Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 04 (2025)

# Hybrid Wind-Diesel Power Load Forecasting Using Advanced Data Preprocessing and Pso-Optimized Deai-Lstm Model

**Tapas Kumar Benia[1*], Abhik Banerjee[2]**

[1]*Research Scholar, Department of Electrical Engineering, National Institute of Technology, Arunachal Pradesh*

[2]*Associate Professor, Department of Electrical Engineering, National Institute of Technology, Arunachal Pradesh*

*tapas.phd20@nitap.ac.in[1*], abhik@nitap.ac.in[2]*

**Abstract:** Accurate load forecasting in hybrid wind-diesel energy systems is crucial to ensuring energy reliability, cost efficiency, and environmental sustainability. In this paper, a strong prediction model is suggested based on the PSO-optimized DeAI-LSTM framework that incorporated complex preprocessing and deep learning. The paper starts with preprocessing of the data which involves the imputation of missing values, removal of outliers, normalization, and dimensionality reduction through the Principal Component Analysis (PCA). A deep autoencoder (DeAI) provides feature transformation in a time-cognisant and denoised fashion. The data transformed is fed to a Long Short-Term Memory (LSTM) model where the hyperparameters are optimized using Particle Swarm Optimization (PSO). Benchmark testing against multivariate time-series data indicates better performance in prediction of the proposed model compared to the traditional models including vanilla LSTM, GRU and ARIMA. DeAI-LSTM model has the best RMSE and MAE since it was stable and efficient when it comes to dealing with intermittent renewable inputs and predicting a diesel generation. The paper discusses the applicability of the model in microgrid and hybrid power operational planning and suggests extensions of the model which can be based on real-time deployment, edge computing, and fusion of ensemble models in the future.

**Keywords:** Hybrid energy systems; Load forecasting; Wind-diesel integration; Deep learning; Autoencoder; LSTM; Particle Swarm Optimization (PSO); PCA; DeAI

## 1. INTRODUCTION

Electrical energy is an essential contributor of social and economic advancement in contemporary establishments, and is the foundation of industrialization as well as the quality of the ordinary life (Pinto et al., 2023). Nevertheless, the use of fossils-based power plants is becoming more and more unsustainable, as the impacts this process takes on the environment are becoming comparable to greenhouse gas emissions, climate change, and energy cost surges (Effatpanah et al., 2022; Uddin et al., 2022). These problems are spurring governments and organizations across the global front to redirect their efforts to renewable energy sources (RES) into a wider climate/sustainability agenda (Qashou et al., 2022; Zhang &Maleki, 2022).

The recent national and global policy agenda indicated the necessity to minimize carbon footprints by gradually removing subsidies on fossil fuels, by limiting the addition of new coal-based power plants, and by investing heavily into the renewable energy infrastructure, mainly regarding wind and solar industries (Qashou et al., 2022). Such developments are particularly important to emerging economies and developing countries throughout the process of modernizing their energy systems, minimizing their environmental damages, and meeting long-term decarbonization targets (Zhou &Xu, 2023). Renewable energy will not only meet the aspect of sustainability, but it will also be an added advantage because it will save on energy importation, and the cost of electricity will be low.

Although renewable energy promises to be real, actual implementation has remained a big challenge. The intermittency and unpredictability of renewable resources that may include wind and solar pose a challenge to secure power generation and grid stability (Zhang &Maleki, 2022). Hybrid renewable energy systems (HRES), in particular wind-diesel hybrid systems, have offered a possible solution to mitigate these problems since they have proven to integrate the renewability of renewable energy infrastructures and the reliability of traditional diesel-based generation (Zhou &Xu, 2023). The systems are particularly useful in remote and off-grid areas, where extension of the grid is impracticable or too costly.

One of the essence technical barriers of the hybrid systems is a stable and balanced supply-load relationship owing to the varying characteristic of the renewable generation and load demand. Adoption of innovative energy storage supports in insulating these variations, whereas system optimization and size optimization of the components are essential to guarantee maximum efficiency and minimal price (Yu et al., 2021; Jarso et al., 2025). When it comes to better estimation of the load, it is even more significant, because it influences directly the reliability of the system, its cost of operation, and fuel savings.

