Sridaran Rajagopal · Priti Sajja ·
Rohit Thanki · Ajay Kumar (Eds.)

Communications in Computer and Information Science 2821

Artificial Intelligence Based Smart and Secured Applications

4th International Conference, ASCIS 2025
Gujarat, India, September 11–13, 2025
Revised Selected Papers, Part III

Part 3

Springer

Contents

Artificial Intelligence and Machine Learning

Voice Veritas: Fake Voice Detection Using Deep Learning 3
A. Anju, J. Assmiya, D. Archana, S. Heshikaa, S. Muthuselvan, and M. Beema Mehraj

An In-Depth Analysis of Emerging Technologies in Contemporary Healthcare- A Review 13
Kamal Saluja, Tanya Khaneja, Sunil Gupta, Reema Gupta, and Vikas Solanki

Hybrid Knowledge Base Architecture for Real-Time Conversational AI 36
A. Vijaya Bharathi and Prashant Nitnaware

Multimodal Transformer-Based Classification of Breast Cancer Laterality 49
Shaesta Mujawar, Aisha Shaikh, and Anita Chaware

Deep Dive into Machine Learning for Early Breast Cancer Classification 67
D. Indhushree and R. Kokila

A Hybrid AI Framework for Personalized Foundational Learning Using Decision Trees and Fuzzy Logic 87
V. Prem Kumar and K. Rajakumari

Real-Time Knowledge Access Using Large Language Models and Web Technologies 102
M. Jayaram, Ch Jyothi, T Deepthi, Sri Ramya, and Paka Srikanth

Enhancing Bird Flu Outbreak Predictions Using Data Mining Techniques 117
K. Deeba, K. Jamberi, and B. Saratha

Meta-heuristic Ensemble Feature Selection (MEFS) and Stacking Deep Ensemble Classifier (SDEC) Model for Weather Prediction and Renewable Energy Demand Forecasting 133
Lekshmi Mohan and R. Durga

HySCAF: An Explainable and Efficient Feature Selection Strategy for Gestational Diabetes Prediction 159
Thejaswi Nandyala and K. Baalaji

Meta-heuristic Ensemble Feature Selection (MEFS) and Stacking Deep Ensemble Classifier (SDEC) Model for Weather Prediction and Renewable Energy Demand Forecasting

Lekshmi Mohan[1] and R. Durga[2(✉)]

[1] Department of Computer Science, Vels Institute of Science, Technology and Advanced Studies, Chennai, Tamil Nadu, India

[2] Department of Advanced Computing and Analytics, Vels Institute of Science, Technology and Advanced Studies, Chennai, Tamil Nadu, India

durga.scs@vistas.ac.in

Abstract. Wind speed and solar radiation varies is a predictable part of generating electricity from renewable resources. However, renewable energy data is unpredictable and disorganized is a difficult process. Deep learning is a technique for identifying the high-level invariant structures and intrinsic nonlinear features in the dataset. Enhancing the effectiveness of forecasting models requires feature selection, which is the process of identifying and choosing the most relevant elements from the dataset. In this paper, Meta-heuristic Ensemble Feature Selection (MEFS) model, and Stacking Deep Ensemble Classifier (SDEC) model are presented for feature selection and classification of wind speed and weather forecasting. The dataset, which was collected from Kaggle, contains four years' worth of Spanish weather, price, generating, and electrical consumption data. The dataset was sourced from the Transmission Service Operators (TSO) public portal, the European Network of Transmission System Operators for Electricity (ENTSOE). Min-max normalization was applied to pre-process the dataset. Entropy Binary Dragonfly Algorithm (EBDA), Adaptive Weight Dung Beetle Optimization (AWDBO), and Inertia Weight Wild Horse Optimizer (IWWHO) were combined into MEFS. SDEC is an ensemble method that combines the procedure of several models, such as the Peephole Long Short-Term Memory Network (PLSTM), Conditional Generative Adversarial Network (CGAN), and Lagrange Contractive Auto Encoder (LCAE)). SDEC model improves forecasting performance by training a meta-learner (Weighted Averaging Ensemble (WAE)) on the outputs of the underlying models. PLSTM gates can enter the cell since they have direct links or peepholes to the cell state. CGAN creates a corresponding number of samples to predict electricity from renewable sources including forecasted and actual samples. LCAE is an unsupervised Artificial Neural Network (ANN) with a regularization term using a Lagrange multiplier for forecasting prediction. Finally, WAE is used to merge the results of individual models. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Pearson Correlation Coefficient (r), Nash Sutcliffe Efficiency (NSE), precision, recall, f-measure, and accuracy metrics is used to evaluate the efficiency of forecasting approaches.

S. Rajagopal et al. (Eds.): ASCIS 2025, CCIS 2821, pp. 133–158, 2026.
https://doi.org/10.1007/978-3-032-17840-4_9

Keywords: Renewable energy forecasting · Stacking Ensemble Model · Meta-heuristic Ensemble Feature Selection (MEFS) · Stacking Deep Ensemble Classifier (SDEC) · Peephole Long Short Term Memory Network (PLSTM) · Conditional Generative Adversarial Network (CGAN) · Lagrange Contractive Auto Encoder (LCAE) · Weighted Averaging Ensemble (WAE) · weather prediction

1 Introduction

Renewable energy is essential for lowering greenhouse gas emissions and slowing down climate change [1, 2]. Research and development in renewable energy has received a lot of attention lately due to the rising demand for clean and sustainable energy. The possibilities for cost savings, job development, and reduce in dependence on significance energy are the benefits of renewable energy sources (RES). However, RES intrinsic instability and unpredictability pose a serious difficulty to the broad use of renewable energy [3, 4]. The weather has a big impact on wind energy production, and it can be hard to anticipate with any degree of accuracy [5, 6].

Physical models estimated the evolution of meteorological observations using atmospheric motion equations to forecast wind speed in the future [7, 8]. However, the time and cost required to create a physical model could result in poor forecasting accuracy [9, 10]. Wind speed prediction is essential for integrating wind turbines into smart networks and enhancing control over energy generation. A number of data-based methods have been created to enhance wind energy production estimation [11, 12]. Then, utilizing wind energy curves unique to each turbine, these projections are translated into anticipated energy outputs. Integration into a power grid is challenging, nevertheless, because wind speed varies greatly and is mostly affected by weather [13].

Accurate wind speed estimates are necessary for wind turbine integration into the electrical grid. Data-driven models, historical datasets can be categorized as classical statistical techniques, and deep learning-based models for prediction [14]. Traditional statistical tools can explain linear dependencies in time series data, including autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) are very useful for short-term forecasting [15]. However, these models effort to reflect the difficult and nonlinear dynamics intrinsic in long-term wind energy forecasting, and they decline as the forecast possibility rises. Addressing this issue involves developing accurate forecasting models for the renewable energy demand forecasting.

Numerous sophisticated models that can capture complex nonlinear interactions and handle complex time-series data have demonstrated great potential to the development of deep learning. Complex input-output mappings can be learned by the Multilayer Perceptron (MLP) [16], a Feedforward Neural Network (FFNN). Recurrent Neural Network (RNN) [17, 18] is especially made to keep an internal state that enables the model to take historical data when generating predictions to manage sequential input. The ability of RNN variants including Gated Recurrent Unit (GRU), Bidirectional Long Short-Term Memory (Bi-LSTM), and LSTM networks to efficiently handle long-term dependencies has led to their exceptional effectiveness in wind energy forecasting [18, 19]. DL approaches has been introduced for the renewable energy forecasting. Creating reliable

and understandable models that can shed light on the variables affecting the production of renewable energy is also crucial.

Therefore, feature selection is a crucial step to achieve a reliable solution for wind speed forecasting and weather prediction [12]. Aim is to create a new ensemble model that uses categorization and a novel meta-heuristic optimization technique to predict wind speed. In terms of search behavior, meta-heuristics exhibit two traits: diversification and intensification [20]. Diversifying involves developing several solutions that examine the search field globally, whilst intensifying involve restricting the search field to a find location. A strong balance between intensity and diversity should be maintained when choosing solutions. Ensemble methods are learning algorithms that classify new samples by employing a set of classifiers and weighted votes of their predictions. The way the underlying models are merged and trained are two of the many variables that affect how effective an ensemble approach is. In the literature, there are standard methods for creating ensemble models that have been effectively shown in a number of fields [21, 22]. Ensemble approaches work well when there is a demand for high predicted accuracy, complicated relationships, and a lot of data. Typical techniques for the ensemble method include stacking, blending, boosting, and bagging.

Bagging is a technique that lowers variation and increases predicted accuracy by utilizing numerous models. By providing additional weight to the data points that earlier models misclassified, the process known as "boosting" iteratively creates a powerful prediction model. By averaging across several models, bagging lowers variance; boosting concentrates on lowering bias by using weighted data points. The process of blending entails utilizing a different model that was trained on the validation set to combine the predictions of several different methods on the same dataset. A meta-model is created by stacking many basic models together [33]. Because the meta-model is more robustly trained, stacking frequently produces marginally superior results among them. Furthermore, an ensemble may produce more stable findings than a single approach, which would produce a subset that could be regarded as a local optimum.

