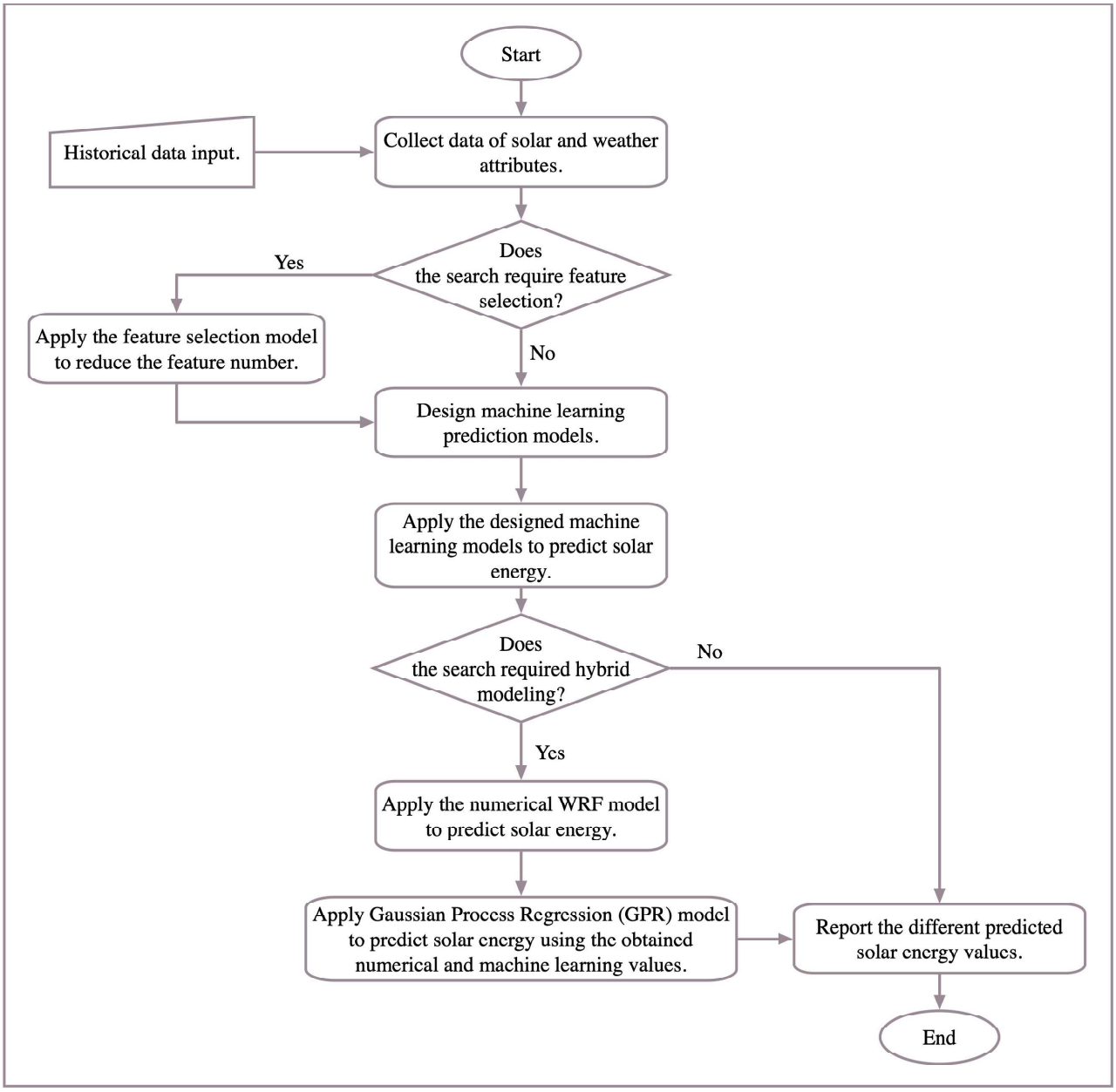

Figure 1. The flowchart of the proposed solar prediction models.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

flow of the presented method are highlighted in Figure 2, and detailed in the following subsections.

2.1.1. Solution Representation

The FS method encodes its solutions in binary vectors. The dimension of these vectors is equal to the number of conditional features. Therefore, if the entity of the coding-vector has a value of 1, it implies that the corresponding feature is included in the solution represented by this vector. Otherwise, this feature is not included in that solution.

2.1.2. Feature Selection Evaluation

The dependency degree concept in the rough set theory [36] is invoked to evaluate the goodness of reducts or solutions. Therefore, the feature selection problem can be defined in terms of maximizing the dependency degree values of the solutions and minimizing their cardinality. The dependency degree function of a solution (feature reduct) can be computed using the following definitions [36]:

- **Indiscernibility Relation.** Given a set A of all condition features and a feature subset $P \subseteq A$, the indiscernibility relation is denoted by $IND(P)$ and defined as:

$$IND(P) = \{(\xi, \eta) \in U \times U \mid \forall a \in P, a(\xi) = a(\eta)\}.$$

- **Indiscernibility Equivalence.** The relation $IND(P)$ forms an equivalence relation on the set U. The relation $IND(P)$ represents a partition of U denoted by $U/IND(P)$. For any pair $(\xi, \eta) \in IND(P)$, it can be said that ξ and η are indiscernible by features of P. The P-indiscernibility equivalence classes are denoted by $[\xi]_P$.
- **Lower and Upper Approximation.** Given a subset $\Xi \subseteq U$, we define the P-lower approximation of Ξ by:

$$\underline{P}\Xi = \{\xi | [\xi]_P \subseteq \Xi\}.$$

  Moreover, we define the P-upper approximation of Ξ by:

$$\overline{P}\Xi = \{\xi | [\xi]_P \cap \Xi \neq \emptyset\}.$$

- **Positive Region.** The positive region of the partition of $U/IND(Q)$ with respect to P is defined as the set of all members of U that can be uniquely classified to blocks of the partition $U/IND(Q)$ using the knowledge in P, which can be determined by:

$$POS_p(Q) = \bigcup_{\Xi \in U/IND(Q)} \underline{P}\Xi.$$

- **Dependency Degree (γ).** The dependency degree is the ratio of all objects of U that can be appropriately classified to the blocks of the partition $U/IND(Q)$ by means of P. This dependency degree is denoted by $\gamma_P(Q)$ and determined by:

$$\gamma_P(Q) = \frac{|POS_P(Q)|}{|U|},$$

  where $|\cdot|$ is the cardinality measure.

Therefore, the dependency degree can be stated as the ratio of all objects of U that can be classified to the blocks of the partition $U/IND(Q)$ using P.

A feature subset Q is said to depend totally or partially on another feature subset P, if $\gamma_P(Q) = 1$, or $\gamma_P(Q) < 1$, respectively. In order to measure the quality of a solution x, we can use the dependency degree $\gamma_x(D)$ of decision attribute D. Therefore, for two solutions x and y, we can say the x is better than y, if one of the following conditions holds:

- $\gamma_x(D) > \gamma_y(D)$,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

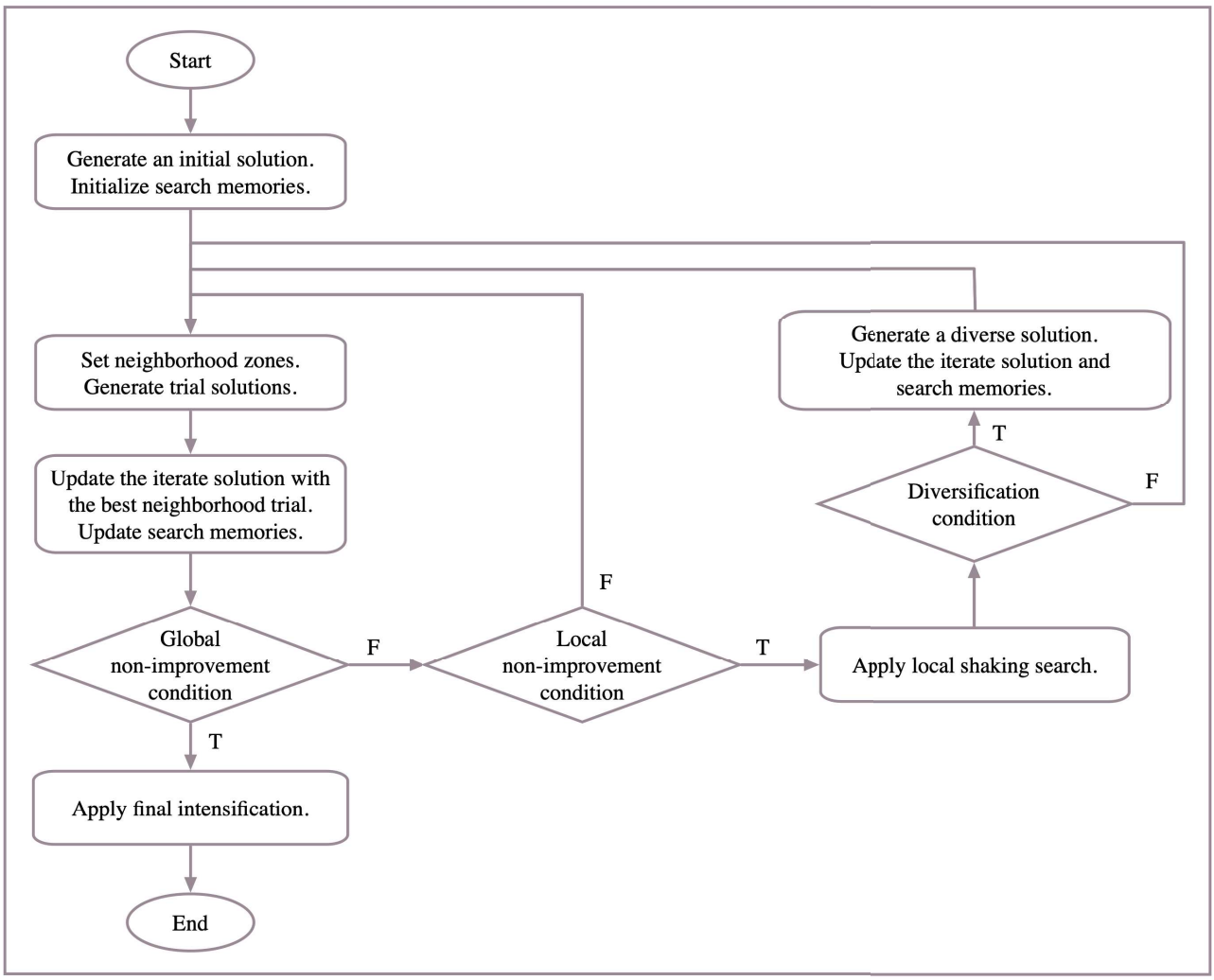

Figure 2. The flowchart of the proposed feature selection *m*TSAR method.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

- $\gamma_x(D) = \gamma_y(D)$, and $|x| < |y|$,

where $|x|$ and $|y|$ are the number of features in x and y, respectively.

2.1.3. Initialization

A random binary vector is generated as an initial solution. Lists of tabu and elite are initialized as empty lists. The most recently visited solutions are placed in the tabu list to prevent being trapped in local optimal solutions. Moreover, the elite list contains the best solutions found up to now. Then, they can be included in the steps of intensification.

2.1.4. Search Procedures

The key search procedures of the proposed FS system are identical with minor modifications to those of our previously-published system [35]. Specifically, the proposed method begins with an initial solution and continues to produce trial solutions within the neighborhood of the current one. The stop criterion is met when no improvement is accomplished during of a predefined number of consecutive iterations. Thereafter, the search process initiates a diversification step from a new diverse solution. If the number of such consecutive non-improvement iterations reaches another pre-defined number of iterations, an intensification step is initiated to refine the best solution achieved so far. If the number of iterations reaches a maximum permitted iteration limit, the search is terminated. Lastly, the search process uses a final step for diversification-intensification to obtain the final solution. The details of the neighborhood and local search procedures are explained below.

- **Neighborhood Search.** The neighborhood of the current iterate solution $x = (x_1, \ldots, x_n)$ is broken down into a fixed number of neighborhood zones denoted by Z^j, $j = 1, \ldots, \ell$, and expressed as:

$$Z^j(x) = \{x^j : x^j = (x_1^j, \ldots, x_n^j)\},$$

where

$$x_i^j \begin{cases} \neq x_i, & \forall i_1, \ldots, i_j \in \{1, \ldots, n\}, \text{ and } i_1 \neq \cdots \neq i_j, \\ = x_i, & \text{otherwise.} \end{cases}$$

Within each zone, the search process generates a trial solution in accordance with the tabu restriction to avoid revisiting recent solutions.

- **Solution and Memory Updates.** The next iterative solution is selected as the best trial one among the generated solutions. Thereafter, the tabu and elite lists are revised.
- **Local Search.** Using a local search technique called Shaking [35], the best solution is improved by attempting to sequentially eliminate its attributes without raising its degree of dependency value. The steps of the shaking procedure are depicted in Figure 3, which is a modified version of the original shaking procedure in [35]. The standard shaking technique [35] is used only to lower the cardinality of the best reducts whose γ function values are equal to 1. However, the modified shaking procedure applies the feature reductions to both total or partial reducts.

2.1.5. Diversification

Whenever diversification is required, it is possible to generate a new diverse solution. The attributes included in a diverse solution are generated with probability that is inversely proportional to their appearance in the previously generated solutions.