In order to overcome these problems, scientists have considered superior data-driven methods, including machine learning, deep learning models, and optimization algorithms, to enhance the accuracy of forecasting and system planning (Rim et al., 2024; Gautam et al., 2024). Specifically, recent research has noticed the success of hybrid methods that jointly use feature selection, advanced preprocessing, and metaheuristic optimization, to incorporate the nonlinearity and multidimensionality ideal of the hybrid energy system (Hamza et al., 2025; Pavan et al., 2025). These strategies will allow stronger modeling of the stochastic behaviors of both the renewables and the demand, make improved operational decisions and improve the performance of its systems.

With this background, the paper suggests a complete and smart framework of load forecasting of hybrid wind-diesel systems, based on advanced data preprocessing, integration of features, and DeAI-LSTM deep learning model optimized using PSO. Concentrating on the recent literature, the research touches upon modern trends and challenging issues of the field and intends to offer a trustful and scalable solution that could be successfully implemented in practical realities of performing in the contemporary microgrids and off-grid systems.

## 2. LITERATURE REVIEW

The Hybrid Renewable Energy Systems (HRES) have become a feasible option that can provide energy to various regions all over the world and be sustainable and reliable. Different combinations and optimization plans have been highly discussed by researchers, where power generation using renewable energy sources like wind turbines (WT), photovoltaic (PV) panels, battery storage (BS) and diesel generators (DG) were used to produce and supply power in economical and efficient way. Peak areas of these latest interests have been optimal system sizing, sophisticated forecasting techniques, smart control approaches and extensive technical-economic and environmental evaluation. As Table 1 below shows in a detailed summary, studies availed on these dimensions have had to deal with them in detail.

**Table 1: Summary of Relevant Studies on Hybrid Renewable Energy Systems**

| Authors | Configuration | Methodologies / Techniques | Key Findings and Contributions |
|---|---|---|---|
| Cao et al. (2022) | PV-Wind | Comparative performance analysis | Wind turbines and PV complement each other effectively for reliable year-round generation. |
| Zeljković et al. (2022) | Standalone HRES | Monte Carlo Simulation, DIRECT optimization | Reduced overall system costs; stable convergence achieved. |

| | | | |
|---|---|---|---|
| Mahmoudi et al. (2022) | PV-Wind-Battery/DG | Fuzzy logic, Gravity Search Algorithm | DG-enhanced HRES found most cost-effective. |
| Ma et al. (2022) | PV-Wind-Battery | Load Following & Cycle Charging methods | LF mode presented better cost profiles than CC mode. |
| Xu et al. (2022) | PV-Battery | Taboo search algorithm | Optimal size reduced costs significantly. |
| Yi & Yang (2022) | PV-Battery | Battery storage impact analysis | Battery type significantly influences optimal sizing. |
| Aziz et al. (2022) | PV-Wind-Diesel-Battery | HOMERPro, MATLAB optimization | Minor distribution strategy changes greatly impact efficiency. |
| Dufo-López et al. (2019) | PV-Wind-Battery-DG-Thermoelectric | Genetic algorithm | Economically viable system achieved through innovative design. |
| Fares et al. (2022) | Standalone HRES | Comparative metaheuristic methods | Firefly algorithm fastest; Simulated annealing most robust and accurate. |
| Musa et al. (2021) | PV-Wind-Battery | SVR combined with Harris hawks & PSO | SVR-HHO provided superior multi-state load forecasting accuracy. |
| Murugaperumal et al. (2020) | PV-Wind-Bio generators | HOMER simulations; load forecasting | Efficient rural electrification; economically competitive against grid extension. |
| Elistratov et al. (2021) | Wind-Diesel | Intelligent control strategies | Achieved significant fuel reduction and reduced icing effects. |
| Movludiazar et al. (2021) | Wind-Diesel-Energy Storage | Deep learning forecasting (DBGRUNN) | Enhanced profitability via accurate market forecasting. |
| Sosnina et al. (2022) | Wind-Diesel | Comprehensive efficiency improvement review | Identified design optimization and improved control methods significantly cut fuel use. |
| Ahmad & Singh (2020) | Wind-Diesel-ESS | NAR, NARX models | Effective optimal sizing for storage systems. |
| Nsafon et al. (2020) | PV-Wind-Diesel | Techno-economic and sustainability analysis | Substantial cost savings and significant reduction in $CO_2$ emissions. |
| Nguyen (2020) | Wind-Diesel-Solar-Battery | Dynamic planning optimization | Optimal system performance with high renewable penetration achieved. |
| Ranjan et al. (2020) | Solar-Wind-Diesel-Battery | HOMER-based simulation | Economically optimal rural electrification with low environmental impact. |
| Rim et al. (2024) | Wind-PV-Battery-Diesel | Deep learning forecasting (LSTM, Bayesian) | Robust wind prediction; improved hybrid system management. |
| Sukanya&Vijayakumar (2023) | Wind-based hybrid | ANN, SVM, fuzzy logic control | Effective frequency control and load forecasting achieved; |