In this paper, min-max normalization, which rescales data values to a predetermined range. MEFS was established to integrate several feature selection techniques, such as the Inertia Weight Wild Horse Optimizer (IWWHO), Adaptive Weight Dung Beetle Optimization (AWDBO), and Entropy Binary Dragonfly Algorithm (EBDA). SDEC model is worked by training base models (PLSTM, CGAN, and LCAE) with meta-learner (Weighted Averaging Ensemble (WAE)). Two datasets for wind speed and weather prediction are used to evaluate the suggested method.

2 Literature Review

Ragupathi et al., [24] presented the Deep Energy Predictor Model (DEPM) that enhances the accuracy and effectiveness of energy consumption prediction by combining the benefits of XGBoost and Deep Neural Network (DNN). The Cascaded Residual Network (ResNet) is introduced which takes use of DL capabilities to ensure that the finds most intricate and relevant patterns in the dataset. XGBoost reliable performance in classification tests with DNN ability to recognize patterns in the prediction tool. Measurements to household electric power usage were gathered via Kaggle and are included in the Household Electric Power usage Dataset. Based on the extensive (47-month) time span from

December 2006 to November 2010, it has a wide range of information (Kagglec.om, 2024). The model is evaluated using precision, recall, f1-score, and accuracy. DEPM significantly improves the accuracy of energy forecasting, and it is powerful tool for energy growth.

Lei et al., [25] proposed a Entropy-Weighted K-Means- Random Forest (EWKM-RF) method, and the RF-SSA-BiLSTM prediction model for building energy consumption. Sparrow Search Algorithm (SSA) is used to optimize the Bidirectional Long Short-Term Memory (BiLSTM). Data on energy usage from a civil public building in Dalian city is collected and evaluated to evaluate the accuracy of the proposed model. EWKM-RF technique enables exclusive feature selection and categorization of the features influencing energy usage. The result of forecasting model has been assessed using the MAE, Mean Average Percentage Error (MAPE), and RMSE.

Karijadi and Chou [26] suggested a hybrid method for predicting energy consumption that combines LSTM and RF using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). First, a number of components are extracted from the initial energy consumption data using CEEMDAN. Second, LSTM is used to forecast the other components, whereas RF is utilized to anticipate the component with the highest frequency. Thirdly, all of the components prediction findings are combined to get the final prediction results. The proposed method has been experimented on real-world building energy usage dataset.

Alghamdi et al., [27] suggested a novel technique that can generalize the forecast accuracy for data on solar radiation and wind speed. Genetic Algorithm Al-Biruni Earth Radius (GABER) technique is used to improve the forecasting accuracy and ability of the ensemble model by optimizing its parameters. Experiments are carried out to determine the effectiveness and superiority of the proposed strategy using two datasets (wind speed and solar radiation). Wilcoxon signed-rank tests and Analysis of Variance (ANOVA) are used to assess the consequence and variation of the proposed approach. NSE, Pearson Correlation Coefficient (r), covariance coefficient (R2), RMSE, Mean Bias Error (MBE), and identify agreement (WI) has been used to measure the model.

Atef and Eltawil [28] proposed to investigate into the systematic effects of BiLSTM and deep-stacked unidirectional (Uni-LSTM) networks on energy consumption prediction. Two and three LSTM layers are explicitly contrasted to the single-layered LSTM to demonstrate the significance of adding the stacked layers. To select the optimal model, each proposed model was submitted to a hyperparameter optimization process. A deep-stacked LSTM layer outperforms single-layered models in terms of predictive accuracy. Empirical four-year dataset from Switzerland (2015–2018) was used to assess proposed model. The training approach required a total of 28,032-time steps. Furthermore, a 5606-time step from the dataset was used to validate the generated model with 50 iterations, demonstrating the related efficiency. Trained network was then fed the 7032-time steps to determine the matching predictable output. Outcomes are measured using the RMSE, MAPE, and MAE.

Eseye et al., [29] suggested an integrated feature selection based on Machine Learning (ML) to identify the most relevant and non-redundant features for accurate energy demand forecasting. The fitness score of the features was determined using Gaussian Process Regression (GPR), and feature selection using Binary Genetic Algorithm (BGA).

A Feed Forward Artificial Neural Network (FFANN) model is introduced to energy demand forecasting. The model was trained on a two-year hourly dataset and tested on a one-year hourly dataset in the Otaniemi neighbourhood of Espoo, Finland. The total electricity consumption of the buildings peaked in 2015 at 221 kW, 592 kW, 29 kW, and 86 kW, respectively. This model improves the effectiveness by choosing fewer predictors to attain better prediction accuracy and a lowest MSE.

El-Kenawy et al., [30] introduced an ensemble model by weighted to wind speed prediction. Forecasts pertaining to wind speed and weight value optimization are made using a new ensemble model and the new Adaptive Dynamic Grey Wolf-Dipper Throated Optimization (ADGWDTO) technique. GWO algorithm mimics the dynamic group-based cooperative to balance between exploration and exploitation. K-Nearest Regressor (KNR), LSTM, and MLP regression models with hyperparameters are all improved by the ADGWDTO algorithm. The robustness and stability of the proposed technique are verified using one-way Analysis of variance (ANOVA) and Wilcoxon rank-sum.

Jiang et al., [31] suggested a multi-step wind power forecasting method called the Bidirectional Gated Recurrent Unit (BiGRU). Neural Network clustering and Multilabel integrated Random Forest (MLRF) feature selection from its basis (NNClustering).The MLRF technique provides RF based on many criteria to find the best time steps and input features for multistep forecasting task of multifactor time series. In addition to reducing the computational cost, this enhances the model ability for generalization. NNClustering strategy, which generates a unique convolution-based clustering and optimizes the clustering centers by adjusting the parameters using the gradient descent method. Whale Optimization Algorithm-Bidirectional Gated Recurrent Unit (WOA-BiGRU) forecasting model is introduced which reducing modelling complexity. BiGRU model has been worked based on data sequences in both directions to extract features, WOA is introduced which optimizes the parameters of BiGRU model to produce the best forecasting model. Hybrid strategy offers a new and effective way to predict wind energy due to its superior forecasting capabilities.

Abbasimehr et al., [32] introduced a new two-stage forecasting process consisting of model generation and data training. Using the first input data and the learned features, the building model generates the forecasting model. Temporal Convolution Neural Network (TCNN) and Multi-Head Attention (MHA) are applied to renewable energy consumption prediction. The data training model, statistical features are extracted from the input data, and an eXtreme Gradient Boosting (XGBoost) is used to extract the best feature subset. Tests reveal that TCNN benefits from feature addition, and deep learning models based on MHA perform better than both the benchmark models and their conventional methods.

3 Methodology

In this paper, Meta-heuristic Ensemble Feature Selection (MEFS), and Stacking Deep Ensemble Classifier (SDEC) is introduced for feature selection, wind speed classification, and weather forecasting. Firstly, the weather and electricity usage datasets were gathered from Kaggle. Secondly, the dataset is normalized using min-max normalization. Thirdly, MEFS is developed based on various approaches such as EBDA, AWDBO, and IWWHO. Fourthly, SDEC model is introduced based on number of base models like

PLSTM, CGAN, and LCAE, are combined to create the SDEC model. The outcomes of these techniques are combined using the SDEC model and a meta-learner (WAE). The performance of forecasting techniques has been evaluated using measures such as RMSE, MAE, r, NSE, precision, recall, f-measure, and accuracy. The overall flow diagram is illustrated in Fig. 1.