2.1.6. Final Intensification

In order to produce new promising solutions, the common features that appear in the elite solutions can be utlilized. In particular, the reducts obtained are stored in a package called the Reduct Set (RedSet). Then, the term “core” is defined as the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

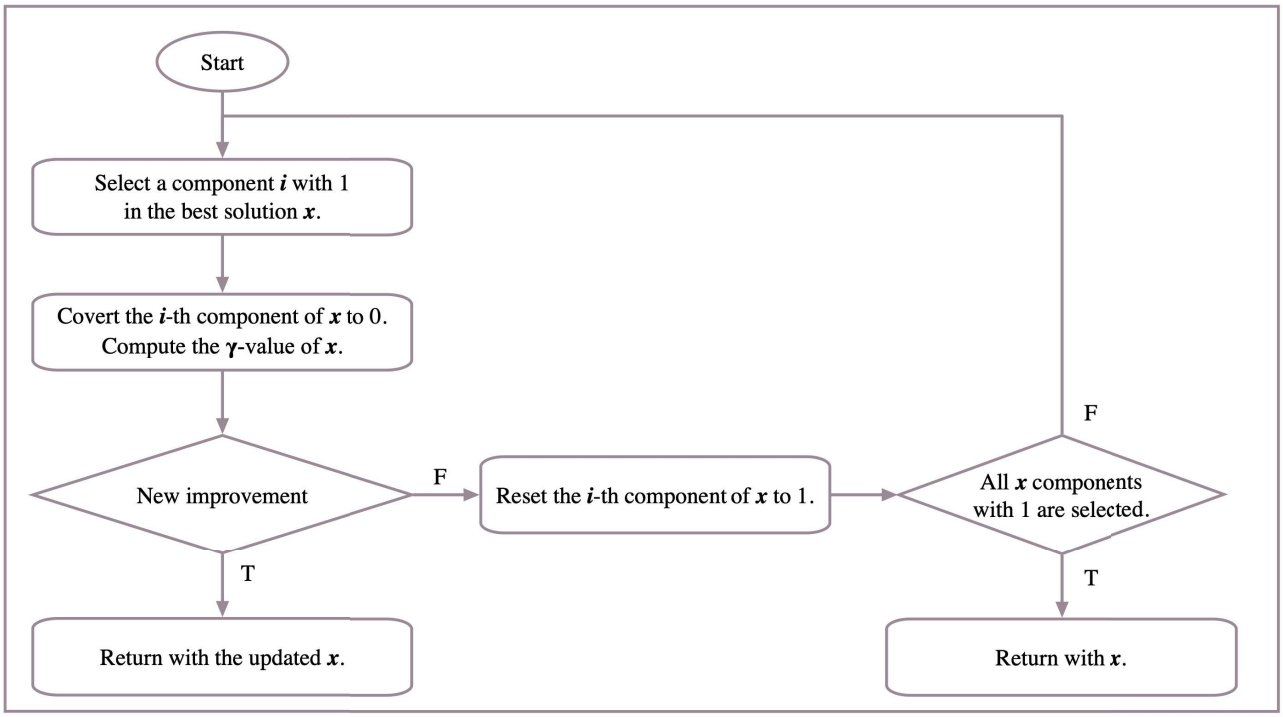

Figure 3. Reducing the cardinality of the best solution using the shaking procedure.

intersection of all reducts saved in the RedSet. Thus, a test solution x^{Final} is constructed as the intersection of the best m reducts in RedSet. Therefore, a new solution candidate x^{Final} is generated to contain the core. The trial solution x^{Final} is only considered if its number of features is less by at least two than the best obtained reducts. Thereafter, new features will be added to x^{Final} by converting the zero positions in x^{Final} in which the highest γ-value is given one. This upgrading process continues until a suitable new solution is found.

2.1.7. Control and Termination

Three non-improvement counters (I_{local}, I_{div}, I_{global}) are used to control the processes of applying the local search, diversification, and final intensification, where $I_{local} < I_{div} < I_{global}$. Specifically, if a number of non-improvement iterations, I_{local}, is reached, the shaking procedure is utilized. Then, if the number of non-improvement iterations is increased and reaches I_{div}, a new diverse solution is generated. Finally, if the number of non-improvement iterations exceeds I_{global}, then the final intensification is employed to refine the best solutions in RedSet.

2.2. Prediction Models

Several solar radiation prediction models for the global horizontal irradiance (GHI) values are proposed on the basis of classification, regression, numerical and hybrid techniques. The target is to obtain predicted values of the GHI (in Wh/m^2 per day) through regression models or predicted classes of different ranges of the GHI values through classification models. All prediction models are created in two versions; with or without feature selection. The numerical models of weather forecasting are inlaid within hybrid regression models to obtain improved predicted values that are hopefully better than the values obtained by the pure numerical or machine learning models.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

2.2.1. Machine Learning Models

The range of solar radiation energy can be discretized into a certain number of classes. Then, several classifiers are used to predict the classes of solar radiation energy. The following classifiers are invoked in this study.

- **Decision Trees.** Binary decision trees are multi-class learners in which decisions are followed in the shape of a tree, from its root node down to its leaf nodes that contain the response [37]. Different structures of decision trees can be used in classification based on the number of leaves used to make distinctions among classes. The number of leaves could be low, medium, or high corresponding to classification models coarse, medium or fine decision trees, respectively. The optimizable model employs certain techniques to automatically tune model hyper-parameters.
    1. Fine Decision Tree
    2. Medium Decision Tree
    3. Coarse Decision Tree
    4. Optimizable Decision Tree
- **Discriminate Analysis.** Discriminant analysis assumes that data are generated by different classes based on different Gaussian distributions. Therefore, these classifier models attempt to estimate the parameters of a Gaussian distribution that fit each class [38]. Two common types of such classifier models are linear and quadratic discriminant analysis apart from the optimizable discriminate analysis in which model hyper-parameters are automatically tuned:
    5. Linear Discriminate Analysis
    6. Quadratic Discriminate Analysis
    7. Optimizable Discriminate Analysis
- **Naïve Bayes Classifiers.** These classifiers classify data in two steps. In the first step, the classifier estimates the probability distribution parameters assuming that the predictors are conditionally independent, given the class based on some training data. In the second step, the classifier considers other unseen test data samples and computes the posterior probability of that samples belonging to each class [39]. Then, the method classifies the test data according to the largest posterior probability. Such a classifier model invokes different techniques, such as Gaussian, kernel predictors, or an optimizable technique:
    8. Gaussian Naïve Bayes
    9. Kernal Naïve Bayes
    10. Optimizable Naïve Bayes
- **Support Vector Machine (SVM).** This classifier uses a separate hyper-plane to classify data into two classes [40]. Different trick kernels can be utilized, if the data is not linearly separable. Moreover, the classifier can deal with multi-class classification through different upgrading criteria. Several kernels and modifications can be used to design the following SVM models.
    11. Linear SVM
    12. Quadratic SVM
    13. Cubic SVM
    14. Fine Gaussian SVM
    15. Medium Gaussian SVM
    16. Coarse Gaussian SVM
    17. Optimizable SVM
- **Nearest Neighborhood Classifiers.** A nearest neighbor can find other nearest neighbors within a defined distance to search data points based on specified distance metrics such as Euclidean and Hamming [41,42].
    18. Fine KNN
    19. Medium KNN

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

20. Coarse KNN
21. Cosine KNN
22. Cubic KNN
23. Weighted KNN
24. Optimizable KNN Classifiers

- **Ensemble Classifiers.** A classification ensemble is a prediction model that comprises a weighted combination of several models for classification. In general, combining multiple classification models improves predictive performance. Ensemble classifiers use boosting, random forest, bagging, random subspace, and error-correcting output codes ensembles for multi-class learning [43].

25. Ensemble Boosted Decision Trees
26. Ensemble Bagged Decision Trees
27. Ensemble Subspace Discriminant Analysis
28. Ensemble Subspace KNN
29. Ensemble RUS Boosted Decision Trees
30. Optimizable Ensemble Classifiers

- **Neural Networks.** Artificial neural networks are a subset of machine learning that are at the heart of deep learning algorithms. Their name and structure are based on the human brain, and they mimic the way in which organic neurons communicate starting from the input layer to output layer passing through hidden layers. They prove great success in different applications [44]. The following neural network models are used with various size and number of hidden layers.

31. Narrow Neural Networks
32. Medium Neural Networks
33. Wide Neural Networks
34. Bilayered Neural Networks
35. Trilayered Neural Networks

Regression models can be implemented to predict certain amounts of solar radiation energy by estimating GHI values. One of the most powerful regression models is the Gaussian Process Regression (GPR) model. The GPR model is a non-parametric kernel-based probabilistic model [45,46], which measures the similarity between training data to predict the value for test data.

2.2.2. Numerical Model

Numerous prediction models using NWP of solar radiation have been applied [4,5]. A recent and effective numerical model is the WRF mesoscale model [47]. The WRF model serves both needs for atmospheric research and operational forecasting. The WRF model is fitted with two dynamic (computational) cores, a data assimilation system and a software architecture, thereby enabling for parallel computation and system scalability.

The WRF model is a mesoscale model developed by a group of scientists from different institutes and centers, such as the National Center for Atmospheric Research (NCAR), National Centers for Environmental Prediction (NCEP), National Oceanic and Atmospheric Administration (NOAA), and a number of other collaborating institutes and universities. The WRF is a fully compressible nonhydrostatic three-dimensional (3D) primitive equation model that is designed for simulating atmospheric phenomena across scales. These scales varies from large eddies ($\sim$100m) to mesoscale circulations and waves (from $\sim$ 100m to $>$ 1000km).

The WRF system provides different physics options for cloud parameterization, planetary boundary layer (PBL) turbulence physics, atmospheric radiation, and land surface models (LSMs). It also incorporates various initialization routines and data assimilation techniques that numerous weather agencies and research centers have extensively tested. Additional manuals and descriptions of the WRF model are fully documented in [48,49].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

Figure 4. The layout of the hybrid prediction regression model

2.2.3. Hybrid Model

In order to improve the prediction process, the machine-learning models can use the output of the numerical models. In this research, a hybrid model is designed by using known weather data and the estimated future data of solar radiation energy obtained by the WRF model to build a new hybrid model for short-term solar radiation energy. Therefore, a Gaussian Process Regression (GPR) model is utilized to do this job. Actually, the GPR models are probabilistic ones with non-parametric kernel-based structures [46]. The proposed hybrid model is highlighted along with a corrector model that enhances the prediction values of the WRF model by using the machine learning of the GPR.

Figure 4 illustrates a high-level structure of the hybrid model that uses two types of input data. The first input data is the historical solar features including the values of GHI of the m previous days, denoted by $x_1, x_2, \ldots, x_m$. The other input data is the predicted value of GHI on the considered day for prediction, denoted by y. Then, the GPR model uses these inputs to predict a new GHI value.

Specifically, the designed GPR model predicts new GHI values, which are expected to be more accurate than the ones predicted by the WRF models. Consider the input vector $X = (x_1, x_2, \ldots, x_m, y)$, then a new GHI value y' can be computed from the following regression model:

$$y' = X^T\beta + \epsilon,$$

where ϵ is generated for the normal distribution $\mathcal{N}(0, \sigma^2)$, and β, σ are estimated from the training data[46]. In order to deal with the non-linearity, a kernel-based structure can be used to modify the above-mentioned model to be:

$$y' = h(X)^T\beta + f(X),$$

where $h(\cdot)$ are basis transformation functions and $f(\cdot)$ is a Gaussian function [46].

3. Experimental Setup and Evaluation

We exploit available observed historical data in order to measure the performance of the reduction in the input feature space. In particular, the impact of dimension reduction on the solar radiation estimation process is investigated. This investigation is conducted by measuring weather data variables, such as temperature, wind speed, humidity and direct normal irradiance, as well as other environmental data. Table 2 enumerates the attributes used for evaluation purposes. Evaluation of the proposed system is performed by setting the GHI for the current day as the objective output.

We design an experiment to evaluate the discrete energy class prediction with and without feature selection. For compliance with typical classification frameworks, we discretize the GHI measurements into a finite number of levels. The range of the recorded GHI expands between 0 and 9000. We generated two discrete sets. The first set is called 5-class, which comprises five levels of GHI values. Each level contains approximately 2000 values of GHI. Similarly, the other set, 10-class, contains ten different classes representing ten discrete GHI levels of approximately 1000 for each one.

In order to measure the candidate reduction of the input feature space, we use three data sets that were collected at distant stations distributed around Saudi Arabia. The locations of these stations exhibit diverse climatic conditions. The diversity of these cities in terms of locations and topographies has supported the choice of these cities for our experimentation. Furthermore, the research nature of the installed stations in these

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

Table 3: Dates of widespread dust storm events covering different areas of Saudi Arabia on 2014.

Month	Days
January	19, 27
February	24
March	3, 9, 12, 16, 24, 27, 31
April	1, 3, 11, 15, 19, 27, 30
May	3, 7, 10, 13, 19, 23
June	5, 12, 16, 18, 22
July	5, 9, 13, 18, 20, 31
August	18
October	9, 11, 14, 16, 21
November	4

cities makes it easy to obtain the necessary solar data. King Abdullah City for Atomic and Renewable Energy (KACARE) has installed and monitors these stations under the Renewable Resource Monitoring and Mapping (RRMM) Program [50,51]. The main weather measurement that is used for evaluation is the GHI. The data sets are collected for three Saudi cities on a daily basis from mid-2013 to the end of 2014. Comprehensive evaluation is performed using these data sets. However, because of technical issues with some of these recently-installed KACARE stations, two important readings are missing during this period: visibility and sky cover parameters. This is apart from the obvious uncertainty associated with all other measurements. Therefore, to overcome this issue, we used another source to obtain the visibility variable data — the Presidency of Meteorology and Environment stations. As depicted in Table 2, only two cities out of the used three have the visibility parameter recorded.

In order to evaluate the GHI prediction performance of the proposed hybrid learning model, which is used for regression in this case, other data sets were selected to cover four different levels of challenge: clear, cloudy, dusty, and dusty-cloudy [8]. Table 3 [8] presents the challenging cases, including 41 dust storms. The solar attributes are collected on those dates and the following days thereby leading to the creation of data sets with 81 records at three stations KAU, QU, and TU. If one day after the storm is recorded in addition to the storm days, it should add up to 82 days rather than 81. There is one day missing because one of the storms lasted for two consecutive days. The prediction performance is done by feeding measurements of preceding days to the regression process in order to predict GHI in these specific 81 days.

These 41 cases reveal a clear seasonality changes in the observed frequency of dust storms during 2014. The highest frequency of events are during the spring and summer (March – August), whereas the lowest number of dust storms events took place in the autumn and winter (September – November). More details are found in [52].

The simulations of the severe dust storm events over Saudi Arabia are performed using the WRF with the dynamic core of the Advanced Research WRF (ARW). The WRF model provides two-day hourly forecasting for surface solar irradiance for specific cases in 2014. The atmospheric dust aerosol is indirect data that is highly correlated with the solar radiation at the surface. Specifically, an increase in atmospheric aerosol dust will immediately turn into solar irradiance reduction on the surface. Consequently, improving the aerosol forecasting leads to more accurate prediction of surface solar radiation.

For the solar irradiance prediction process, we use the forceasting data of the National Centers for Environmental Prediction (NCEP), which follows the Global Forecast System (GFS) model [53]. As a preprocessing step, we downscale these GFS forecasts both spatially and temporally. Four daily samples of NCEP GFS are given at 0 UTC, 6 UTC, 12 UTC, and 18 UTC. The temporal and spatial resolutions are three hours and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

Table 1: The data of the stations and their recorded measurements.