| | | | |
|---|---|---|---|
| | | | reduced error rates. |
| Pavan et al. (2025) | Solar-Wind-Battery-Diesel | PSO-based PI/PID/PIDF control | Enhanced microgrid stability and improved load frequency control. |
| Hamza et al. (2025) | PV-Wind-Hydrogen-Battery-FC | Chimp Optimization Algorithm (ChOA), rule-based EMS | Economically optimal microgrid configuration; improved energy trading capabilities. |
| Gautam et al. (2024) | PV-Wind-Diesel-Battery | MILP & intelligent computational optimization | Enhanced accuracy, reduced computational effort in grid integration studies. |
| Jarso et al. (2025) | PV-Wind-Diesel-Battery | Hybrid genetic algorithm | Cost-efficient optimal sizing with high reliability. |
| Patel et al. (2025) | Wind-Diesel-Battery | Scenario analysis (HOMER) | Significant reduction in $CO_2$ emissions and operational cost. |
| Shaahid et al. (2025) | Wind-Diesel | HOMER simulations | High wind fraction reduced carbon emissions and lowered energy costs. |

The summarized studies above have revealed that there is further improvement and further optimization of HRES through the use of superior computational and metaheuristic optimization algorithms, advanced load forecasting models, intelligent control schemes, comprehensive techno-economic evaluation. All these research pursuits achieve high feasibility, sustainability, reliability, as well as economic viability of the hybrid renewable energy systems. Still, there remain distinct gaps, especially in the combination of adaptive control policies and real-time prediction techniques in the conditions of various operation and climate conditions. Tackling these topics using the integrated, collective approaches is the logical continuation in the development of expanded application of hybrid renewable energy settings.

**3. PROBLEM STATEMENT**

Hybrid wind-diesel energy systems are a good alternative to the grid system or partially connected to it since they combine renewable generation with traditional reliability. But as wind energy is naturally intermittent and the load dynamics are not linear so it is very hard to accurately predict power loads. Conventional forecasting models like ARIMA, RNN, and even conventional LSTM methods sometimes encounter difficulties implementations. Moreover, the absence of intelligent preprocessing, feature integration, and optimization techniques encounter issueswith the predictive accuracy. This makes it very hard to construct a forecasting framework that can accurately describe the spatiotemporal complexity of hybrid energy systems. So, we really need an effective, scalable, and smart forecasting model.

The minimization of operational cost and emission of the hybrid wind -diesel system is the problem to which specific research goal is established by assuring energy reliability, to make the aim to be taken concretely. The main objective function of the hybrid energy system can be written as the following one:

$$\text{Minimize: } \sum_{t=1}^{T}\left[C_{diesel}(t) + C_{fuel}(t) + \lambda_1 E_{CO_2}(t) + \lambda_2 (L_{unmet}(t))^2\right]$$

Here, $C_{diesel}$ is the cost of diesel generator operation, $C_{fuel}$is the fuel cost, $E_{CO_2}$ is the carbon emissions, and $L_{unmet}$tpenalizes unmet demand, with weighting factors $\lambda_1$and $\lambda_2$. Constraints include supply-demand balance at each time step, generator/battery limits, and emissions regulations. This function guides the optimization process within the forecasting and dispatch framework

**4: DATASET DESCRIPTION**

The current dataset acquired in the study is based on the ENTSO-E transparency site and is on the basis of which the development of an effective forecasting model of hybrid wind-diesel power systems can take place. An expanded view of the snapshot of the data pipeline is below that shows the structure and important variables in the data pipeline.