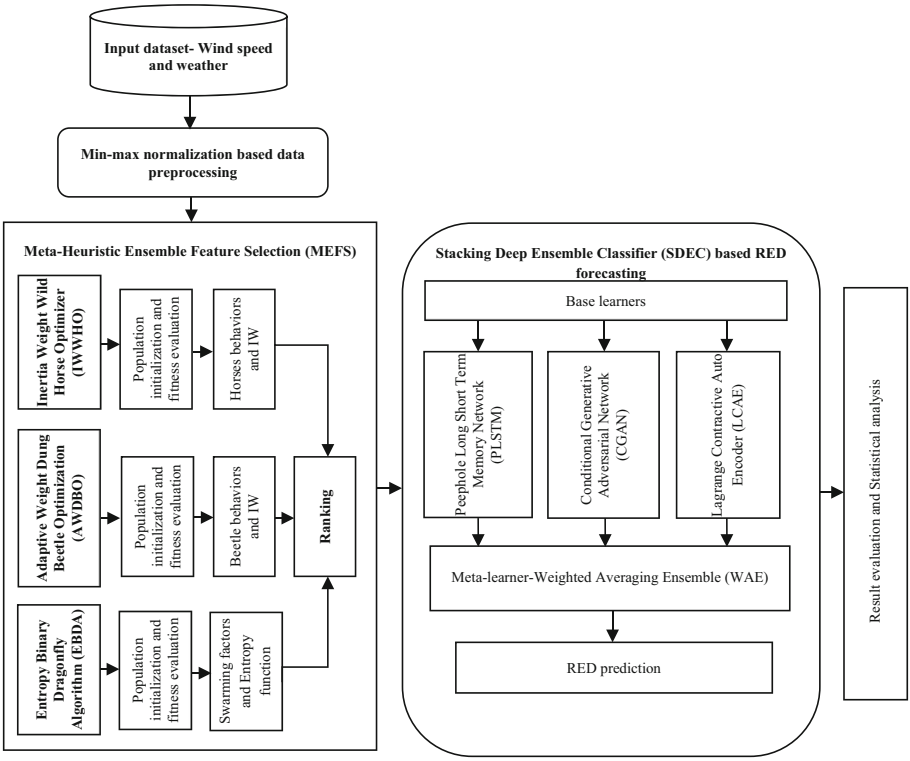

Fig. 1. Flow Diagram of Proposed System

3.1 Dataset and Preprocessing

Forecasting the energy market is one of the ML/DL domains that have the most influence on the shift to renewable energy-based electrical infrastructure. This dataset includes weather, pricing, generation, and consumption data for Spain during a four-year period. The TSO public portal provided the generating and consumption data. https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather/?select=weather_features.csv.is the source of the dataset. The hourly electrical usage statistics and the related TSO estimations for pricing and consumption make the dataset unique.

Min-Max Normalization: A normalizing method called min-max normalization rescales numbers to fall inside a given range, often 0 to 1. Using Eq. (1), min-max

normalization is performed by taking into account the features with the greatest and lowest values (x_{max}, x_{min}),

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

All other values are converted to decimal numbers between 0 and 1, with the highest value of a feature assigned to 1 and the minimum value set to 0.

Meta-heuristic Ensemble Feature Selection (MEFS): IWWHO, AWDBO, and EBDA results are combined using the MEFS technique. MEFS has been used to aggregate the attributes of multiple algorithms into a rating. It is also required to specify a threshold in order to obtain a functional subset of features.

Feature Selection. Feature selection is the method of removing an irrelevant number of features, and redundant features to create an accurate forecasting model. The transfer function generates probability values for swapping vector elements, which can be either 0 or 1. The quality of the given qualities is assessed using the fitness function [33, 34]. Its evaluation is done using Eq. (2).

$$\text{Fitness} = v_1 \text{Error} + v_2 \left| \frac{S}{T} \right| \tag{2}$$

where |T| is the total number of features, |S| is the chosen number of features, and Error is the classification error. With $v_1 = 1 - v_2$. The factors $v_1 \text{and} v_2$ have a variety of [0, 1].

EBDA: EBDA is developed depending on the swarming behavior of dragonflies [35]. The two behaviors of hunting and migration are the basis for choosing the most desirable features from the RED dataset. To find and capture the finest feature collection from the dataset, dragonflies fly in small subgroups. By following superior subgroups and pursuing swarms, dragonflies distinguish themselves for the best feature selection from the RED dataset [35]. A step vector is used in Eq. (3) to represent the dragonflies flight direction,

$$\Delta X_{t+1} = (sS_i + aA_i + cC_i + fF_i + eE_i) + w\Delta X_t \tag{3}$$

where the separation S_i, alignment A_i, cohesion C_i, attraction F_i, and distraction E_i in the ith search agent are weighted by S_i, A_i, C_i, F_i and E_i. The entropy function is denoted by inertia weight *w*.

AWDBO: DBO method was used to choose the optimal features from the dataset based on the rolling, dancing, foraging, thieving, and reproducing behaviors of the dung beetle. Four methods for population renewal are created in light of these developments [36].

Ball-Rolling Dung Beetles (BRDB): BRDB, forecasting accuracy is used to continuously update its location of the current feature in sunlight.

Breeding Dung Beetles (BDB): Boundary selection is the process by which dung beetles will eggs in a safe area to give their offspring a safe habitat.

Foraging Dung Beetles (FDB): Young dung beetles need to identify the best foraging area and guide themselves during foraging.

Stealing Dung Beetles (SDB): SDB is based on updating the dung beetle location using iterations.

Firstly, better balancing local exploitation and global exploration, adaptive weight continuously adjusts the weight factor, increasing the search efficiency and identifying optimal solutions. Secondly, promoting algorithm is introduced by enhancing its capacity to adjust features and reduces the likelihood of local optima convergence. DBO method is useful for balancing local exploitation and global exploration with dynamic weight factor alteration.

IWWHO: IWWHO is introduced which imitates the horse behavior to their group and connect another before achieving majority to avoid sister or daughter mating and achieve the best feature selection from the dataset. Horses social organization frequently allows for the division of them into two classes: defensive and non-defensive. In order to obtain her ideal feature selection from the dataset, mares and stallions mingle and interact socially while grazing. Foals leave their groups to form their own families as they reach adulthood and join other groups. This behavior prevents siblings and stallions from mating with the best qualities. IWWHO algorithm, which includes five main phases that are described below, is inspired by horse social behavior, including grazing, dominance, leadership hierarchy, and mating, while selecting features from the dataset [37].

Initial Population: If there are N individuals and G groups, then the number of leaders is G and the number of non-leaders (mares and foals) is N–G. The percentage of stallions expressed as G/N is called PS.

Grazing Behavior: throughout the best feature selection from the dataset, foals graze close to their herd throughout the most of their lives.

Horse Mating Behavior: One of the unique behaviors that distinguish horses from other animals is the separation of foals from their natal groups prior to puberty and mating.

Group Leadership: The other members of the group will follow the leaders or stallions to a suitable feature, such as a waterhole.

Leader Selection and Exchange: First, leaders are selected at random and their fitness levels are taken into consideration.

Ensemble Ranking: A suitable aggregation function is used to combine the ranking features into a single feature, and the rank determines the final score. Based on their overall ranking, the features are subsequently arranged in the final ensemble subset from most significant to least essential. Using the proper threshold, an ensemble subset of highly significant characteristics has been obtained. By varying threshold values till it satisfies stopping conditions, it is carried out independently to single and ensemble subsets. It is crucial to assess the selection process durability across a broad range of threshold values, as screening criteria may differ according on the dataset and the ranking criteria used. Lastly, the forecasting procedure uses the chosen attributes as input.

3.2 Stacking Deep Ensemble Classifier (SDEC) Model

SDEC model, multiple iterations of the base learners and the meta-learner are used to RED forecasting. The SDEC model, multiple base learners (PLSTM, CGAN, and

LCAE) are trained to produce a final output. Weighted Averaging Ensemble (WAE), a meta-learner is trained and combined using the output of these models. SDEC is the ability to create an ensemble using heterogeneous DL models, and stacking model is created by combining the results them. SDEC architecture is depicted in Fig. 2.

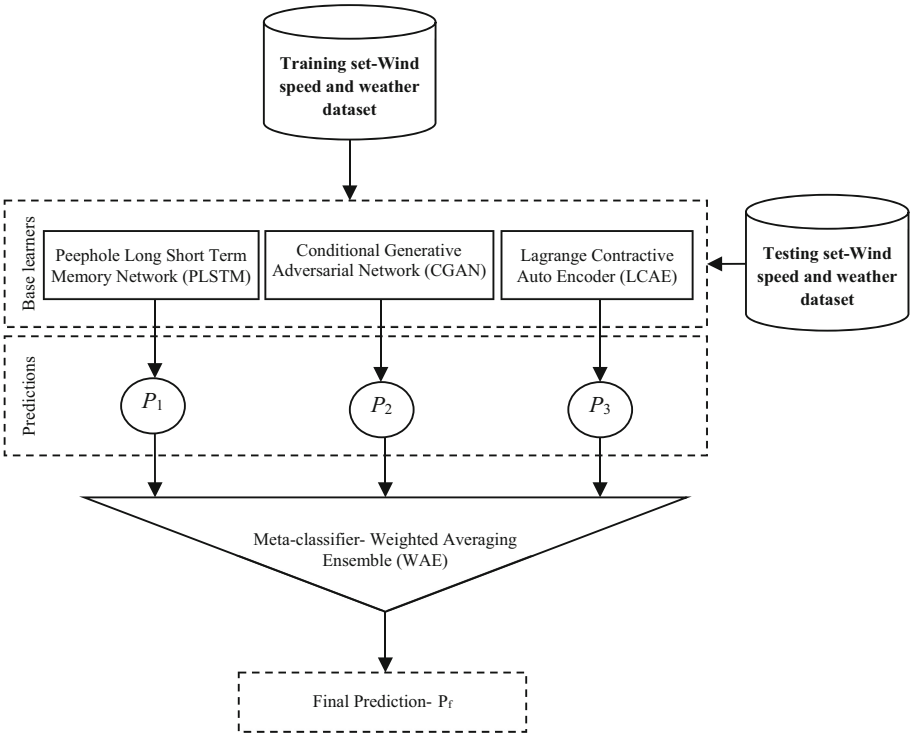

Fig. 2. Architecture of Stacking Deep Ensemble Classifier (SDEC)

3.3 Ensemble Model

The meta-learner training phase and the base model training phase are the two stages that typically comprise the SDEC model [23]. To begin, separate the original dataset (wind speed and weather) into training and testing sets. Then k-fold cross validation is introduced to train, and test the dataset. The training set is then split into k parts, each of which trains the model and uses the remaining $(k - 1)$ components of the dataset (weather and wind speed) to simulate the predictions for that portion of the meta-learner. The second step involves reassembling the predictions dataset (wind speed and weather) as the initial training dataset (wind speed and weather) after the basic model using k-fold cross validation.