Station	City	Latitude (N)	Longitude (E)	Elevation (m)	Data Samples
King Abdulaziz Univ. (KAU)	Jeddah	21.49604	39.24492	75	582
Qassim Univ. (QU)	Qassim	26.34668	43.76645	688	576
Taif Univ. (TU)	Taif	21.43278	40.49173	1518	575

Table 2: Solar attributes used in the current experiment.

Attributes	Abbreviation	KAU	QU	TU
Air Temperature (Degrees C)	T	✓	✓	✓
Average Wind Direct at 3m (Deg North)	WD	✓	✓	✓
Average Wind Speed at 3m (m/s)	WS	✓	✓	✓
Diffuse Horizontal Irradiance (Wh/m2)	DH	✓	✓	✓
Direct Normal Irradiance (Wh/m2)	DN	✓	✓	✓
Peak Wind Speed at 3m (m/s)	PWS	✓	✓	✓
Relative Humidity (Percent)	H	✓	✓	✓
Station Pressure (mB (hPa equivalent))	P	✓	✓	✓
Visibility	V	✓	✓	×

$0.5^\circ \times 0.5^\circ$, respectively. The forecast accuracy evaluation is performed by comparing the GHI forecasts of WRF with the obtained ground measurements. Land cover and elevation and land cover data were obtained from the digital terrain model of the United States Geological Survey [54].

In order to represent different weather conditions, simulations of the aforementioned cases were obtained using non-hydrostatic WRF-ARW mesoscale model (version 3.7.1). These simulations were based on the NCEP GFS. Two nested domains are included in the model configuration, as depicted in Figure 5. Unevenly spaced vertical levels are used with 27 km and 9 km of grid spacing for the coarser grid of domain 1 and for domains 2 and 45, respectively. In the evaluation procedure, we used estimates corresponding to domain 2 grid points that enclose the experimental radiometric stations. Two-day with one-hour resolution forecasting simulations were performed on a daily basis. The starting point was ate midnight UTC. The two-way nesting option between domains 1 and 2 was selected to allow the grids to interact in both directions.

In this study, the used a scheme known as Grell convective scheme, which represents an advanced version of the Grell-Devenyi ensemble convection scheme [55]. The rapid radiative transfer scheme (RRTM) is selected to control parameterization for long-wave radiation [56]. The RRTM scheme represents the influences of the detailed absorption spectrum, accounting for carbon dioxide, ozone, and water vapor as well as a scheme for short-wave radiation [57] and PBL scheme [58] of Yonsei University (YSU).

Specific days of the year, with distinct sky conditions, are selected to analyze the performance of the WRF model. The main objective of such selection is to evaluate the model's forecasting accuracy under different meteorological conditions. Therefore, the condition of the sky is the main basis of the analysis. In particular, we considered four different daily scenarios: clear sky, cloudy, dusty, and dusty-cloudy. From an operational perspective, it is more practical to forecast on a day-ahead mean basis than an hour-ahead basis.

4. Results and Discussion

In this section, the discussion of the obtained results is presented in the context of the main hypothetical questions raised in this study. Particularly, the first question of which prediction paradigm performed the best: numerical, machine learning, or hybrid.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

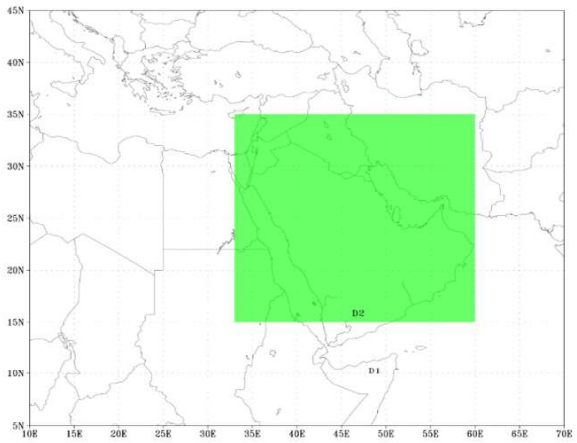

Figure 5. Successive nested domains for model configuration.

The second question is related to investigating the impact of feature space dimension-reduction on the prediction performance. The later aspect has been considered partly in an earlier study [59]. However, we extend this study here by considering the hybrid model for regression of real-time non-discretized data.

4.1. Feature Selection Results

One of the main aims of this research is to optimize the input feature space. Before discussing the proposed models, we emphasize the impact of the proposed feature-space reduction on efficiency. Specifically, a discussion is raised regarding the potential effects of this reduction based on the γ-values. We consider the following three different forms of the output space:

- a continuous real-number space,
- a 5-class decision space, and
- a 10-class decision space

Figures 6–8 illustrate the precision independence of the input attributes when fed individually to the classifier — that is a single-reduct input space. Every one of these three figures displays the γ-values separately for each of the three output spaces listed above, for each single attribute. From these figures, it is clear that the *DH* and *DN* attributes yield the best γ-values, then *H*. *WS* and *PWS* have the worst γ-values.

Figures 9–11 illustrate how dual-attribute reductions function in relation to the γ-values. The left diagonals in these figures reflect the top view of Figures 6–8. The non-linearity of the input data is proved here. This implies that the best quality is not always achieved when effective cuts are taken individually, and vice versa. This reflects the complex nature of the problem under consideration herein.

More comprehensive results for real, 5-class, and 10-class data are presented in Tables 4 and 5. The 5-class and 10-class results are combined in Table 5, as both results are similar. These results indicate how a combination of a very low-γ single-reduct attribute and other attributes may provide better prediction quality than a combination of good single reducts. In the KAU case, for example, combining *H* and *DH* with *PWS* boosted the value of γ to 100%. A similar effect appears when the low *P* attribute is combined with other attributes.

4.2. Prediction Results with Classification

For prediction and energy level classification, we follow a five-fold cross-validation evaluation scheme. The results with reducts ignore the attributes that have not been

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

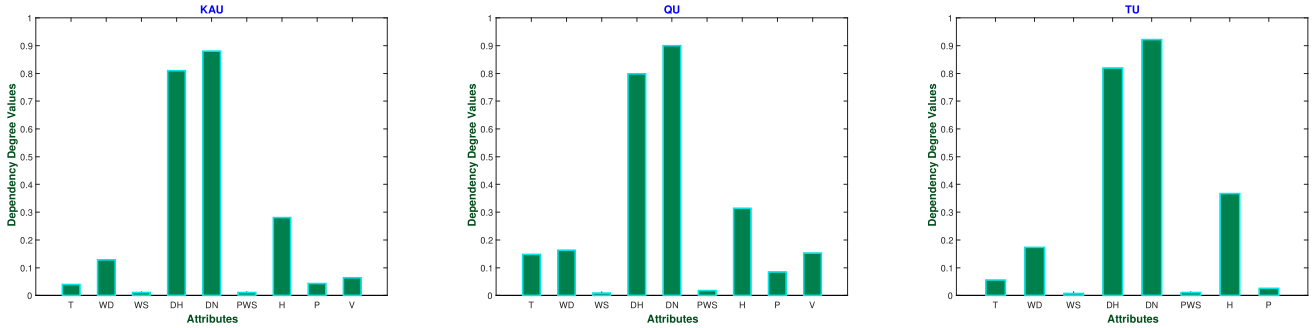

Figure 6. The real values of the decision attribute: γ-values of reducts with a single attribute.

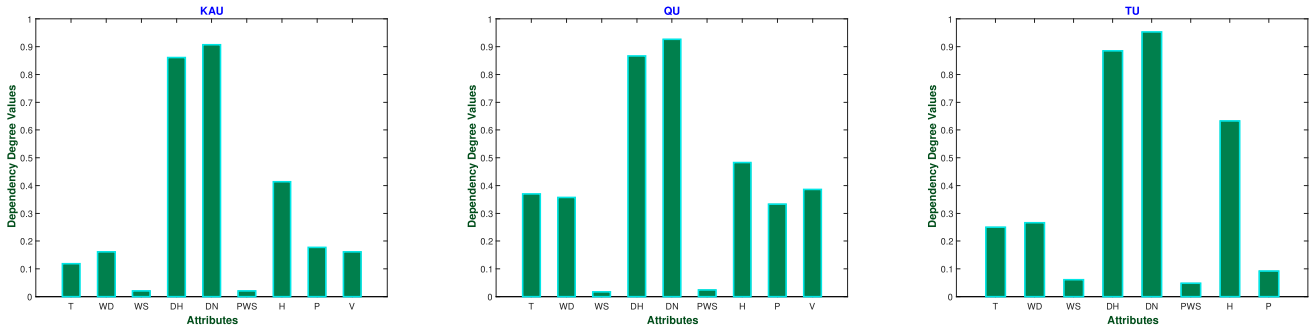

Figure 7. The 5-class decision attribute: γ-values of reducts with a single attribute.

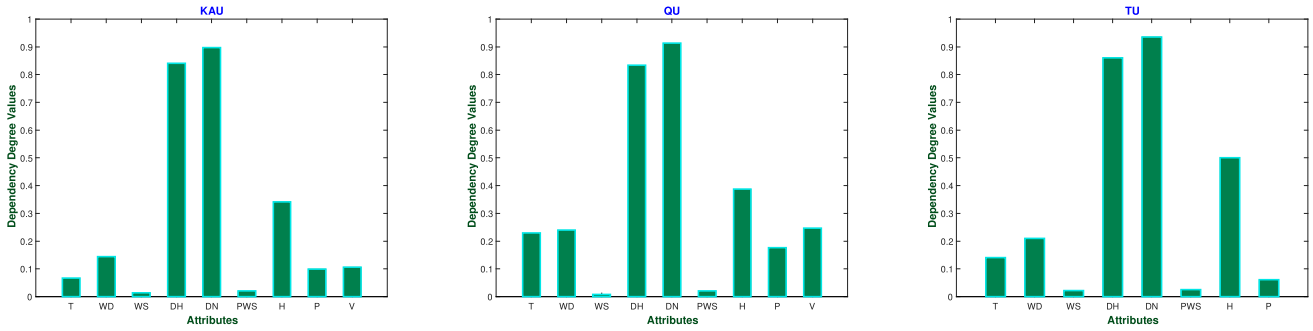

Figure 8. The 10-class decision attribute: γ-values of reducts with a single attribute.

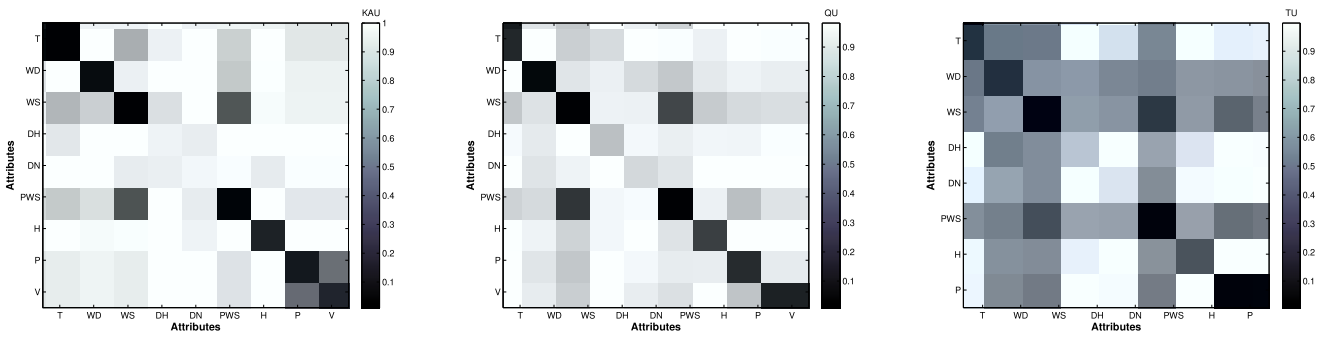

Figure 9. The real values of the decision attribute: Distributions of γ-values of reducts with dual and single attributes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

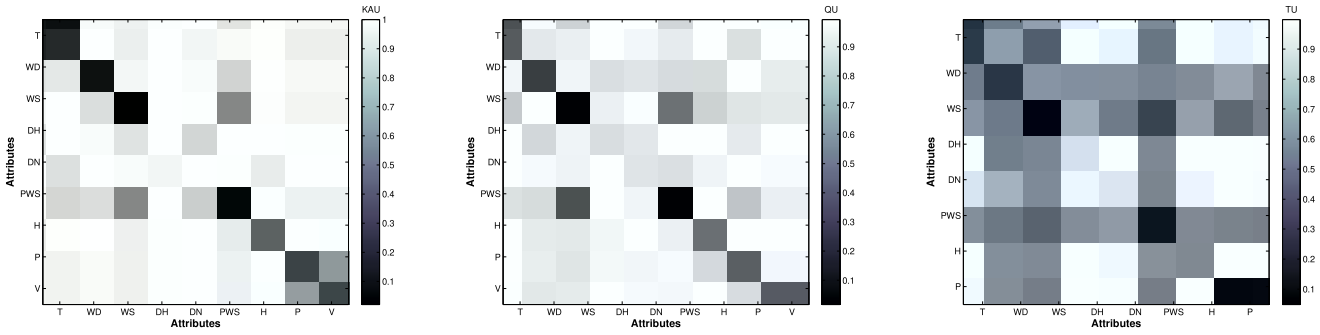

Figure 10. The 5-class decision attribute: Distributions of γ-values of reducts with dual and single attributes.

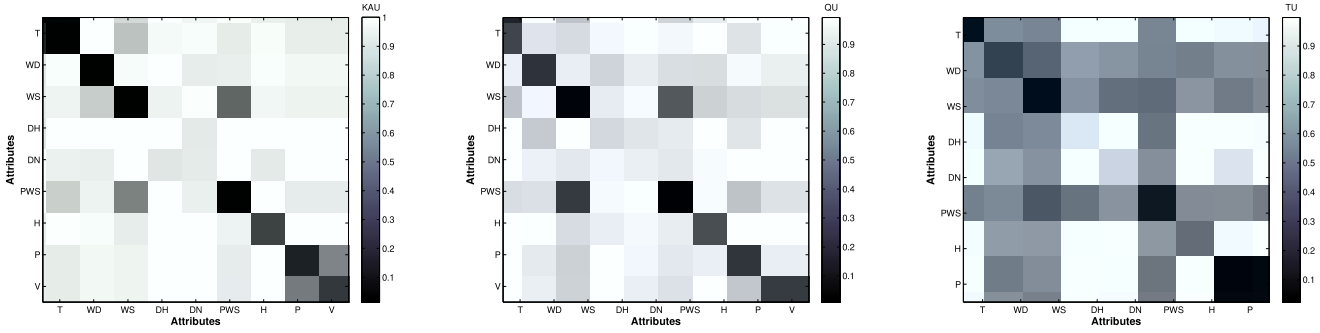

Figure 11. The 10-class decision attribute: Distributions of γ-values of reducts with dual and single attributes.