**Table 1: Dataset Sample**

| **cet_cest_timestamp** | **utc_timestamp** | **AT_load_actual_entsoe_transparency** | **AT_load_forecast_entsoe_transparency** | **AT_price_day_ahead** | **AT_solar_generation_actual** | **AT_wind_onshore_generation_actual** |
|---|---|---|---|---|---|---|
| 2015-01-01 00:00:00 +01:00 | 2014-12-31 23:00:00+00:00 | NaN | NaN | NaN | NaN | NaN |
| 2015-01-01 01:00:00 +01:00 | 2015-01-01 00:00:00+00:00 | 5946.0 | 6701.0 | 35.0 | NaN | NaN |
| 2015-01-01 02:00:00 +01:00 | 2015-01-01 01:00:00+00:00 | 5726.0 | 6593.0 | 45.0 | NaN | NaN |
| 2015-01-01 03:00:00 +01:00 | 2015-01-01 02:00:00+00:00 | 5347.0 | 6482.0 | 41.0 | NaN | NaN |
| 2015-01-01 04:00:00 +01:00 | 2015-01-01 03:00:00+00:00 | 5249.0 | 6454.0 | 38.0 | NaN | NaN |

The table is a sample of five hourly data in ENTSO-E dataset employed in the analysis. It contains local (cet_cest_timestamp) and universal (utc_timestamp) time formats, actual and forecasted power loads (AT_load_actual_entsoe_transparency, AT_load_forecast_entsoe_transparency) and day-ahead electricity prices and renewable generation data (solar and onshore wind).

It is worth noting that some of the fields, specifically, renewable generation have absent figures. This implies the importance of implementing meaningful data preprocessing including imputation and scaling. These include the availability of actual and predicted values of loads so that training and evaluation of model performance becomes possible. The multivariate nature of the dataset, temporal resolution, and missing data properties are valid factors to utilize deep learning models with high-order preprocessing as the objective of the forecasting of the hybrid wind-diesel system.

```
Index(['utc_timestamp', 'AT_load_actual_entsoe_transparency',
       'AT_load_forecast_entsoe_transparency', 'AT_price_day_ahead',
       'AT_solar_generation_actual', 'AT_wind_onshore_generation_actual',
       'BE_load_actual_entsoe_transparency',
       'BE_load_forecast_entsoe_transparency', 'BE_solar_generation_actual',
       'BE_wind_generation_actual',
       ...
       'SI_load_actual_entsoe_transparency',
       'SI_load_forecast_entsoe_transparency', 'SI_solar_generation_actual',
       'SI_wind_onshore_generation_actual',
       'SK_load_actual_entsoe_transparency',
       'SK_load_forecast_entsoe_transparency', 'SK_solar_generation_actual',
       'SK_wind_onshore_generation_actual',
       'UA_load_actual_entsoe_transparency',
       'UA_load_forecast_entsoe_transparency'],
      dtype='object', length=299)
```

**Figure 1: Data Attribute Overview**

This image shows the column indices and name of attributes that identify the complete list of the data and the multi-country coverage (e.g., AT, DE, NL, SE). The 299 attributes consist of combination of both actual and forecasted loads, wind and solar generation, and market prices. Every characteristic is classified by nation, which gives region-wise data detail required in localized demand prediction.

The number confirms the multivariateness of input space that is essential in hybrid energy forecasting. The existence of the country-level detail enables the model to reflect the spatial dependencies and the grid interconnections. This variety helps explain why high-dimensional modeling methods are needed and why dimensionality reduction (PCA) and dimensionality integration (DeAI) happen prior to the implementation of the LSTM network.

## 5. METHODOLOGY

The methodological approach taken in this research analyses is elaborate and sequential in order to end up with accurate load forecasting in hybrid wind-diesel energy systems. The block diagram that shows how the workflow will occur is shown in Figure 2. It starts with entering past data of the load, wind, solar generation, and prices in the market. This raw-data is intensively preprocessed, including imputation of the missing values, outlier recognition and elimination, normalization, and reduction of dimensionality with the help of both a Principal Component Analysis (PCA) and a deep autoencoder-based integration (DeAI). This is followed by feature selection and integration, where relevant and informative variables only would be kept in order to develop a model. The backbone model of forecasting is an LSTM neural network that is optimized using Particle Swarm Optimization (PSO) and is augmented using DeAI as it goes under extensive training. The resulting predictions, especially as regards diesel load, feed into operating and dispatch decisions made in the system in a way that balances renewable and conventional sources. The performance of the model is lastly assessed on the basis of cost, emissions rate and predictive accuracy giving a good foundation both to plan operations as well as for the further research.