3.4 Peephole Long Short Term Memory (PLSTM)

Long Short-Term Memory (LSTM) includes of three layers: the input layer, the hidden layer, and the output layer are used to assess whether an input is expected to be memorized and to create a forecasting prediction. The amount of the forget gate is preserved to this point in time with forecasting final cell state (c_{t-1}) [38]. A cell unit for storing datasets is also included in the gate in the blocks. However, LSTM has a number of drawbacks, including overfitting and training instability, sensitivity to hyperparameters, trouble processing lengthy sequences, and high computing costs. The output gate doesn't define the LSTM memory lifetime; rather, the interaction between the cell gate and the total gate enables learning and long-term dependence. LSTM storage unit now incorporates a peephole connector to overcome these issues. All gates can verify each other cell status to the peephole link.

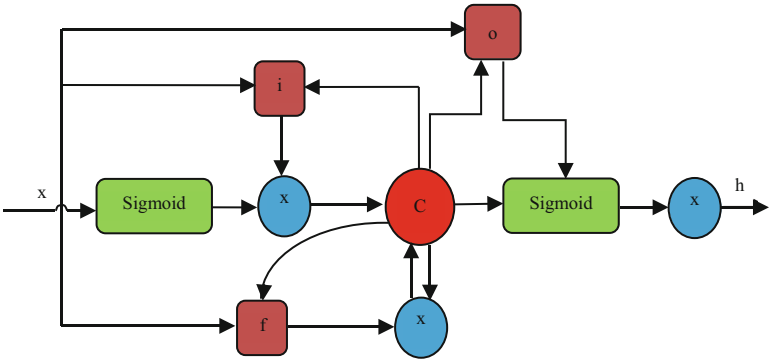

Fig. 3. Architecture of Peephole Connection

The construction of an LSTM memory block with a peephole link is shown in Fig. 3. The three exit arrows c_t to the three gates $i, f, \&o$ in the storage unit symbolizes the peephole connections of the cell state influences these gate values at last. These peephole connections show the input of c_{t-1} to the activation of the memory cells at time step $t-1$. The storage unit activation is considered while determining the input gate activation in the time step t. A little circle contains the symbol for the addition of its inputs and the multiplication of its parts. A PLSTM with a three-layer structure is illustrated in Fig. 4.

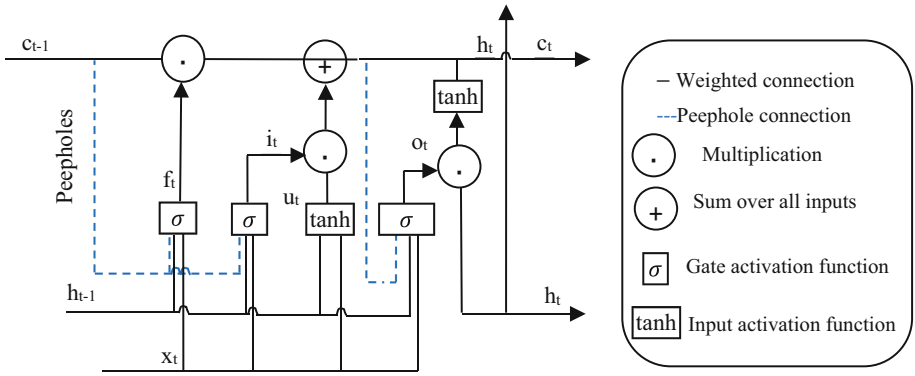

Fig. 4. Architecture of Peephole Long Short Term Memory (Plstm)

Input layer x_t and the previous hidden state h_{t-1}, the gates can also rely on the previous internal state c_{t-1} by creating peephole connections. This adds another term to the gate that also returns from the cell c_t, and the activation function δ_g generates the forget gate f_t. Sigmoid output, which has values between 0 and 1. It is used to evaluate the forget gate output, which indicates the capability to forget to envelop the state of the cell in the top layer. Equation (4) provides the numerical formulation for this output,

$$f_t = \delta_g(W_f x_t + V_f h_{t-1} + U_f c_{t-1} + b_f) \tag{4}$$

The input value, the cell state, and the peephole connection result is denoted by i_t which uses the δ_g,

$$i_t = \delta_g(W_i x_t + V_i h_{t-1} + U_i c_{t-1} + b_i) \tag{5}$$

The cell state update comes next. The sum of the two components yields the cell state. The forget gate output and c_{t-1} are multiplied to provide the first component, and the activation function x_t,c_{t-1},h_{t-1} yields the second portion. Equation (6) expresses the updated cell status,

$$c_t = i_t \delta_c(W_c x_t + U_f c_{t-1} + V_f h_{t-1} + b_i) + c_{t-1} f_t \tag{6}$$

The input value, cell state from the prior time, and the corresponding value that the peephole returned are majorly affect the decision of output gate. The connected cell state is defined by the δ_g, whereas the output gate is expressed by Eq. (7),

$$o_t = \delta_g(W_o x_t + U_o c_{t-1} + V_o h_{t-1} + b_o) \tag{7}$$

Equation (8) illustrates the function updates the hidden state by multiplying the output gate and the cell state,

$$h_t = o_t \delta_h(c_t) \tag{8}$$

Equation (9) which represents the output for the complete PLSTM,

$$y_t = k\left(W_h h_t + b_y\right) \tag{9}$$

where W_h is the hidden output weight matrix, W_f, W_i, W_o, V_f, V_i, V_o, and U_f, U_i, U_o are the weight matrices of forgotten, input, and output gates, and b_f, b_i, b_o, b_y are the bias vectors of forgot, input, output, and prediction gates. During PLSTM model training, the loss function is minimized using three optimization techniques: adaptive moment estimation (Adam), root mean square propagation (RMSProp), and stochastic gradient descent with momentum (SGDm).

3.5 Conditional Generative Adversarial Network (CGAN)

An advanced version of GAN called a CGAN. CGAN, condition is used to control output generation to predict weather and wind speed. CGAN includes of two steps like discriminator and generator. CGAN creates a dataset with the same structure observations corresponding to the same label using a generator with a random array as input and an identified class label. This network tries to classify the observations as actual or predicted datasets using discriminator with known batches of labeled dataset that contain observations from both the training dataset and the generated dataset from the generator [39].

Generator: Generator is an Artificial Neural Network (ANN) stack where each subsequent layer has twice as many cells. Consider concerning adding activation layers in among the dense layers. As a regularization technique improves the forecasting accuracy, the batch normalization layer is located in the last layer before the output layer for RED forecasting using wind speed and weather data. Given an input **z**, the generator is represented by Eq. (10) where **c** is a conditional input and **G** is a non-linear function. The result, $\overrightarrow{y}_g$ is a series of wind speed and meteorological data that illustrates the situation that Eq. (10),

$$G : (\vec{z}, \vec{c}) \mapsto \vec{y}_g \tag{10}$$

The output $\overrightarrow{y}_g$ is augmented and supplied into the discriminator as an extra stabilizing step to prevent mode collapse and overfitting.

Discriminator: Janson-Shannon divergence $\left(JS\left[\overrightarrow{y}_g \middle| \overrightarrow{x}_d\right]\right)$ and the finite log-likelihood-ratio between the expected dataset, $\overrightarrow{y}_g$ and the actual value, $\overrightarrow{x}_d$. For each sample, $y_g \epsilon \overrightarrow{y}_g$ and $x_d \epsilon \overrightarrow{x}_d$, random values are introduced to avoid overfitting on the wind speed and weather dataset. The discriminator model, dropout is used to decrease the overfitting by decreasing the interdependent learning among the neurons. Equation (11) describes the connection among input and output with each other using discriminator,

$$D : (\overrightarrow{y}, \widetilde{x}) \mapsto \overrightarrow{y}_d \text{ where } \overrightarrow{y} \supseteq \overrightarrow{y}_g, \overrightarrow{x}_d \tag{11}$$

Equations (12–13) provide the CGAN an objective loss function during training model which reflects the relationship between the discriminator and generator,

$$L_G = E\left[\log\left(D\left(G\left(\overrightarrow{z}, \overrightarrow{c}\right), \widetilde{x}\right)\right)\right] \tag{12}$$

$$L_D = E\left[\log\left(D\left(\overrightarrow{\mathbf{x}}_{\mathbf{d}}, \widetilde{x}\right)\right)\right] + E\left[\log\left(1 - D\left(G\left(\overrightarrow{z}, \overrightarrow{c}\right), \widetilde{x}\right)\right)\right] \tag{13}$$

where L_G and L_D is denoted as the loss functions of discriminator and generator.