Table 4: The real values of the decision attribute: The best reducts for five independent runs.

Dataset	Attributes in the best reducts									Reduct Size	Reduct Quality
	T	WD	WS	DH	DN	PWS	H	P	V		
KAU				✓			✓			2	100%
QU							✓	✓	✓	3	99.65%
	✓						✓	✓		3	99.31%
					✓				✓	2	97.92%
				✓			✓			2	97.92%
TU								✓	–	1	99.83%
				✓	✓				–	2	99.83%
					✓				–	1	92.17%

Table 5: The 5-class or the 10-class decision attribute: The best reducts for five independent runs.

Dataset	Attributes in the best reducts									Reduct Size	Reduct Quality
	T	WD	WS	DH	DN	PWS	H	P	V		
KAU	✓			✓						2	100%
				✓		✓	✓			3	100%
QU							✓	✓	✓	3	99.65%
							✓	✓		2	99.31%
	✓				✓		✓		✓	4	97.92%
	✓	✓					✓		✓	4	97.92%
TU								✓	–	1	99.83%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

selected in real values of the decision attribute, as shown in Table 4 — for example attributes WD, WS, and PWS.

Figures 12–17 depict the class prediction results using the 35 prediction models that were discussed in Section 2.2. Generally, using reducts for reduced feature selection yields better class predication rates for most cases. The remarkable performance boost occurs with the cases that are originally affected by missing measurements. e.g., the discriminant analysis for TU datasets. Tables 6–8 illustrate the same result sets in a numerical form for a clearer comparison. It can be seen that the levels of classification vary according to the model used, with advantage to the results to the support vector machines and neural networks models.

4.3. Prediction Results with Regression

Two solar forecasting experiments were carried out using regression models. In the first experiment, the GPR models are applied to the KAU, QU and TU datasets. The main results of this experiment are presented in Figures 21-26 and Table 9. The predicted solar irradiance versus the real values for the invoked datasets with and without reduction are shown in Figures 21-26. These figures indicate how promise the proposed regression models are, especially the ones with reduction. Table 9 shows this by comparing the error values between the two models using the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-Squared measures. Given the values of the prediction and observation data, y_i and ζ_i, $i = 1, \ldots, n$, respectively, these comparitive measures are computed as follows:

$$\begin{aligned} e_i &= (y_i - \zeta_i), \\ \text{RMSE} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}, \\ \text{MAE} &= \frac{1}{n}\sum_{i=1}^{n} |e_i|, \\ \bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i, \\ \text{R-Squared} &= 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \end{aligned} \tag{1}$$

The comparison in Table 9 highlights the success of the reduction models in obtaining better results in two cases, and the results were very close in the third one.

In the other regression experiment, the machine learning (GPR), numerical (WRF), and hybrid (GPR with WRF data) prediction models are applied on datasets KAU, QU, and TU with dust storms. The daily solar features including the GHI values of the current and previous days are used to predict the value of the GHI on the next day. Root mean square errors are computed and reported using the five-fold cross-validation criterion, as depicted in Table 10. The conclusion obtained from this result set implies that the use of machine learning generally improves the results. The proposed hybrid model gives the best results in most cases. Even with the only exception, with TU dataset, the recorded error using the hybrid model is slightly larger than GPR. The second conclusion from these results is that feature reduction does not help a lot with the regression process. Figures 18– 20 reveal detailed class prediction results. In most cases, the classification failure occurs with neighboring classes.

Figures 27– 29 comprehensively reveal the detailed deviations between the true and the predicted GHI values. The hybrid model performs well in most cases. There are some days where the error between true and predicted values is large. This is because these days usually follows cloudy ones whose mostly inaccurate measurements.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

Table 6: Class prediction rates of the KAU datasets.

No.	Classifiers	5 Classes without reducts	5 Classes with reducts	10 Classes without reducts	10 Classes with reducts
1	Fine Decision Tree	70.6%	72.7%	57.2%	57.9%
2	Medium Decision Tree	68.0%	68.2%	52.2%	50.2%
3	Coarse Decision Tree	61.7%	61.7%	40.5%	40.5%
4	Optimizable Decision Tree	73.4%	75.4%	57.9%	59.8%
5	Linear Discriminate Analysis	87.6%	88.1%	76.6%	76.8%
6	Quadratic Discriminate Analysis	74.6%	74.7%	–	–
7	Optimizable Discriminate Analysis	87.6%	88.1%	76.6%	76.8%
8	Gaussian Naïve Bayes	60.1%	58.1%	40.5%	39.2%
9	Kernal Naïve Bayes	69.4%	68.7%	50.5%	48.6%
10	Optimizable Naïve Bayes	70.1%	68.7%	50.7%	48.6%
11	Linear SVM	87.5%	88.8%	72.3%	75.1%
12	Quadratic SVM	86.9%	**89.7%**	73.0%	78.4%
13	Cubic SVM	83.2%	88.7%	68.6%	73.5%
14	Fine Gaussian SVM	68.9%	78.7%	45.5%	60.1%
15	Medium Gaussian SVM	80.4%	83.3%	60.7%	65.8%
16	Coarse Gaussian SVM	73.4%	74.9%	50.3%	51.4%
17	Optimizable SVM	**88.1%**	89.3%	**78.9%**	**81.6%**
18	Fine KNN	74.2%	78.9%	57.6%	62.9%
19	Medium KNN	71.8%	79.2%	53.6%	59.5%
20	Coarse KNN	62.8%	67.2%	43.1%	44.8%
21	Cosine KNN	71.5%	75.3%	54.0%	59.2%
22	Cubic KNN	76.3%	78.2%	54.3%	58.2%
23	Weighted KNN	76.5%	81.3%	59.1%	64.6%
24	Optimizable KNN Classifiers	85.6%	86.3%	74.1%	74.9%
25	Ensemble Boosted Decision Trees	76.5%	75.9%	55.0%	56.2%
26	Ensemble Bagged Decision Trees	70.4%	72.3%	62.2%	62.2%
27	Ensemble Subspace Discriminate Analysis	82.0%	82.6%	71.8%	68.7%
28	Ensemble Subspace KNN	74.7%	67.4%	61.9%	49.8%
29	Ensemble RUS Boosted Decision Trees	72.9%	73.4%	50.3%	51.4%
30	Optimizable Ensemble Classifiers	76.3%	79.4%	63.4%	62.4%
31	Narrow Neural Networks	85.6%	86.6%	71.8%	75.6%
32	Medium Neural Networks	87.6%	87.1%	72.3%	68.7%
33	Wide Neural Networks	87.8%	88.7%	75.1%	70.3%
34	Bilayered Neural Networks	84.7%	87.1%	71.6%	77.0%
35	Trilayered Neural Networks	85.4%	85.6%	73.5%	77.7%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

Table 7: Class prediction rates of the QU datasets.

No.	Classifiers	5 Classes		10 Classes	
		without reducts	with reducts	without reducts	with reducts
1	Fine Decision Tree	72.3%	78.6%	55.4%	57.8%
2	Medium Decision Tree	75.3%	76.8%	51.6%	53.7%
3	Coarse Decision Tree	66.2%	65.5%	46.0%	46.0%
4	Optimizable Decision Tree	77.7%	81.5%	60.3%	62.2%
5	Linear Discriminate Analysis	85.8%	93.0%	–	84.0%
6	Quadratic Discriminate Analysis	–	–	–	–
7	Optimizable Discriminate Analysis	85.8%	93.0%	–	84.0%
8	Gaussian Naïve Bayes	65.2%	64.8%	–	–
9	Kernel Naïve Bayes	72.0%	69.9%	–	51.2%
10	Optimizable Naïve Bayes	72.0%	70.6%	–	52.1%
11	Linear SVM	83.1%	92.9%	69.7%	82.9%
12	Quadratic SVM	83.6%	92.0%	70.6%	81.2%
13	Cubic SVM	80.3%	90.8%	67.4%	79.3%
14	Fine Gaussian SVM	68.8%	80.7%	44.4%	65.3%
15	Medium Gaussian SVM	78.0%	87.1%	62.4%	72.8%
16	Coarse Gaussian SVM	72.0%	77.0%	50.0%	52.4%
17	Optimizable SVM	**86.2%**	**93.9%**	**77.9%**	**86.4%**
18	Fine KNN	74.6%	85.7%	57.1%	70.7%
19	Medium KNN	74.6%	81.9%	51.9%	63.8%
20	Coarse KNN	66.9%	70.0%	45.8%	48.4%
21	Cosine KNN	73.0%	81.7%	50.7%	63.2%
22	Cubic KNN	74.2%	82.8%	51.2%	66.4%
23	Weighted KNN	76.3%	85.0%	55.9%	70.9%
24	Optimizable KNN Classifiers	86.1%	91.1%	74.2%	81.2%
25	Ensemble Boosted Decision Trees	81.2%	82.0%	58.4%	60.5%
26	Ensemble Bagged Decision Trees	79.4%	82.2%	64.1%	65.3%
27	Ensemble Subspace Discriminate Analysis	78.6%	79.1%	67.9%	66.0%
28	Ensemble Subspace KNN	79.3%	73.3%	69.5%	51.7%
29	Ensemble RUS Boosted Decision Trees	71.3%	70.9%	39.7%	42.2%
30	Optimizable Ensemble Classifiers	84.8%	84.1%	71.3%	70.6%
31	Narrow Neural Networks	85.9%	92.2%	70.9%	82.1%
32	Medium Neural Networks	86.1%	91.6%	70.6%	80.5%
33	Wide Neural Networks	84.7%	91.6%	71.4%	81.4%
34	Bilayered Neural Networks	85.4%	90.9%	74.4%	81.5%
35	Trilayered Neural Networks	84.7%	93.4%	73.5%	82.1%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

Table 8: Class prediction rates of the TU datasets.

No.	Classifiers	5 Classes		10 Classes	
		without reducts	with reducts	without reducts	with reducts
1	Fine Decision Tree	68.3%	74.2%	54.4%	57.7%
2	Medium Decision Tree	69.0%	73.5%	55.4%	56.4%
3	Coarse Decision Tree	67.8%	66.7%	44.6%	44.6%
4	Optimizable Decision Tree	74.9%	75.3%	59.1%	59.1%
5	Linear Discriminate Analysis	–	86.4%	–	–
6	Quadratic Discriminate Analysis	–	–	–	–
7	Optimizable Discriminate Analysis	–	86.4%	–	–
8	Gaussian Naïve Bayes	–	–	–	–
9	Kernel Naïve Bayes	–	69.2%	–	48.8%
10	Optimizable Naïve Bayes	–	69.2%	–	48.8%
11	Linear SVM	48.6%	87.2%	39.5%	73.7%
12	Quadratic SVM	49.0%	89.9%	42.0%	**76.5%**
13	Cubic SVM	48.4%	85.4%	39.5%	73.7%
14	Fine Gaussian SVM	38.3%	79.6%	7.0%	48.1%
15	Medium Gaussian SVM	46.9%	85.5%	38.3%	66.4%
16	Coarse Gaussian SVM	43.7%	79.4%	32.9%	53.8%
17	Optimizable SVM	51.6%	**91.6%**	39.4%	75.1%
18	Fine KNN	58.0%	80.5%	36.8%	59.1%
19	Medium KNN	57.7%	80.0%	37.1%	58.0%
20	Coarse KNN	53.3%	71.4%	26.3%	46.3%
21	Cosine KNN	55.4%	78.0%	33.8%	55.4%
22	Cubic KNN	57.1%	78.7%	37.1%	57.1%
23	Weighted KNN	58.7%	81.4%	38.3%	61.7%
24	Optimizable KNN Classifiers	62.9%	88.0%	41.8%	69.5%
25	Ensemble Boosted Decision Trees	76.7%	77.0%	59.8%	60.1%
26	Ensemble Bagged Decision Trees	74.0%	76.7%	55.1%	58.7%
27	Ensemble Subspace Discriminate Analysis	77.0%	79.4%	51.0%	61.5%
28	Ensemble Subspace KNN	70.4%	77.0%	46.3%	51.0%
29	Ensemble RUS Boosted Decision Trees	55.7%	57.7%	19.3%	30.8%
30	Optimizable Ensemble Classifiers	**84.5%**	79.4%	**66.9%**	61.3%
31	Narrow Neural Networks	62.2%	85.4%	50.2%	75.1%
32	Medium Neural Networks	65.7%	85.2%	48.3%	70.6%
33	Wide Neural Networks	66.0%	85.2%	48.1%	70.6%
34	Bilayered Neural Networks	63.8%	88.2%	52.8%	74.4%
35	Trilayered Neural Networks	60.5%	85.7%	51.9%	76.1%

Table 9: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and R-Squared of GHI Values Using Regression Prediction Models on KAU, QU and TU Datasets

	KAU		QU		TU	
	without reducts	with reducts	without reducts	with reducts	without reducts	with reducts
RMSE	124.76	**118.11**	146.79	**143.78**	**170.34**	171.80
MAE	93.067	**85.913**	105.110	**102.530**	**126.200**	129.340
R-Squared	0.99	0.99	0.99	0.99	0.98	0.98

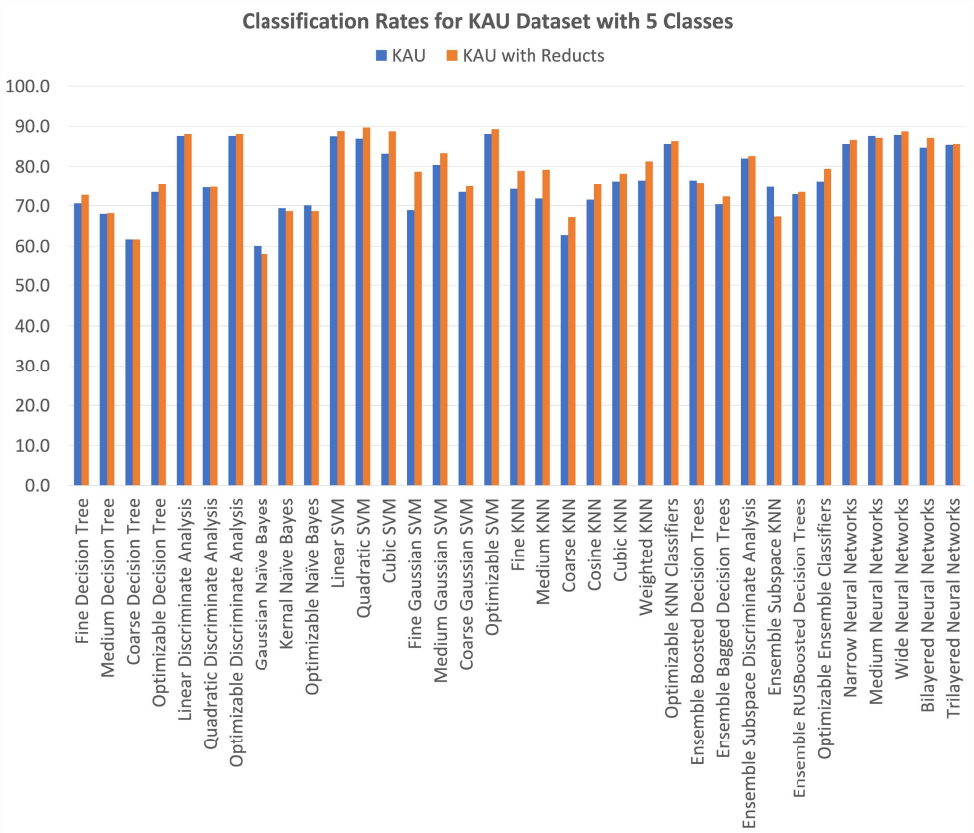

Figure 12. Class prediction rates for the KAU dataset with five classes.