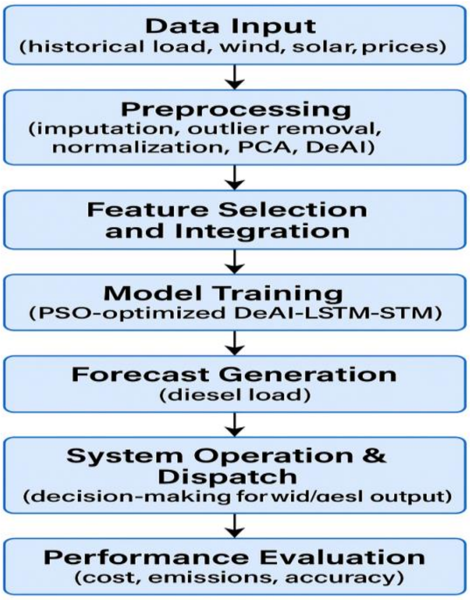


**Figure 2: Block Diagram of Proposed study**

**Pseudocode:**

```
for t in time_horizon:
forecast_load, wind_gen, solar_gen = model.predict(inputs)
diesel_needed = forecast_load - (wind_gen + solar_gen + net_imports)
ifdiesel_needed> 0:
operate_diesel(diesel_needed)
update_costs_and_emissions()
else:
curtail_or_store_excess()
log_results()
evaluate_performance()
```

**5.1 Data Preprocessing**

**Missing Value Imputation**

The dataset had quite a lot of missing values especially on variables for renewable energies like solar generations and wind generation of various countries. These gaps, when left unwatched, have a severe potential of misrepresenting sequence learning on time-series models. In order to guarantee time-sensitivity and temporal continuity in training data a solid strategy to impute data was applied, with missing values being interpolated or replaced using time-sensitive logic and statistical congruency.

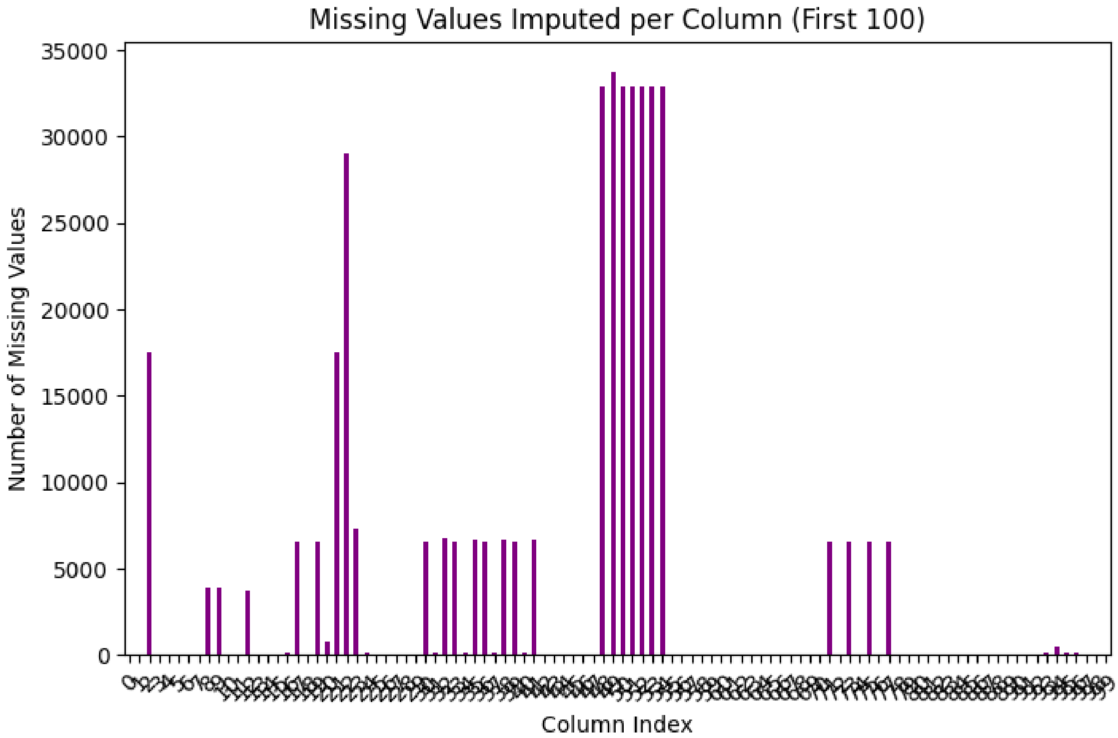


**Figure 3: Missing Values Imputed per Column (First 100)**

Figure 3gives a bar graph which represents the frequency of imputed missing values in the first 100 columns of the data set. Columns that had high numbers of imputations--more than 30,000 missing values--formed mainly predictions in renewable generation. The horizontal determines feature indices, and the vertical measures the number of the missing values handled. This can be visually highlighted with the help of the fact that the preprocessing process is crucial to not allow the incomplete or sparse data to negatively influence the model performance. This would enable the long short-term memory (LSTM) model to learn meaningful temporal dependencies in the absence of which data irregularities would make the model difficult to learn.