Managing Discriminator Overfitting: Equation (14) represents the log-likelihood ratio of GAN.

$$\text{Log} - \text{likelihood ratio} = \log \frac{\overrightarrow{\mathbf{y}}_{\mathbf{g}}(\overrightarrow{\mathbf{y}}_{\mathbf{d}})}{\overrightarrow{\mathbf{x}}_{\mathbf{d}}(\overrightarrow{\mathbf{y}}_{\mathbf{d}})} \tag{14}$$

Nevertheless, G: $(\overrightarrow{Z}, \overrightarrow{C}) \mapsto \overrightarrow{\mathbf{y}}_{\mathbf{g}}$ where the support is $\left\{z \in \overrightarrow{z}, c \in \overrightarrow{C} : G(\mathbf{z}, \mathbf{c}) \neq 0\right\}$ in high-dimensional space, the intersection of the generator's support and the distribution's support that generated $\overrightarrow{\mathbf{x}}_{\mathbf{d}}$. Therefore, the goal of adding synthetic noise, as indicated in Eq. (11) is to give the two underlying distributions an overlapping support. This guarantees that JS divergence generates a continuous function that doesn't saturate to a constant value and that the log-likelihood stays finite hence decreasing the discriminator overfitting.

Embedding: This results in a wind and weather forecasting prediction ($\overrightarrow{y}_g$) of the similar dimension as the predicted samples ($\overrightarrow{x}_d$). A binary class called CGAN is created by embedding and produces an equal number of samples for test wind speed and weather conditions. Thus, the binary class CGAN ($\overrightarrow{y}_g$) generates as many weather and wind predictions as real data. To make sure that samples in the high-dimensional space match both distant points between them and locally embedded lower-dimensional points, the stochastic neighbor embedding technique is created. The neighbor embedding approach is used to calculate the conditional probabilities in Eq. (15), where $\left\{\overrightarrow{y}_g, \overrightarrow{x}_d\right\} \epsilon \overrightarrow{y}$, $P_{j/i}$ indicates the likelihood that predicted samples j would choose actual samples i as a neighbor.

$$P_{j/i} = \frac{\exp\left(-\frac{|y_i - y_j|^2}{2\sigma_i^2}\right)}{\sum_k \neq i \exp\left(-\frac{|y_i - y_k|^2}{2\sigma_i^2}\right)} \tag{15}$$

Equation (16) is used to determine the joint probability P_{ij} of the similarity among samples i and j to guarantee symmetry. $P_{j/i} = 0, P_{j/j} = 0$ the neighbor embedding approach generates two-dimensional embedding output in order to examine the CGAN performance using Eq. (16),

$$P_{ij} = \frac{P_{j/i} + P_{i/j}}{2N} \tag{16}$$

where N is the number of y rows.

3.6 Lagrange Contractive Auto Encoder (LCAE)

LAE is type of Artificial Neural Network (ANN) architecture for predicting wind speed and weather data. It has a decoder that reconstructs the wind speed and weather dataset and an encoder that converts a wind speed and weather dataset to a latent form [40].

Equation (17) provides a mathematical image of the encoding process f to convert the wind speed and weather dataset x into a concealed representation y,

$$y = f(x) = \phi(w_x + b_h) \tag{17}$$

In this case, h is denoted as the hidden layer. The activation function is denoted by ϕ, the bias term by b_h, and the input weights by w_x. Equation (18) has been used to express the decoding function g to map the hidden representation y back to a reconstructed input r,

$$r = g(y) = \phi'\left(w_y + b_r\right) \tag{18}$$

where ϕ' is the activation function, b_r is the decoding bias, and w_y are the weights applied to y. The input is converted to a hidden representation y by the encoder f, and then reconstructed as r by the decoder g. Reconstruction error is the ability to replicate the original input following encoding and decoding. Equation (19) is used to create the cost function with reconstruction error R given the input dataset $D_i = [x_1, x_2, x_3, \ldots..x_n]$,

$$J_{AE}(\theta) = \sum_{x \in D_i} R(x, r) \tag{19}$$

R(x, r) of the input x and output y is represented by Eq. (20),

$$R(x, r) = \|x - r\|^2 \tag{20}$$

AE cost function is modified to include the Lagrange regularization term in LCAE. Between the actual and reconstructed expected output, it reduces reconstruction error. A special kind of regularization that particularly encourages stability and robustness in the learned feature representations are offered by the contractive term. The model is able to actually concentrate on strong features that reflect the underlying causative causes to the contractive regularization. Via the cost function, the contractive penalty immediately introduces robustness and invariance into the optimization process. The regularization technique reduces the reconstruction error $R(x, r)$ which increases accuracy and stability. Equation (21) can be used to express the overall CAE cost function.

$$J_{LCAE}(\theta) = \sum_{x \in D_i} \left(R(x, r) + \lambda \|J_f(x)\|_f^2 \right) \tag{21}$$

where $J_f(x)$ is the Jacobian matrix of hidden layer activations to x, and λ is the Lagrange multiplier that regulates the contractive term weight.

3.7 Weighted Averaging Ensemble (WAE)

Weighted Averaging Ensemble (WAE) is introduced to combine the results of predictions from several techniques with increased forecasting accuracy [23]. Equation (22) can be used to define the WAE model's output.

$$WAE(x) = \sum_{k=1}^{K} f_i(x), x = 1, 2, \ldots N \tag{22}$$

where N is the dataset length and $\omega_i(x)$ is the weight of the k[th] base model with input (x). First, the sum of squared dispersions of each base model is determined with predictable values. Next, Eq. (23) is used to determine each model with their corresponding weights. Equation (23) can be used to compute the weight,

$$\omega_i(x) = \frac{D_k^{-1}(x)}{\sum_{k=1}^{K} D_k^{-1}(x)} \tag{23}$$

where $D_k^{-1}(x)$ is the square of the k[th] model prediction's deviance

4 Results and Findings

This section compares the performance of various RED forecasting techniques. The tests were carried out using MATLAB R2020a on a Windows 10 computer powered by an Intel® CoreTM i5-8500 Processor 9M Cache clocked at 4.10 GHz. The trials employed two separate datasets that were eventually combined to form a new dataset. To compare the outcomes, employed metrics such as RMSE, MAE, r, NSE, precision, recall, f-measure, and accuracy.

4.1 Prediction Results

The results of forecasting techniques are quantified using metrics such as RMSE, MAE, r, NSE, precision, recall, accuracy, and f-measure. In the dataset, $\widehat{\mathrm{V}}_\mathrm{n}$ and V_n is denoted as the n[th] observed and estimated values, respectively, N is the number of samples, $\overline{\widehat{V}_n}$ and $\overline{V_n}$ are the mathematics means of the $\widehat{V}_n$ and V_n values. True Positive (TP) which correctly predicts the positive outcome as positive by the model. True Negative (TN) which correctly predicts a negative outcome as negative by the model. False Positive (FP) that predicts a positive outcome but produces a negative by the model. False Negative (FN) that predicts a negative outcome but the actual result is positive by the model. The expected metrics are shown in Table 1.

Table 1. Evaluation Metrics for Forecasting Methods

METRICS	FORMULA
RMSE	$\sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\widehat{\mathrm{V}}_\mathrm{n}-\mathrm{V}_\mathrm{n}\right)^2}$
MAE	$\frac{1}{\mathrm{N}}\sum_{\mathrm{n}=1}^{\mathrm{N}}\left\lvert\widehat{\mathrm{V}}_\mathrm{n}-\mathrm{V}_\mathrm{n}\right\rvert$
R	$\frac{\sum_{\mathrm{n}=1}^{\mathrm{N}}(\widehat{\mathrm{V}}_\mathrm{n}-\overline{\widehat{\mathrm{V}}_\mathrm{n}})(\mathrm{V}_\mathrm{n}-\overline{\mathrm{V}_\mathrm{n}})}{\sqrt{\left(\sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\widehat{\mathrm{V}}_\mathrm{n}-\overline{\widehat{\mathrm{V}}_\mathrm{n}}\right)^2\right)\left(\sum_{\mathrm{n}=1}^{\mathrm{N}}(\mathrm{V}_\mathrm{n}-\overline{\mathrm{V}_\mathrm{n}})^2\right)}}$

(*continued*)

Table 1. (*continued*)

METRICS	FORMULA
NSE	$1 - \frac{\sum_{n=1}^{N} \left(V_n - \widehat{V}_n\right)^2}{\sum_{n=1}^{N} \left(V_n - \widehat{\overline{V}}_n\right)^2}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F-measure	$\frac{2.Pre.Rec}{Pre+Rec}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$

Lagrange Contractive Auto Encoder Ensemble (LCAEE), Conditional Generative Adversarial Network Ensemble (CGANE), Peephole Long Short Term Memory Ensemble (PLSTME), binary Genetic Algorithm Al-Biruni Earth Radius-Hermite Neural Network Ensemble (bGABER-HNNE) [27], Stacked Ensemble Model (SEM), and the proposed Stacking Deep Ensemble Classifier (SDEC) are among the wind speed forecasting techniques whose performance is compared in Table 2.