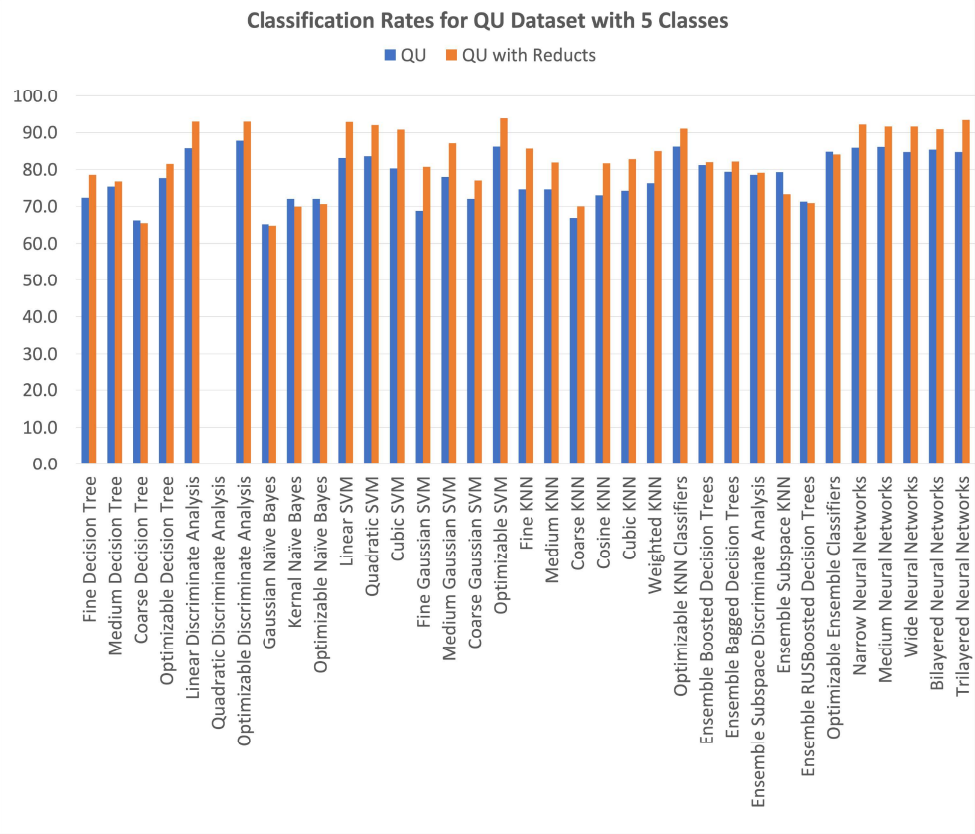

Figure 13. Class prediction rates for the QU dataset with five classes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

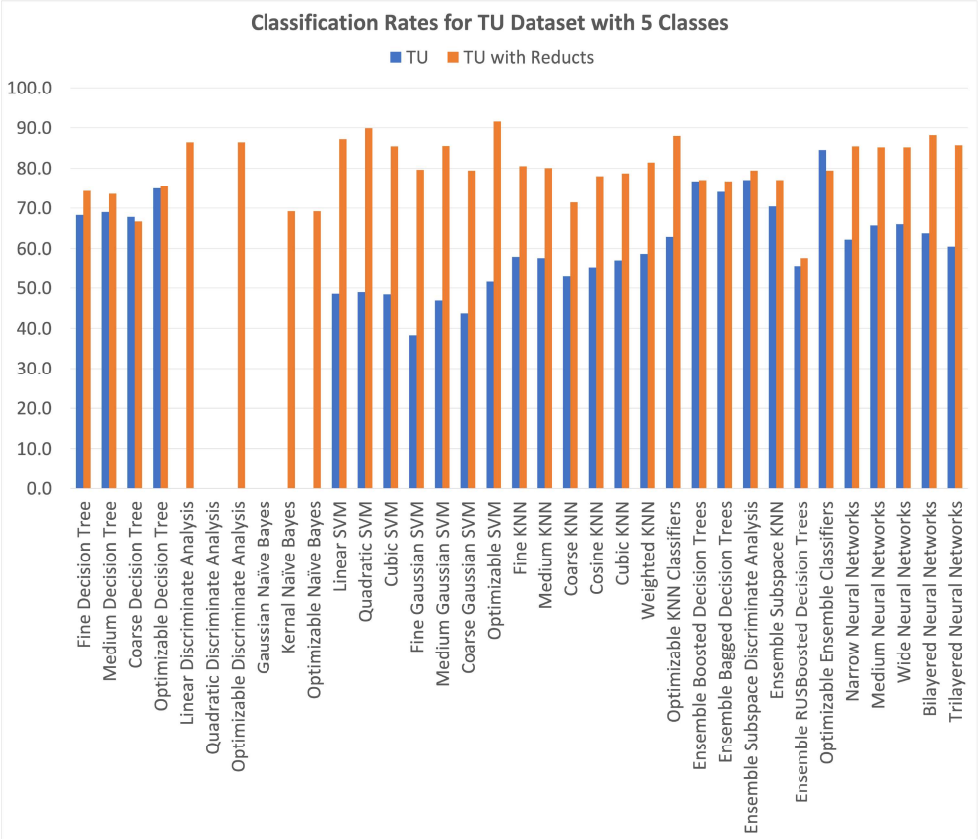

Figure 14. Class prediction rates for the TU dataset with five classes.

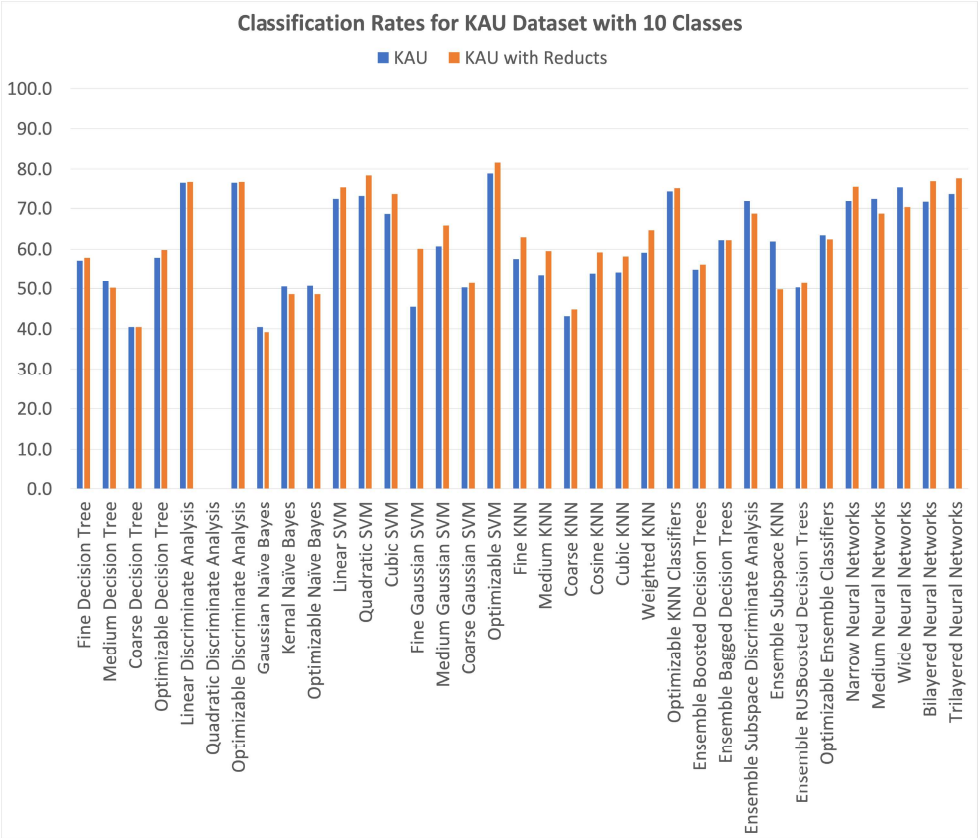

Figure 15. Class prediction rates for the KAU dataset with ten classes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

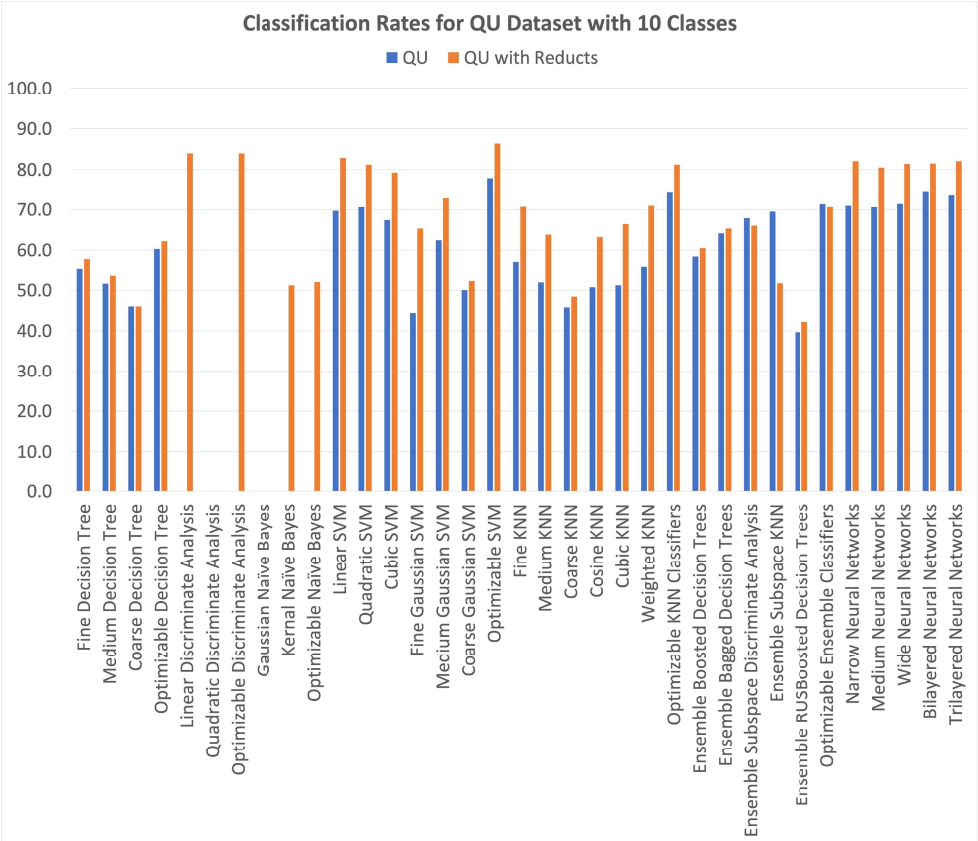

Figure 16. Class prediction rates for the QU dataset with ten classes.

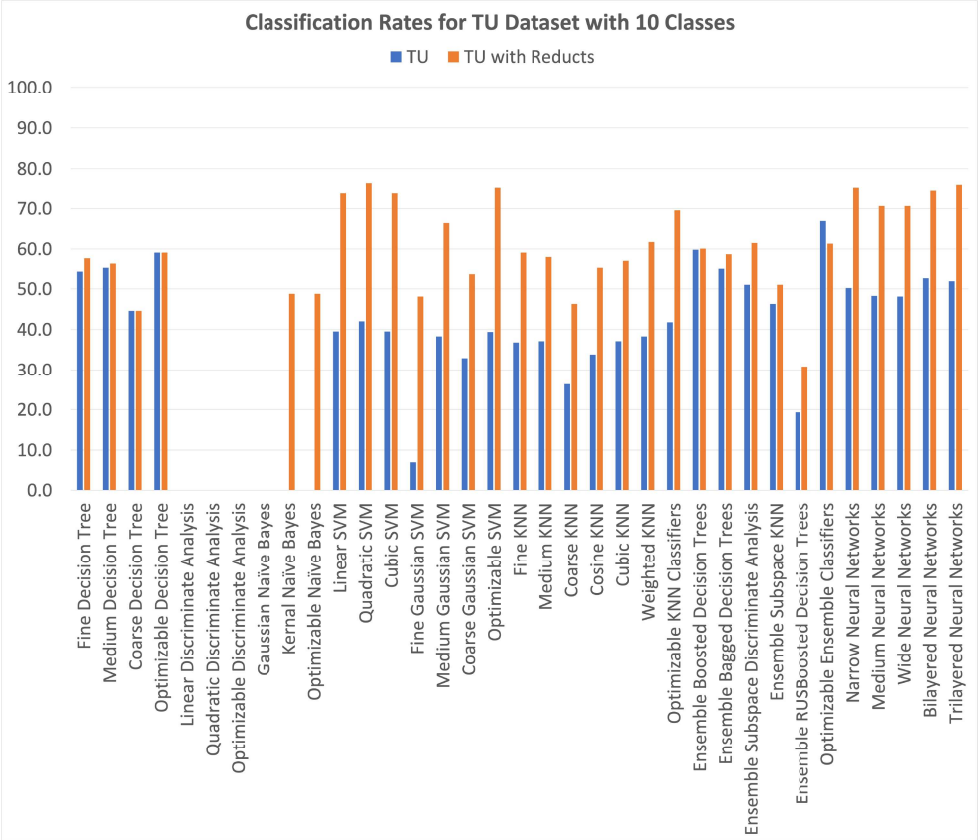

Figure 17. Class prediction rates for the TU dataset with ten classes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

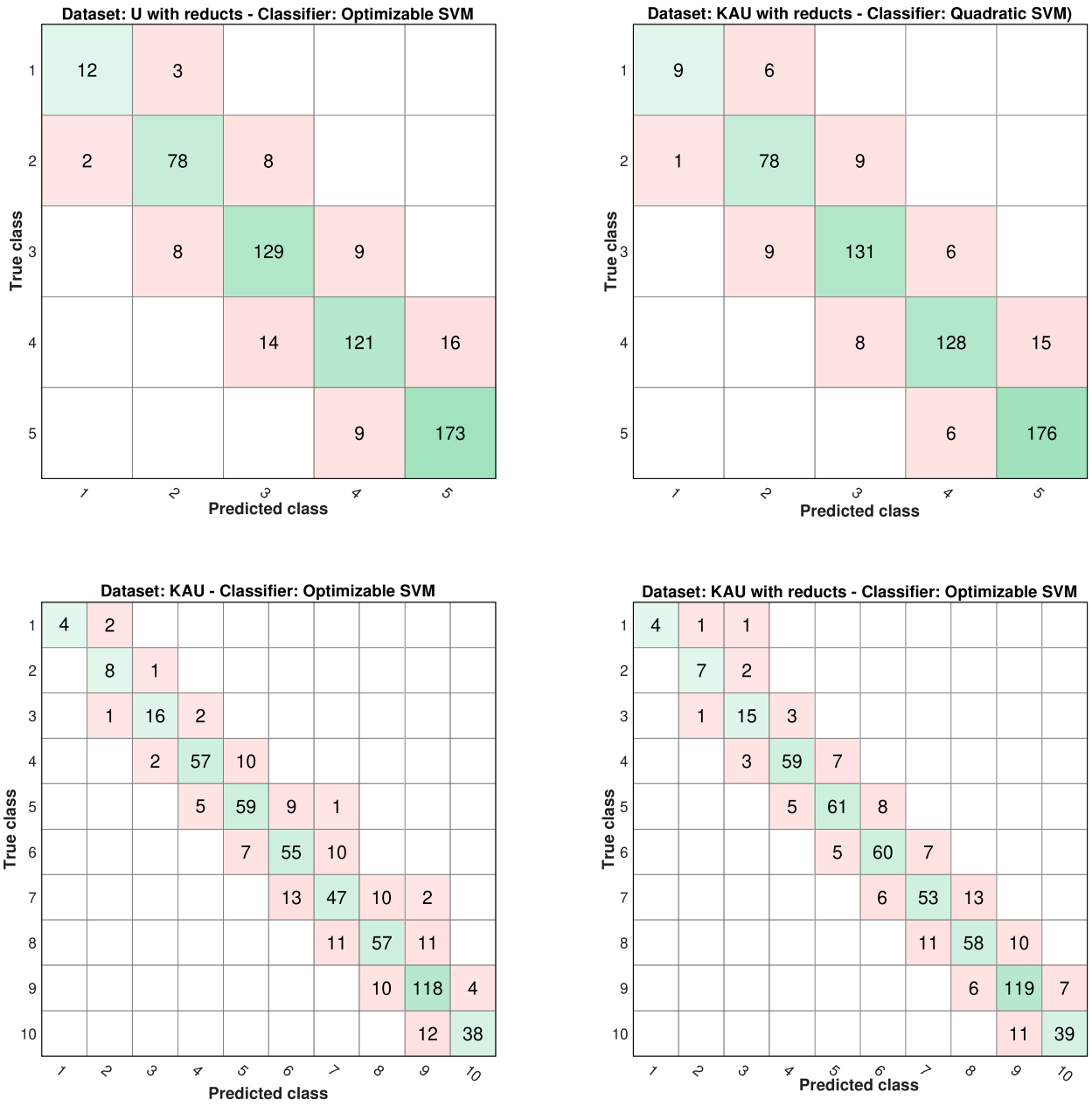

Figure 18. Distributions of KAU predicted classes.

Table 10: RMSE of GHI values using different regression prediction models on dust storm datasets.

Prediction Models	KAU	QU	TU
Numerical Model (WRF)	1412.06	1196.76	1354.16
Machine Learning Model (GPR)	440.02	737.70	**334.46**
Hybrid Model	**421.15**	695.41	365.43
Machine Learning Model with Feature Selection	514.19	645.37	666.37
Hybrid Model with Feature Selection	559.03	**631.13**	641.65

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

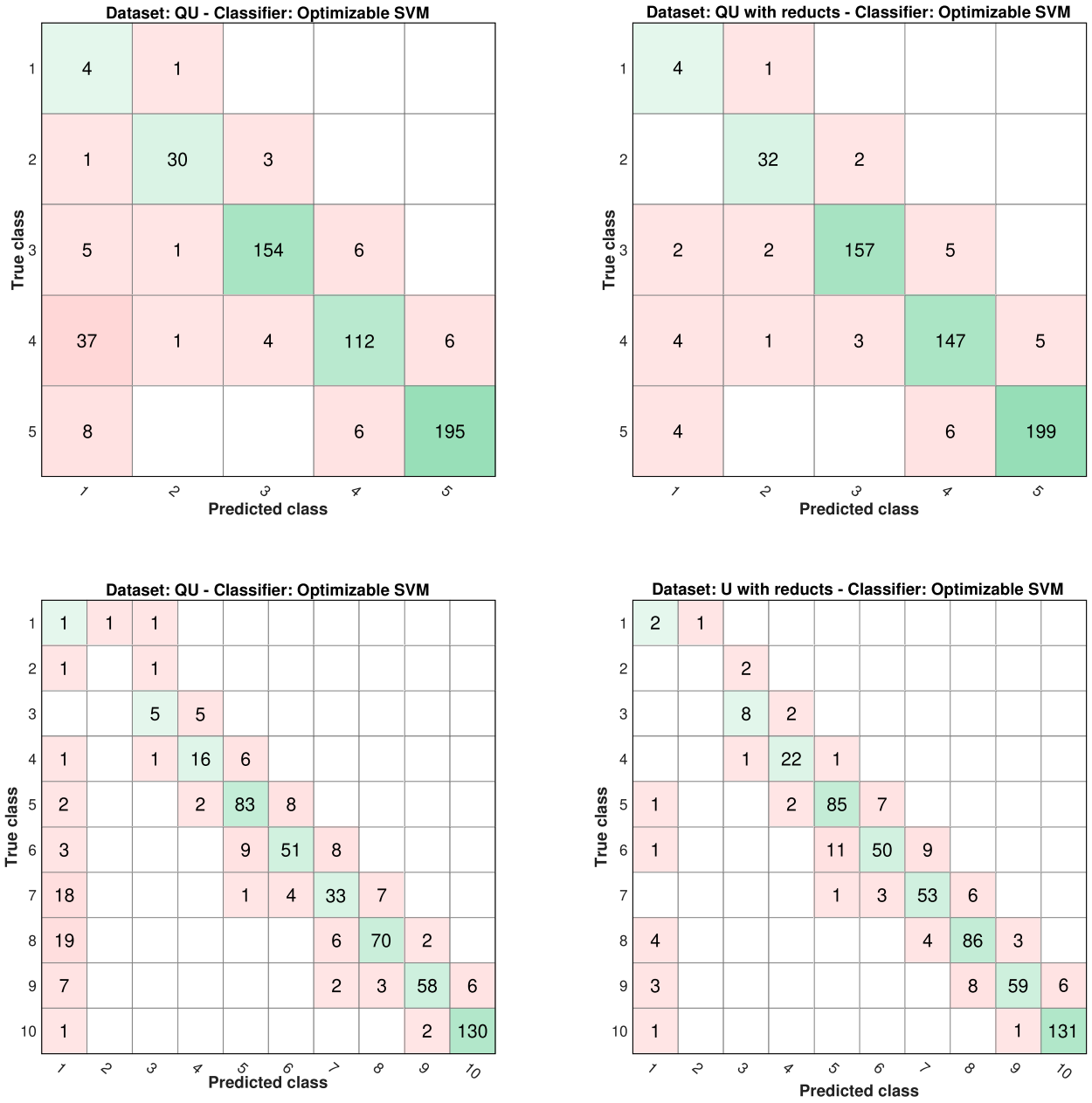

Figure 19. Distributions of QU predicted classes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

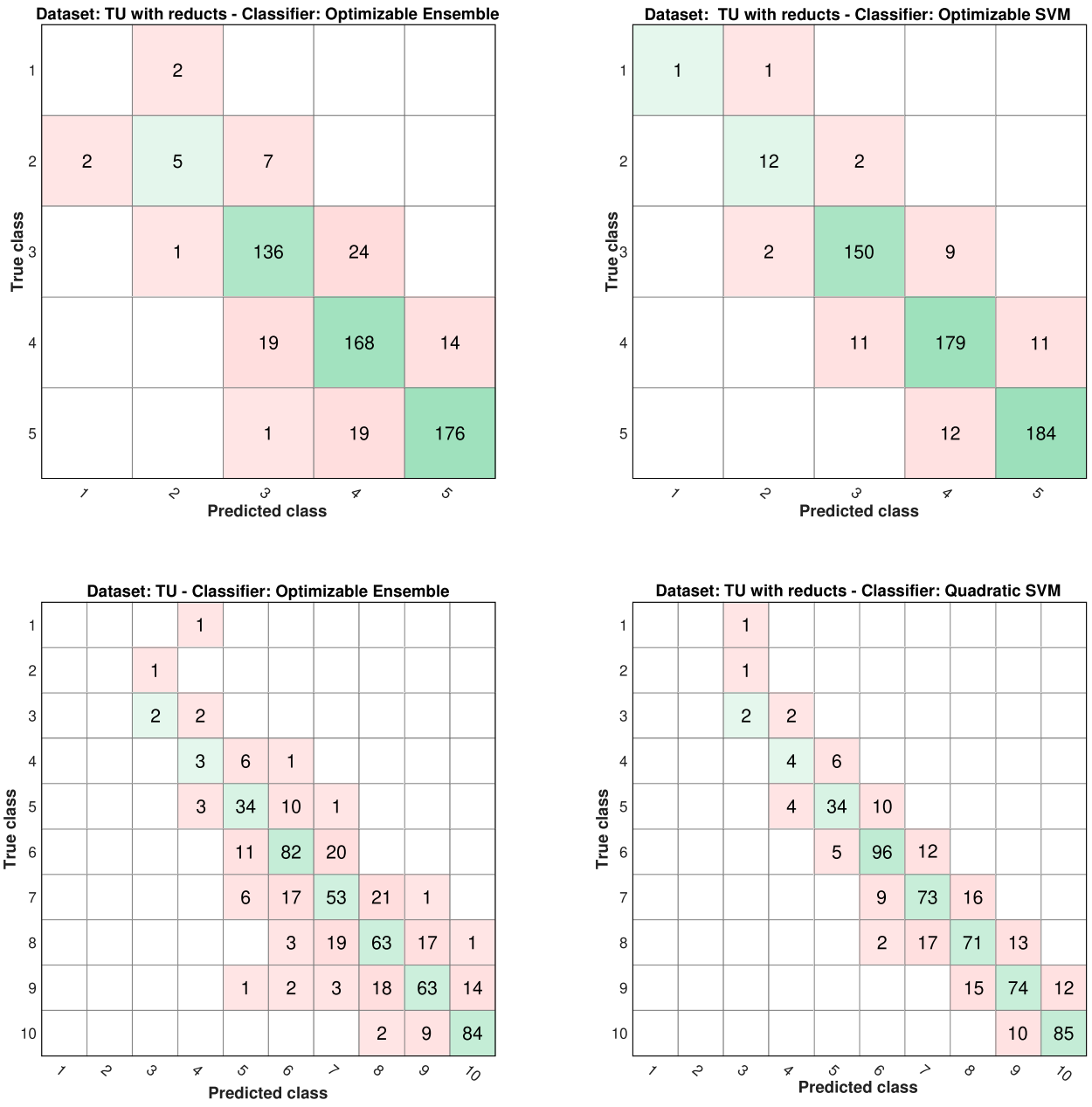

Figure 20. Distributions of TU predicted classes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

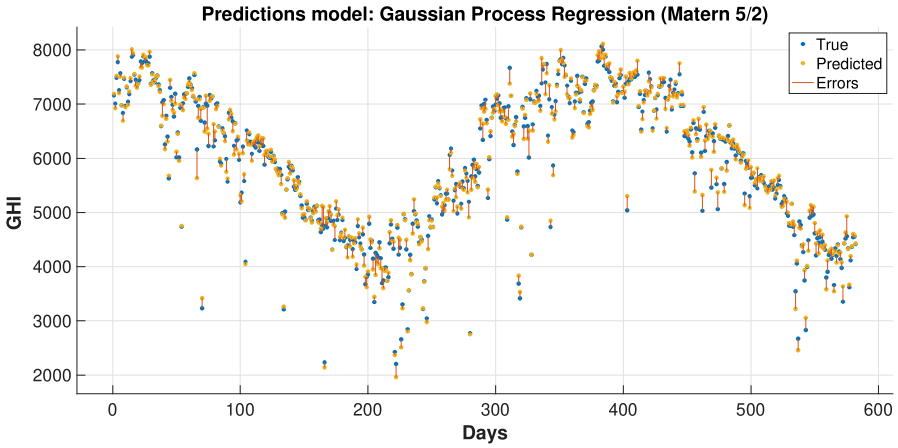

Figure 21. The predicted solar irradiance versus the real values for KAU dataset without reduction.