Table 2. Results Comparison of Forecasting Methods

Forecasting methods	RMSE	MAE	NSE	r	Results (%)			
					Precision	Recall	F-measure	Accuracy
LCAEE	0.0112	0.0064	0.8782	0.8895	89.95	88.14	89.03	88.65
CGANE	0.0105	0.0061	0.8941	0.9067	90.18	90.62	90.89	90.12
PLSTME	0.0097	0.0057	0.9125	0.9258	91.35	91.89	92.12	92.07
bGABER-HNNE	0.0085	0.0052	0.9481	0.9526	92.37	93.05	93.01	94.54
SEM	0.0044	0.0023	0.9717	0.9794	93.35	96.16	94.62	96.08
SDEC	0.0028	0.0011	0.9852	0.9875	97.48	99.23	98.35	98.99

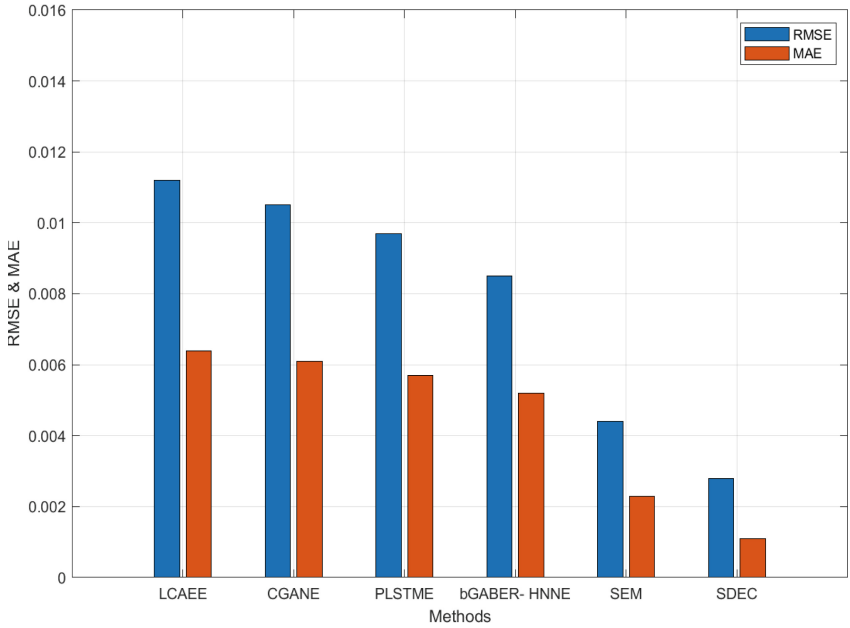

Fig. 5. RMSE & MAE Analysis of Forecasting Methods

RMSE and MSE comparison of CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM, and the SDEC methods are illustrated in Fig. 5. Proposed system has lowest RMSE results of 0.0028, other methods such as the CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM have highest RMSE values of 0.0112, 0.0105, 0.0097, 0.0085, and 0.0044. The proposed technique has decreased MAE of 0.0011, whereas CGANE, PLSTME, LCAEE, bGABER-HNNE, and SEM have highest MAE values of 0.0064, 0.0061, 0.0057, 0.0052, and 0.0023 respectively (See Table 2).

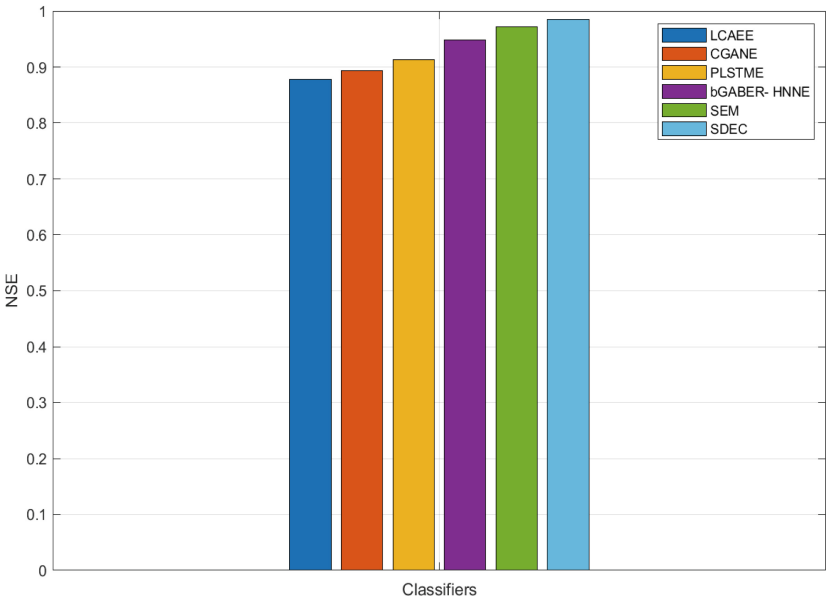

Fig. 6. NSE Comparison of Forecasting Methods

NSE comparison of CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM, and the SDEC are illustrated in Fig. 6. The proposed technique has a higher NSE value of 98.52% than alternative techniques such as CGANE, PLSTME, LCAEE, bGABER-HNNE, and SEM gives lowest NSE values of 87.82%, 89.41%, 91.25%, 94.81%, and 97.17%. CGANE, PLSTME, LCAEE, bGABER-HNNE, and SEM have10.86%, 9.25%, 7.38%, 3.77%, and 1.37% increased NSE value when compared to SDEC model (Table 2).

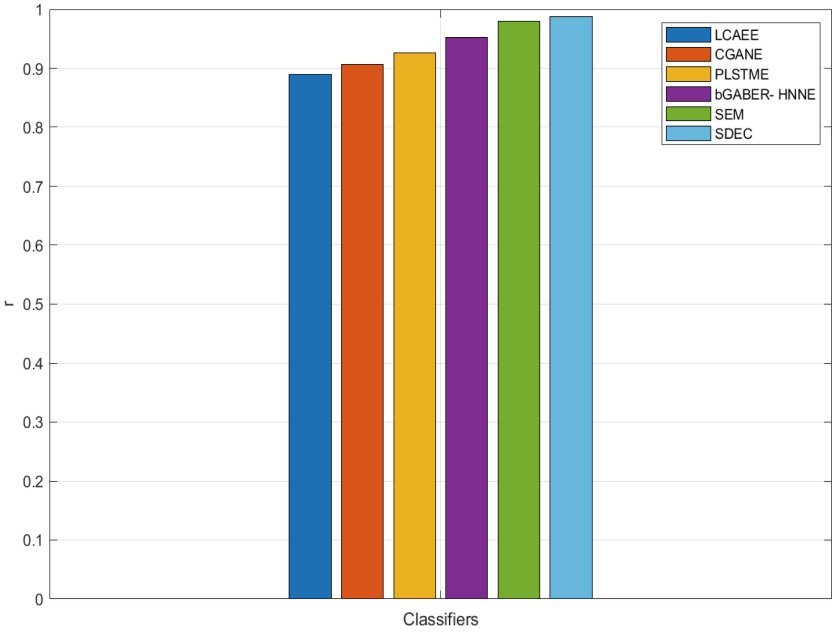

Fig. 7. Pearson Correlation Coefficient (R)Comparison of Forecasting Methods

Pearson Correlation Coefficient of CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM, and the SDEC are shown in Fig. 7. The proposed system has highest r value of 98.75%, other methods such as CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM have lesser r values of 88.95%, 90.67%, 92.58%, 95.26%, and 97.94%. The proposed system has 9.92%, 8.18%, 6.25%, 3.53%, and ~1.00% increased r value when compared CGANE, PLSTME, LCAEE, bGABER-HNNE, and SEM respectively (See Table 2).