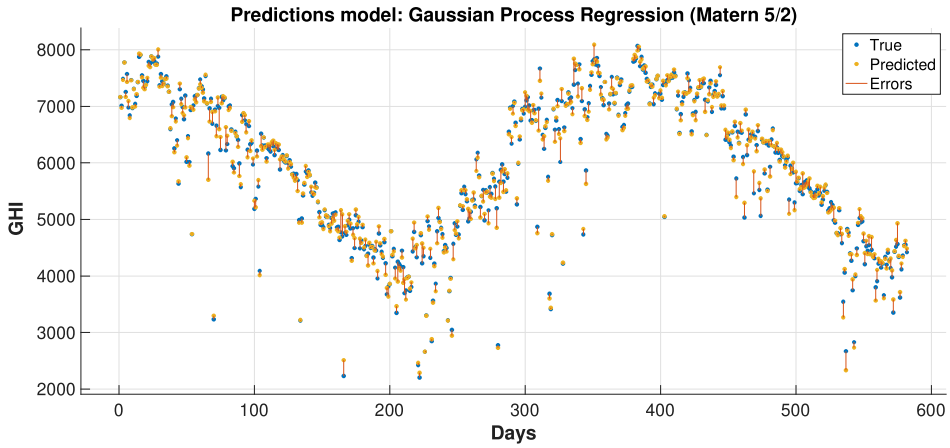

Figure 22. The predicted solar irradiance versus the real values for KAU dataset with reduction.

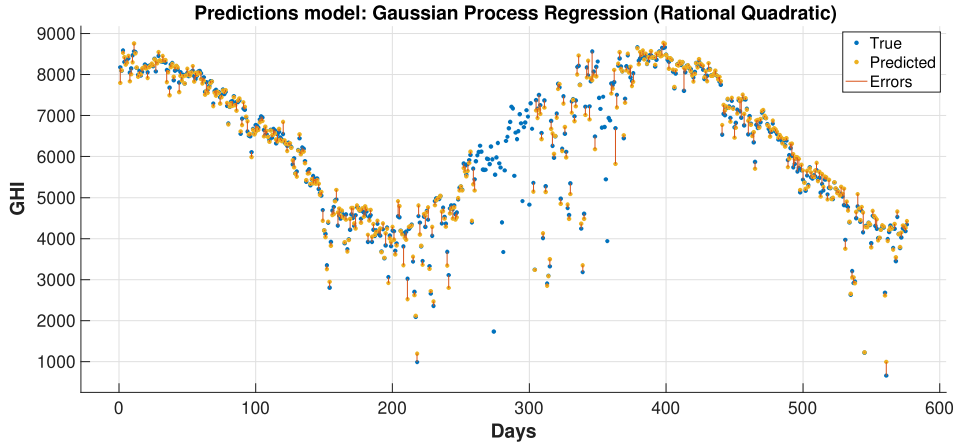

Figure 23. The predicted solar irradiance versus the real values for QU dataset without reduction.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

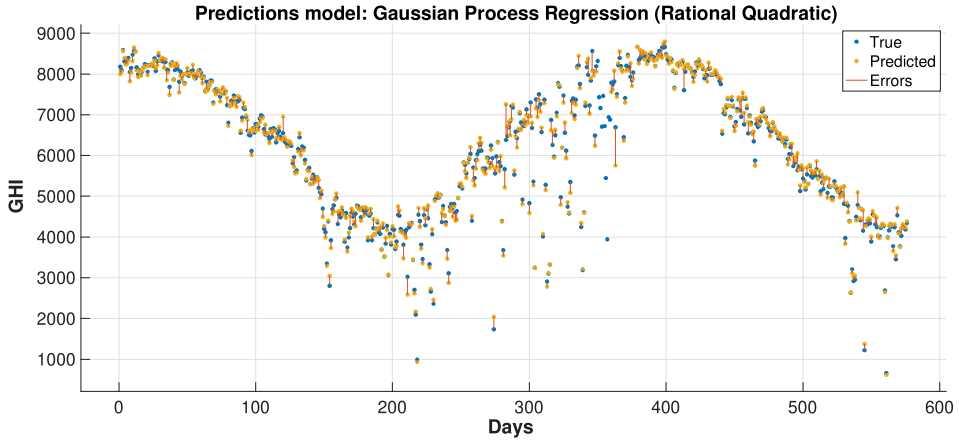

Figure 24. The predicted solar irradiance versus the real values for QU dataset with reduction.

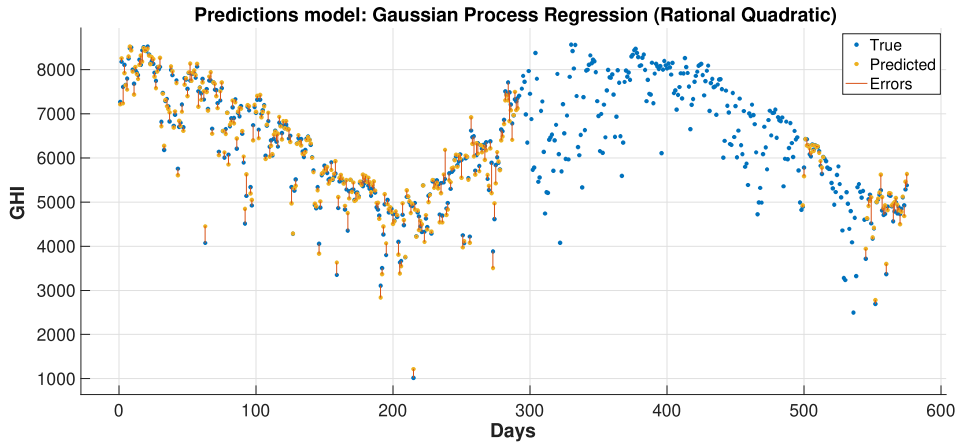

Figure 25. The predicted solar irradiance versus the real values for TU dataset without reduction.

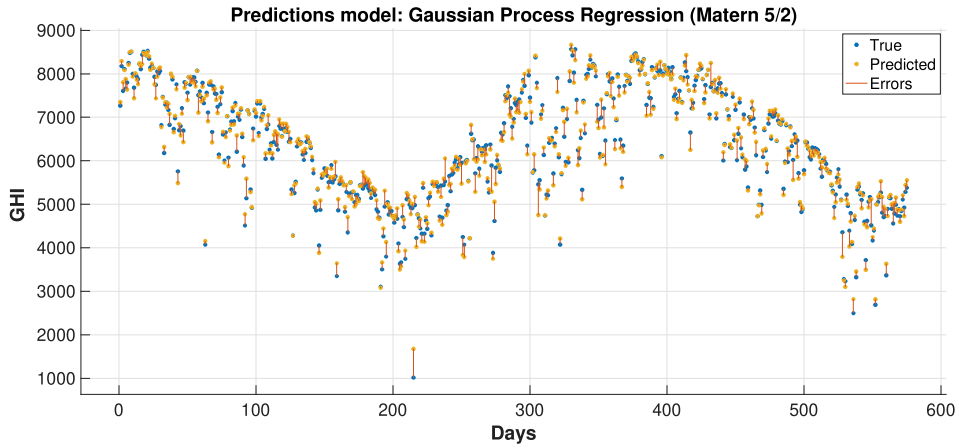

Figure 26. The predicted solar irradiance versus the real values for TU dataset with reduction.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

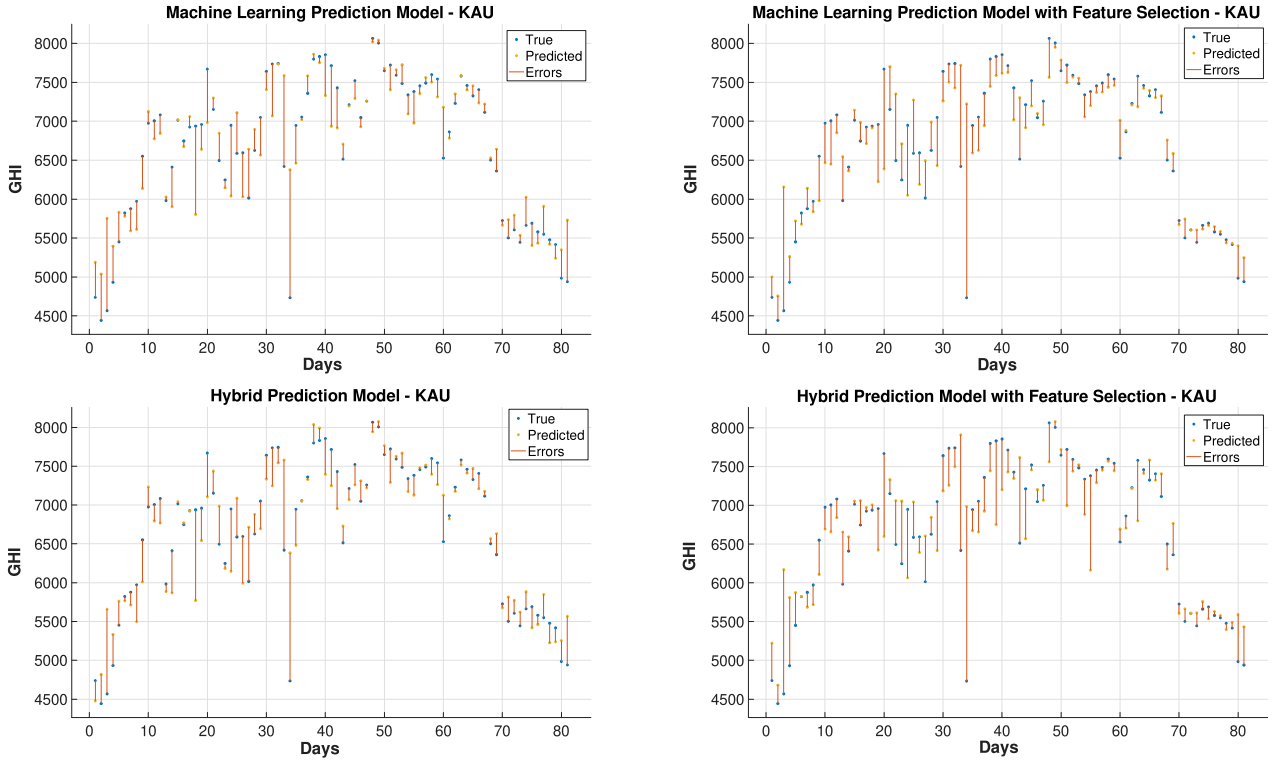

Figure 27. Prediction models for KAU with the dust storm dataset.

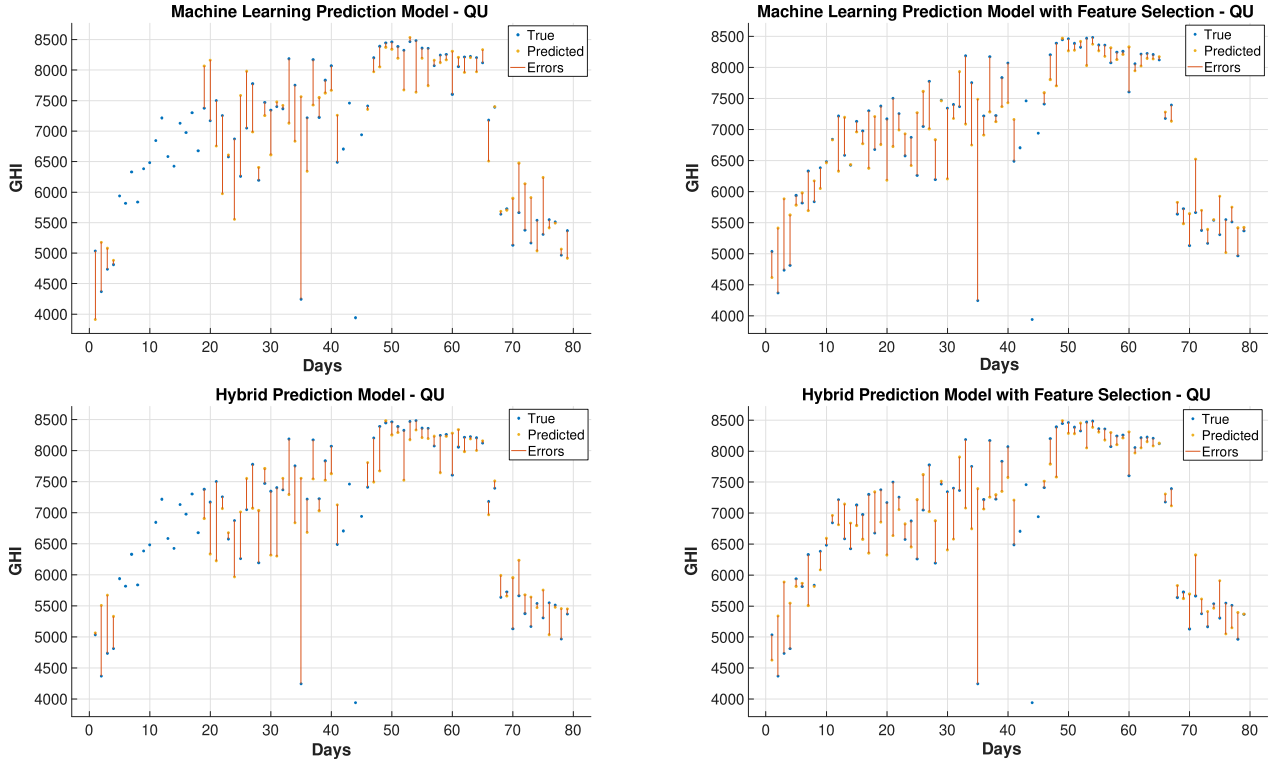

Figure 28. Prediction models for QU with the dust storm dataset.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2021 doi:10.20944/preprints202111.0202.v1