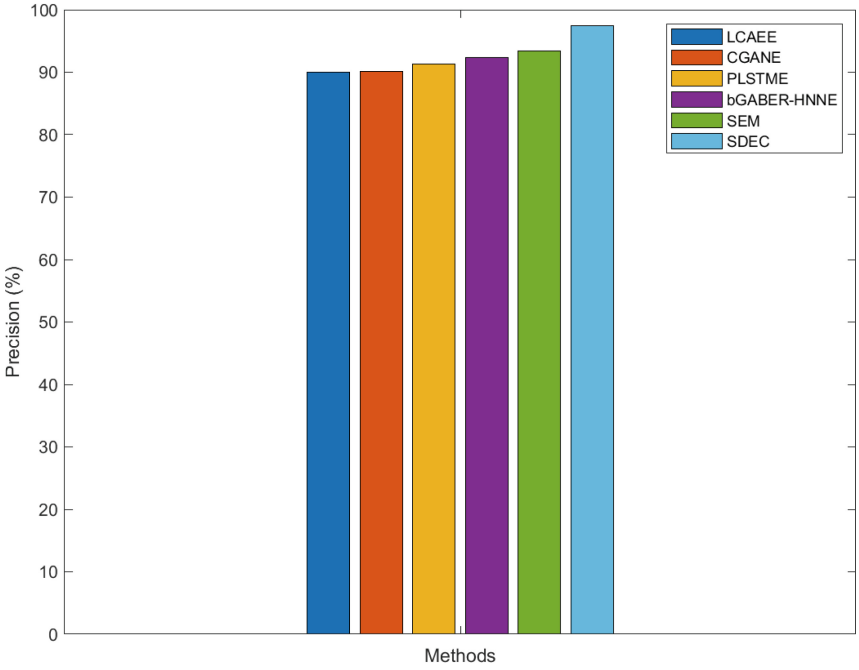

Fig. 8. Precision Comparison of Forecasting Methods

Precision comparison of CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM, and the SDEC are shown in Fig. 8. SDEC model has highest precision of 97.48%, other methods such as the CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM have lower precision value of 89.95%, 90.18%, 91.35%, 92.37%, and 93.35%. SDEC model has 7.53%, 7.30%, 6.13%, 5.11%, and 4.13% increased precision value when compared CGANE, PLSTME, LCAEE, bGABER-HNNE, and SEM respectively.

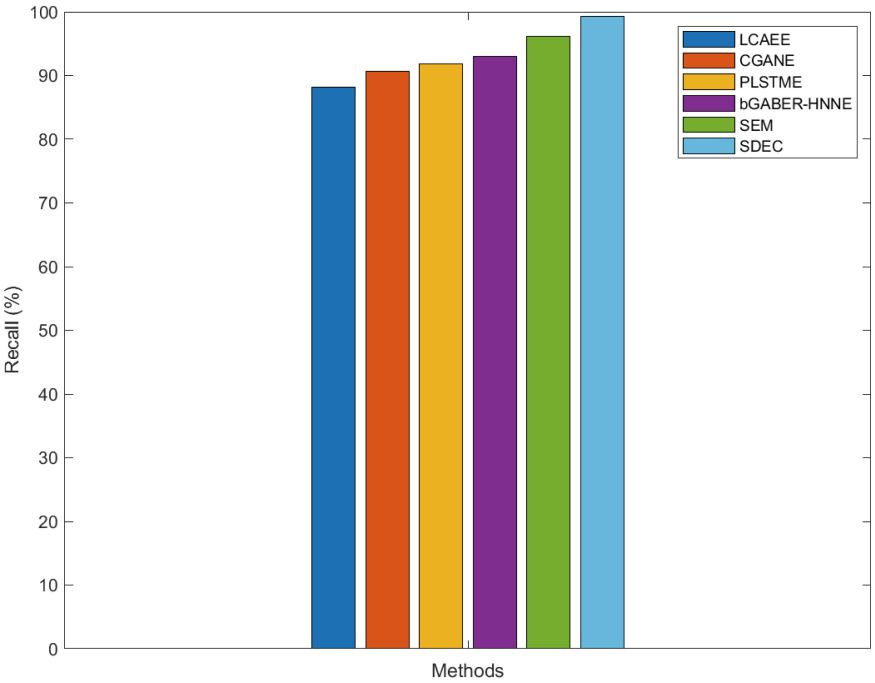

Fig. 9. Recall Comparison of Forecasting Methods

CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM, and the SDEC with respect to recall comparison are shown in Fig. 9. SDEC model has highest recall of 99.23%, other methods such as the CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM have lower recall values of 88.14%, 90.62%, 91.89%, 93.05%, and 96.16%. SDEC system has 11.09%, 8.61%, 7.34%, 6.18%, and 3.07% increased recall when compared CGANE, PLSTME, LCAEE, bGABER-HNNE, and SEM respectively.

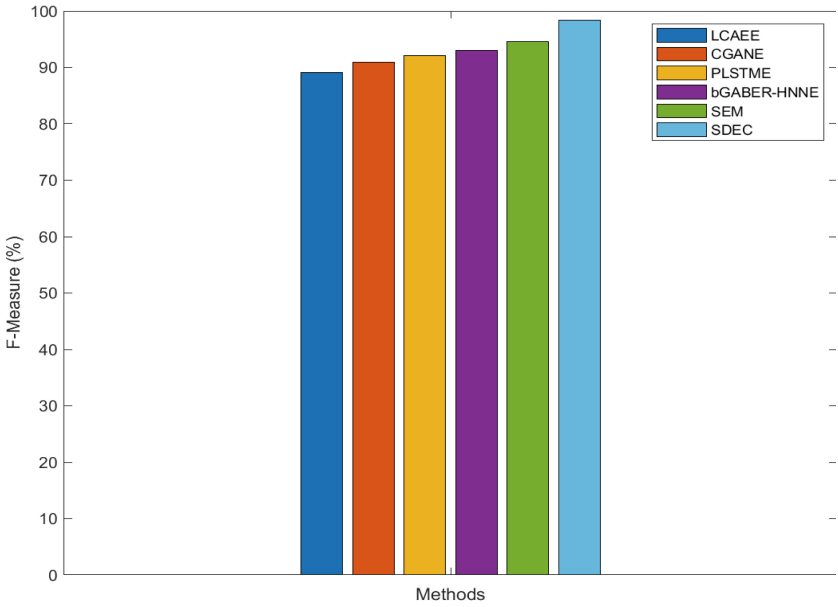

Fig. 10. F-Measure Comparison of Forecasting Methods

CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM, and the SDEC with respect to F-measure are shown in Fig. 10. SDEC model has highest f-measure of 98.35%, other methods such as the CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM have lower f-measure of 89.03%, 90.89%, 92.12%, 93.01%, and 94.62%. SDEC system has 9.32%, 7.46%, 6.23%, 5.34%, and 3.73% increased f-measure when compared CGANE, PLSTME, LCAEE, bGABER-HNNE, and SEM respectively.

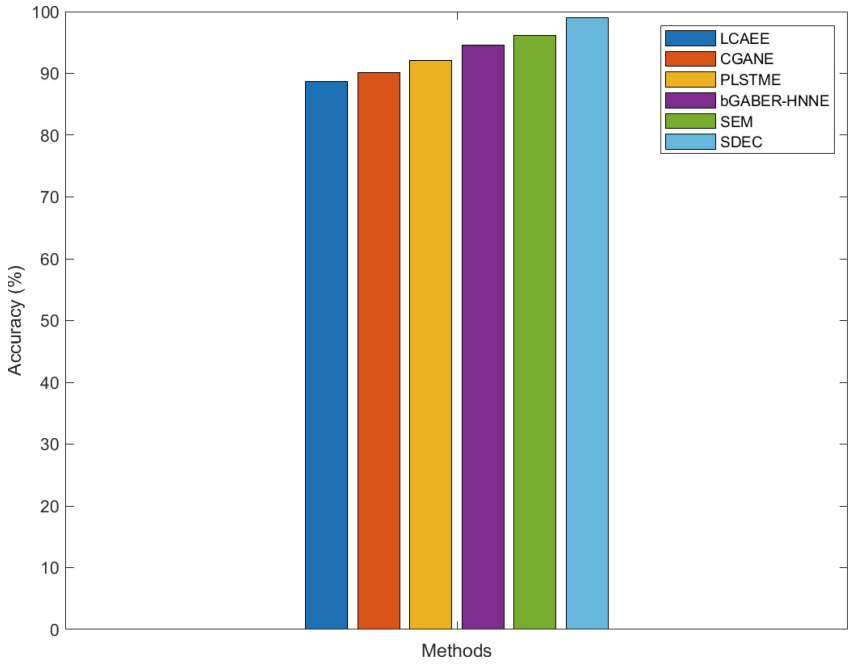

Fig. 11. Accuracy comparison of Forecasting Methods

Accuracy comparison of CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM, and the SDEC are shown in Fig. 11. SDEC model has highest accuracy of 98.99%, CGANE, PLSTME, LCAEE, bGABER-HNNE, SEM have lower accuracy of 88.65%, 90.12%, 92.07%, 94.54%, and 96.08%. SDEC system has 10.34%, 8.87%, 6.92%, 4.45%, and 2.91% increased accuracy when compared CGANE, PLSTME, LCAEE, bGABER-HNNE, and SEM respectively.

5 Conclusion

In this paper, Meta-Heuristic Ensemble Feature Selection (MEFS), and ensemble classification model is introduced for estimating wind speed from meteorological data. The EBDA, WDBO, and IWWHO are the foundations of the MEFS method. The findings of each approach are combined through the ensemble ranking. SDEC model, several base learners (Peephole Long Short-Term Memory Network (PLSTM), Conditional Generative Adversarial Network (CGAN), and Lagrange Contractive Auto Encoder (LCAE)) are trained, and meta-learner (Weighted Averaging Ensemble (WAE)) is trained using the output of base learners. PLSTM, the LSTM storage unit now incorporates a peephole association allows all gates to ensure the cell status of them. CGAN includes of both the generator and discriminator is conditioned on some additional data. CGAN is an advanced type of GAN and its output generation is controlled using a condition. LCAE, Lagrange regularization term, and contractive term is introduced into the cost function. It minimizes reconstruction error between the actual and reconstructed predicted output.

WAE model, predictions of individual methods is combined to improve the forecasting accuracy. Metrics like r, RMSE, MAE, NAE, precision, recall, f-measure, and accuracy are used to assess the results of RED methods. Hyperparameters optimization of classifiers plays a significant role for enhancing RED prediction. In addition, other classifiers such as Reinforcement Learning (RL), and other DL with transform learning methods have been also introduced for forecasting.

References

1. Gielen, D., Boshell, F., Saygin, D., Bazilian, M.D., Wagner, N., Gorini, R.: The role of renewable energy in the global energy transformation. Energy Strategy Rev. **24**, 38–50 (2019)
2. Strielkowski, W., Civín, L., Tarkhanova, E., Tvaronavičienė, M., Petrenko, Y.: Renewable energy in the sustainable development of electrical power sector: a review. Energies **14**, 1–24 (2021)
3. Kumar J.C.R., Majid, M.A.: Renewable energy for sustainable development in India: Current status, future prospects, challenges, employment, and investment opportunities. Energy Sustain. Soc. **10**, 1–36 (2020)
4. Denholm, P., et al.: The challenges of achieving a 100% renewable electricity system in the United States. Joule **5**, 1331–1352 (2021)
5. Nazir, M.S., et al.: Wind generation forecasting methods and proliferation of artificial neural network: a review of five years research trend. Sustainability **12**, 1–27 (2020)
6. Lledó, L., Torralba, V., Soret, A., Ramon, J., Doblas-Reyes, F.: Seasonal forecasts of wind power generation. Renew. Energy **143**, 91–100 (2019)
7. Ding, F., Tian, Z., Zhao, F., Xu, H.: An integrated approach for wind turbine gearbox fatigue life prediction considering instantaneously varying load conditions. Renew. Energy **129**, 260–270 (2018)
8. Han, S., Hui Qiao, Y., Yan, J., Qian Liu, Y., Li, L., Wang, Z.: Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. Appl. Energy **239**, 181–191 (2019)
9. Bouyeddou, B., Harrou, F., Saidi, A., Sun, Y.: An Effective Wind power prediction using latent regression models. In: Proceedings of the 2021 International Conference on ICT for Smart Society (ICISS), Bandung City, Indonesia, 2–4 August 2021, pp. 1–6 (2021)
10. Yan, J., Ouyang, T.: Advanced wind power prediction based on data-driven error correction. Energy Convers. Manag. **180**, 302–311 (2019)
11. Hu, J., Heng, J., Wen, J., Zhao, W.: Deterministic and probabilistic wind speed forecasting with de-noising-reconstruction strategy and quantile regression based algorithm. Renew. Energy **162**, 1208–1226
12. Hanifi, S., Liu, X., Lin, Z., Lotfian, S.: A critical review of wind power forecasting methods—past, present future. Energies **13**, 1–24 (2020)
13. Alkesaiberi, A., Harrou, F., Sun, Y.: EfficientWind power prediction using machine learning methods: a comparative study. Energies **15**, 1–24 (2022)
14. Bouabdallaoui, D., Haidi, T., El Jaadi, M.: Review of current artificial intelligence methods and metaheuristic algorithms for wind power prediction. Indonesian J. Electr. Eng. Comput. Sci. **29**(2), 626–634 (2023)
15. Wang, J., Zhou, Q., Zhang, X.: Wind power forecasting based on time series ARMA model. In: IOP Conference Series: Earth and Environmental Science, vol. 199, no. 2, p. 022015. IOP Publishing, pp.1–6 (2018)
16. Samadianfard, S., et al.: Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. Energy Rep. **6**, 1147–1159 (2020)

17. Zhang, J., Jiang, X., Chen, X., Li, X., Guo, D., Cui, L.: Wind power generation prediction based on LSTM. In: Proceedings of the 2019 4th International Conference on Mathematics and Artificial Intelligence, pp. 85–89 (2019)
18. Kisvari, A., Lin, Z., Liu, X.: Wind power forecasting–a data-driven method along with gated recurrent neural network. Renewable Energy **163**, 1895–1909 (2021)
19. Ding, M., Zhou, H., Xie, H., Wu, M., Nakanishi, Y., Yokoyama, R.: A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. Neurocomputing **365**, 54–61 (2019)
20. Peres, F., Castelli, M.: Combinatorial optimization problems and metaheuristics: review, challenges, design, and development. Appl. Sci. **11**(14), 1–31 (2021)
21. Mohammed, A., Kora, R.: A comprehensive review on ensemble deep learning: opportunities and challenges. J. King Saud Univ. Comput. Inf. Sci. **35**, 757–774 (2023)
22. Pirbazari, A.M., Sharma, E., Chakravorty, A., Elmenreich, W., Rong, C.: An ensemble approach for multi-step ahead energy forecasting of household communities. IEEE Access **9**, 36218–36240 (2021)
23. Lu, M., et al.: A stacking ensemble model of various machine learning models for daily runoff forecasting. Water **15**(7), 1–19 (2023)
24. Ragupathi, C., Dhanasekaran, S., Vijayalakshmi, N., Salau, A.O.: Prediction of electricity consumption using an innovative deep energy predictor model for enhanced accuracy and efficiency. Energy Rep. **12**, 5320–5337 (2024)
25. Lei, L., Shao, S., Liang, L.: An evolutionary deep learning model based on EWKM, random forest algorithm, SSA and BiLSTM for building energy consumption prediction. Energy **288**, 129795 (2024)
26. Karijadi, I., Chou, S.Y.: A hybrid RF-LSTM based on CEEMDAN for improving the accuracy of building energy consumption prediction. Energy Build. **259**, 111908 (2022)
27. Alghamdi, A.A., Ibrahim, A., El-Kenawy, E.S.M., Abdelhamid, A.A.: Renewable energy forecasting based on stacking ensemble model and Al-Biruni earth radius optimization algorithm. Energies **16**(3), 1–30 (2023)
28. Atef, S., Eltawil, A.B.: Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting. Electr. Power Syst. Res. **187**, 1–11 (2020)
29. Eseye, A.T., Lehtonen, M., Tukia, T., Uimonen, S., Millar, R.J.: Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems. IEEE Access **7**, 91463–91475 (2019)
30. El-Kenawy, E.S.M., et al.: Feature selection in wind speed forecasting systems based on meta-heuristic optimization. PLoS ONE **18**(2), 1–28 (2023)
31. Jiang, Z., Tan, Q., Li, N., Che, J., Tan, X.: A novel BiGRU multi-step wind power forecasting approach based on multi-label integration random forest feature selection and neural network clustering. Energy Convers. Manage. **319**, 118904 (2024)
32. Abbasimehr, H., Paki, R., Bahrini, A.: A novel XGBoost-based featurization approach to forecast renewable energy consumption with deep learning models. Sustain. Comput. Inform. Syst. **38**, 100863 (2023)
33. Edelmann, D., Móri, T.F., Székely, G.J.: On relationships between the pearson and the distance correlation coefficients. Statist. Probab. Lett. **169**, 1–6 (2021)
34. Mafarja, M., et al.: Binary dragonfly optimization for feature selection using time-varying transfer functions. Knowl.-Based Syst. **161**, 185–204 (2018)
35. Mafarja, M.M., Eleyan, D., Jaber, I., Hammouri, A., Mirjalili, S.: Binary dragonfly algorithm for feature selection. In: 2017 International Conference on New Trends in Computing Sciences (ICTCS), pp. 12–17 (2017)
36. Ye, M., Zhou, H., Yang, H., Hu, B., Wang, X.: Multi-strategy improved dung beetle optimization algorithm and its applications. Biomimetics **9**(5), 1–30 (2024)

37. Ali, M.H., Kamel, S., Hassan, M.H., Tostado-Véliz, M., Zawbaa, H.M.: An improved wild horse optimization algorithm for reliability based optimal DG planning of radial distribution networks. Energy Rep. **8**, 582–604 (2022)
38. Yang, T., Wang, H., Aziz, S., Jiang, H., Peng, J.: A novel method of wind speed prediction by peephole LSTM. In: 2018 International Conference on Power System Technology (POWERCON), pp. 364–369. IEEE (2018)
39. Souibgui, M.A., Kessentini, Y.: DE-GAN: a conditional generative adversarial network for document enhancement. IEEE Trans. Pattern Anal. Mach. Intell. **44**(3), 1180–1191 (2020)
40. Aamir, M., Mohd Nawi, N., Wahid, F., Mahdin, H.: A deep contractive autoencoder for solving multiclass classification problems. Evol. Intel. **14**, 1619–1633 (2021)