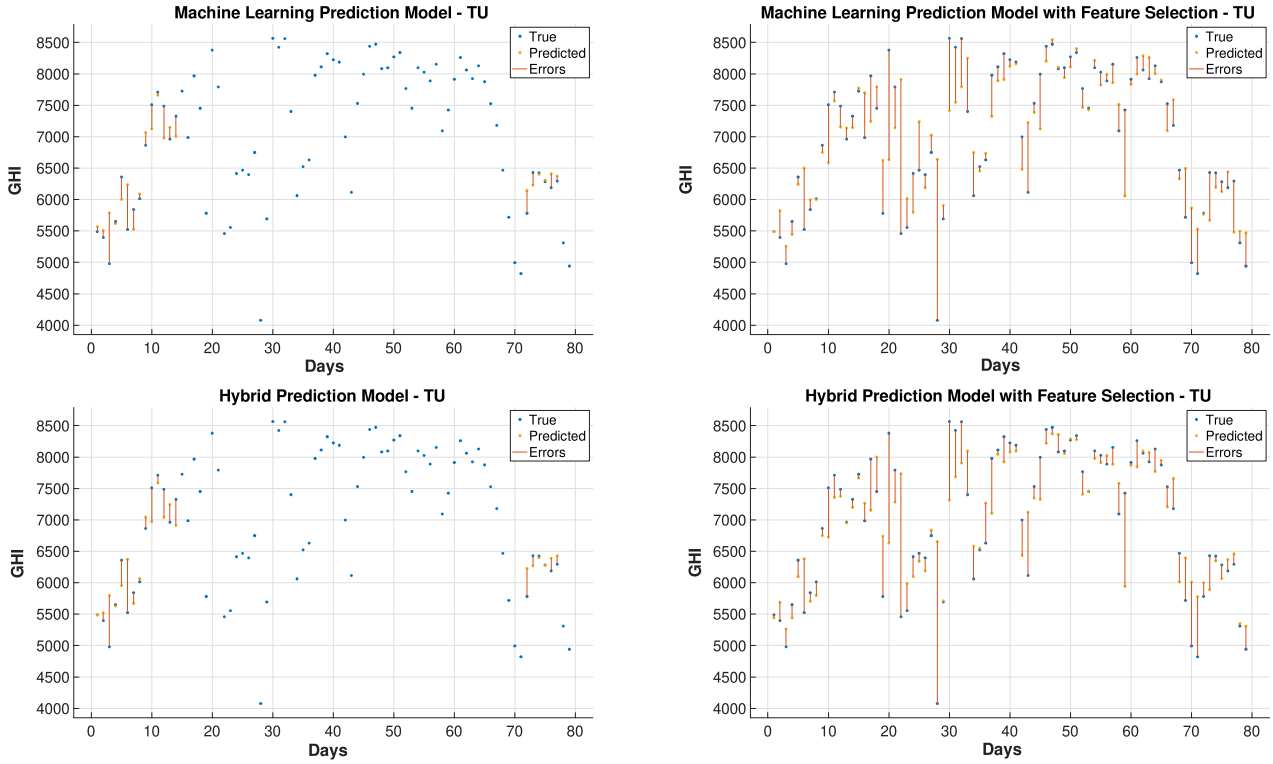

Figure 29. Prediction models for TU with the dust storm dataset.

5. Conclusion

This paper presents hybrid machine learning approaches for solar radiation estimation by utilizing numerical methods. The numerical models, particularly the WRF models, are widely used in forecasting weather data. One of the main achievements of this study is to show the extent to which the use of machine learning models can improve predictions for numerical methods. This has been achieved by building hybrid models through several layers of methodology design. First, a feature selection and dimensionality reduction approach was proposed for parameters associated with solar radiation estimation. The proposed attribute reduction is based on using an adaptive memory programming approach to optimize the input feature space of a solar radiation model. Then, different classification models are used to predict the solar radiation classes. The feature selection has played an important role in increasing the class prediction rates. Finally, the WRF data were used in the proposed regression models to obtain improved prediction results that is generally better than the predictions of pure machine learning and WRF models. The proposed methodologies are evaluated using a real environmental temporal dataset collected from diverse regions in Saudi Arabia. The results proved the effectiveness of the proposed hybrid model in improving the prediction of the GHI values. For discrete classes, attribute reduction, which combines few low-dependency degree single-reduct attributes with other attributes results in very good quality solutions. On the other side, attribute reduction did not contribute much to performance improvement. when discretization is used with the input data classes.

6. Acknowledgments

This work is part of a project funded by King Abdulaziz City for Science and Technology (KACST) with grant number 13-ENES2373-10. In addition, the authors would like to acknowledge King Abdullah City for Atomic and Renewable Energy (KACARE) for supplying data from the stations.

References

1. Cline, W.R. The Economics of Global Warming. *Peterson Institute Press: All Books* **1992**.
2. Riordan, C.; Hulstrom, R.; Cannon, T.; Myers, D. Solar radiation research for photovoltaic applications. *Solar Cells* **1991**, *30*, 489–500.
3. Ibrahim, I.A.; Khatib, T. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. *Energy Conversion and Management* **2017**, *138*, 413–425.
4. Muneer, T.; Younes, S.; Munawwar, S. Discourses on solar radiation modeling. *Renewable and Sustainable Energy Reviews* **2007**, *11*, 551–602.
5. Şen, Z. Solar energy in progress and future research trends. *Progress in Energy and Combustion Science* **2004**, *30*, 367–416.
6. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy* **2017**, *105*, 569–582.
7. Wang, L.; Kisi, O.; Zounemat-Kermani, M.; Salazar, G.A.; Zhu, Z.; Gong, W. Solar radiation prediction using different techniques: Model evaluation and comparison. *Renewable and Sustainable Energy Reviews* **2016**, *61*, 384–397.
8. Abdulrahim, M.; Almaraashi, M. Forecasting of Short-Term Solar Radiation Based on a Numerical Weather Prediction Model over Saudi Arabia. Proceedings of the 6th International Conference on Informatics, Environment, Energy and Applications. ACM, 2017, pp. 16–19.
9. Larson, V.E. Forecasting solar irradiance with numerical weather prediction models. *Solar energy forecasting and resource assessment* **2013**, pp. 299–318.
10. Jimenez, P.A.; Hacker, J.P.; Dudhia, J.; Haupt, S.E.; Ruiz-Arias, J.A.; Gueymard, C.A.; Thompson, G.; Eidhammer, T.; Deng, A. WRF-Solar: Description and clear-sky assessment of an augmented NWP model for solar power prediction. *Bulletin of the American Meteorological Society* **2016**, *97*, 1249–1264.
11. Şen, Z. Fuzzy algorithm for estimation of solar irradiation from sunshine duration. *Solar Energy* **1998**, *63*, 39–49.
12. Bhardwaj, S.; Sharma, V.; Srivastava, S.; Sastry, O.; Bandyopadhyay, B.; Chandel, S.; Gupta, J. Estimation of solar radiation using a combination of Hidden Markov Model and generalized Fuzzy model. *Solar Energy* **2013**, *93*, 43–54. doi:10.1016/j.solener.2013.03.020.
13. Mellit, A.; Arab, A.H.; Khorissi, N.; Salhi, H. An ANFIS-based Forecasting for Solar Radiation Data from Sunshine Duration and Ambient Temperature. *2007 IEEE Power Engineering Society General Meeting* **2007**, pp. 1–6. doi:10.1109/PES.2007.386131.
14. Alobaidi, M.H.; Marpu, P.R.; Ouarda, T.B.M.J.; Ghedira, H. Mapping of the Solar Irradiance in the UAE Using Advanced Artificial Neural Network Ensemble. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2014**, *7*, 3668–3680. doi:10.1109/JSTARS.2014.2331255.
15. Mellit, A.; Pavan, A.M. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Solar Energy* **2010**, *84*, 807–821. doi:10.1016/j.solener.2010.02.006.
16. Khatib, T.; Mohamed, A.; Mahmoud, M.; Sopian, K. Modeling of Daily Solar Energy on a Horizontal Surface for Five Main Sites in Malaysia. *International Journal of Green Energy* **2011**, *8*, 795–819. doi:10.1080/15435075.2011.602156.
17. de Freitas Viscondi, G.; Alves-Souza, S.N. Solar Irradiance Prediction with Machine Learning Algorithms: A Brazilian Case Study on Photovoltaic Electricity Generation. *Energies* **2021**, *14*, 5657.
18. Takamatsu, T.; Ohtake, H.; Oozeki, T.; Nakaegawa, T.; Honda, Y.; Kazumori, M. Regional Solar Irradiance Forecast for Kanto Region by Support Vector Regression Using Forecast of Meso-Ensemble Prediction System. *Energies* **2021**, *14*, 3245.
19. Alzahrani, A.; Shamsi, P.; Dagli, C.; Ferdowsi, M. Solar irradiance forecasting using deep neural networks. *Procedia Computer Science* **2017**, *114*, 304–313.
20. Gbémou, S.; Eynard, J.; Thil, S.; Guillot, E.; Grieu, S. A Comparative Study of Machine Learning-Based Methods for Global Horizontal Irradiance Forecasting. *Energies* **2021**, *14*, 3192.
21. Aslam, M.; Lee, J.M.; Kim, H.S.; Lee, S.J.; Hong, S. Deep learning models for long-term solar radiation forecasting considering microgrid installation: A comparative study. *Energies* **2020**, *13*, 147.

22. Chandola, D.; Gupta, H.; Tikkiwal, V.A.; Bohra, M.K. Multi-step ahead forecasting of global solar radiation for arid zones using deep learning. *Procedia Computer Science* **2020**, *167*, 626–635.
23. Mukhoty, B.P.; Maurya, V.; Shukla, S.K. Sequence to sequence deep learning models for solar irradiation forecasting. 2019 IEEE Milan PowerTech. IEEE, 2019, pp. 1–6.
24. Jayalakshmi, N.Y.; Shankar, R.; Subramaniam, U.; Baranilingesan, I.; Stalin, A.K.B.; Rahim, R.; Ghosh, A. Novel Multi-Time Scale Deep Learning Algorithm for Solar Irradiance Forecasting. *Energies* **2021**, *14*, 2404.
25. de Araujo, J.M.S. Performance comparison of solar radiation forecasting between WRF and LSTM in Gifu, Japan. *Environmental Research Communications* **2020**, *2*, 045002.
26. Husein, M.; Chung, I.Y. Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. *Energies* **2019**, *12*, 1856.
27. Ghimire, S.; Deo, R.C.; Raj, N.; Mi, J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy* **2019**, *253*, 113541.
28. Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* **2018**, *148*, 461–468.
29. Yu, Y.; Cao, J.; Zhu, J. An LSTM short-term solar irradiance forecasting under complicated weather conditions. *IEEE Access* **2019**, *7*, 145651–145666.
30. Abdel-Nasser, M.; Mahmoud, K.; Lehtonen, M. Reliable Solar Irradiance Forecasting Approach Based on Choquet Integral and Deep LSTMs. *IEEE Transactions on Industrial Informatics* **2020**, *17*, 1873–1881.
31. Almaraashi, M. Short-term prediction of solar energy in Saudi Arabia using automated-design fuzzy logic systems. *PloS one* **2017**, *12*.
32. Almaraashi, M. Investigating the impact of feature selection on the prediction of solar radiation in different locations in Saudi Arabia. *Applied Soft Computing* **2018**, *66*, 250–263. doi:https://doi.org/10.1016/j.asoc.2018.02.029.
33. Voyant, C.; Muselli, M.; Paoli, C.; Nivet, M.L. Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. *Energy* **2012**, *39*, 341–355.
34. Boubaker, S.; Benghanem, M.; Mellit, A.; Lefza, A.; Kahouli, O.; Kolsi, L. Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia. *IEEE Access* **2021**, *9*, 36719–36729.
35. Hedar, A.R.; Wang, J.; Fukushima, M. Tabu search for attribute reduction in rough set theory. *Soft Computing* **2008**, *12*, 909–918.
36. Pawlak, Z. *Rough sets: Theoretical aspects of reasoning about data*; Vol. 9, Springer Science & Business Media, 2012.
37. Loh, W.Y. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2011**, *1*, 14–23.
38. Li, T.; Zhu, S.; Ogihara, M. Using discriminant analysis for multi-class classification: an experimental investigation. *Knowledge and information systems* **2006**, *10*, 453–472.
39. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press, 2008. doi:10.1017/CBO9780511809071.
40. Hsu, C.W.; Chang, C.C.; Lin, C.J.; others. A practical guide to support vector classification. *Technical Report, Department of Computer Science, National Taiwan University* **2003**.
41. Hastie, T.; Tibshirani, R. Discriminant adaptive nearest neighbor classification and regression. Advances in Neural Information Processing Systems, 1996, pp. 409–415.
42. Veenman, C.J.; Reinders, M.J. The nearest subclass classifier: A compromise between the nearest mean and nearest neighbor classifier. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2005**, *27*, 1417–1429.
43. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
44. Graupe, D. *Principles of artificial neural networks*; Vol. 7, World Scientific, 2013.
45. Rasmussen, C.E. Gaussian processes in machine learning. Summer School on Machine Learning. Springer, 2003, pp. 63–71.
46. Williams, C.K.; Rasmussen, C.E. *Gaussian processes for machine learning*; MIT press Cambridge, MA, 2006.
47. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Wang, W.; Powers, J.G. A description of the Advanced Research WRF version 3. NCAR Technical note-475+ STR. *NCAR Technical Note* **2008**.

48. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Wang, W.; Powers, J.G. A description of the advanced research WRF version 2. Technical report, National Center For Atmospheric Research Boulder Co Mesoscale and Microscale . . . , 2005.
49. Michalakes, J.; Dudhia, J.; Gill, D.; Henderson, T.; Klemp, J.; Skamarock, W.; Wang, W. The weather research and forecast model: software architecture and performance. In *Use of high performance computing in meteorology*; World Scientific, 2005; pp. 156–168.
50. KACARE. Renewable Resource Atlas. http://rratlas.energy.gov.sa, 2015. Accessed 1st Oct 2015.
51. Zell, E.; Gasim, S.; Wilcox, S.; Katamoura, S.; Stoffel, T.; Shibli, H.; Engel-Cox, J.; Subie, M.A. Assessment of solar radiation resources in Saudi Arabia. *Solar Energy* **2015**, *119*, 422–438. doi: http://dx.doi.org/10.1016/j.solener.2015.06.031.
52. Notaro, M.; Alkolibi, F.; Fadda, E.; Bakhrjy, F. Trajectory analysis of Saudi Arabian dust storms. *Journal of Geophysical Research: Atmospheres* **2013**, *118*, 6028–6043.
53. GFS. *Global Forecast System Model*, 2020 (accessed February 1, 2020).
54. USGS. *U.S. Geological Survey*, 2020 (accessed February 1, 2020).
55. Grell, G.A.; Dévényi, D. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters* **2002**, *29*.
56. Mlawer, E.J.; Taubman, S.J.; Brown, P.D.; Iacono, M.J.; Clough, S.A. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres* **1997**, *102*, 16663–16682.
57. Dudhia, J. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences* **1989**, *46*, 3077–3107.
58. Hong, S.Y.; Noh, Y.; Dudhia, J. A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review* **2006**, *134*, 2318–2341.
59. Hedar, A.R.; Abdel-Hakim, A.E.; Almaraashi, M. Granular-based dimension reduction for solar radiation prediction using adaptive memory programming. Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion. ACM, 2016, pp. 929–936.