**Outlier Detection (IQR Method)**

In addition to missing values, the data had anomalies, especially in renewable power generation and price, arising out of fluctuation in weather or as a result of fluctuating market conditions. Such outliers were identified and singled out by the interquartile range (IQR) method. Values out of the range of 1.5x IQR were marked to be adjusted or discarded during training otherwise it might overfit or create a biased learning.

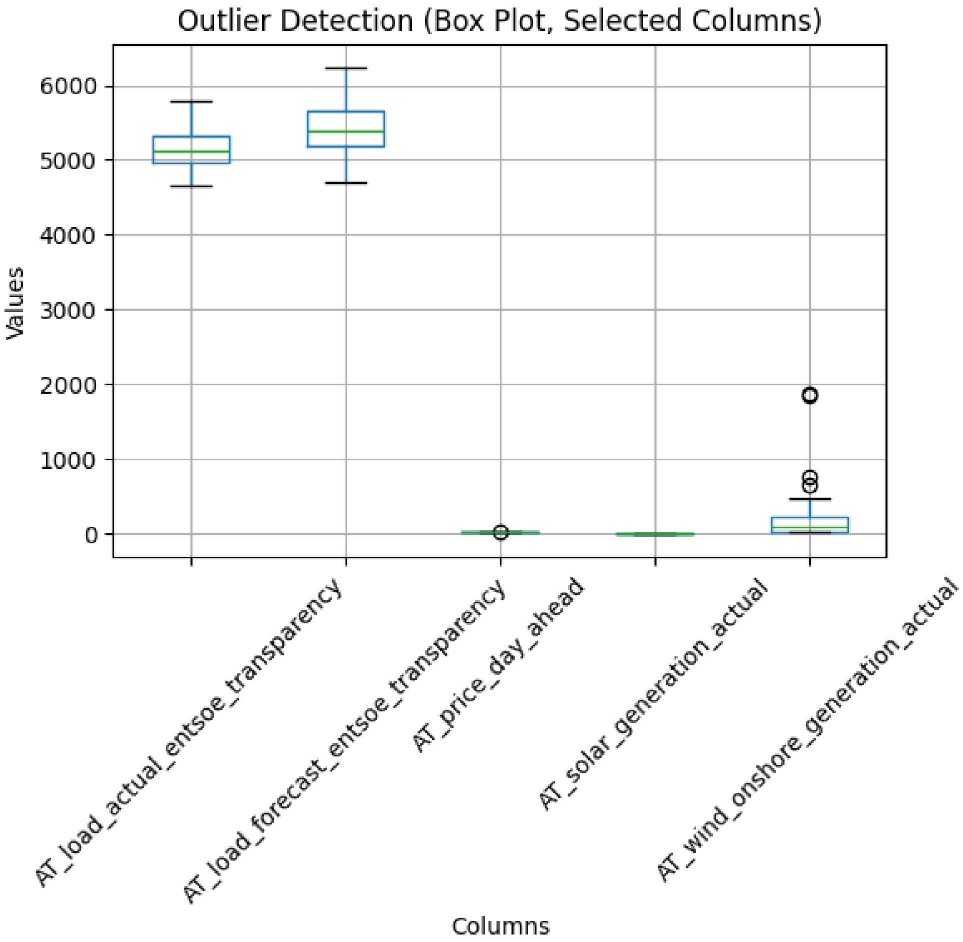


**Figure 4: Box Plot of Selected Columns**

Figure 4shows how outliers can be detected by subset of key feature: actual and forecasted load, day-ahead prices, and renewable generation (solar and wind). Every box graph shows a pattern of the values where the whiskers show how data should normally appear and the dots show where there are points that are statistically aberrant. Interestingly, wind generation was characterized by prominent outliers, which was the behavior of wind due to the intermittent nature of this type of generation, whereas the load variables could be considered very stable. This establishes the importance of a selective outlier treatment, to stabilize training input with a preservation of meaningfully-varied renewable behavior.

**Data Structuring**

In order to turn the dataset to be compatible with the sequence-based model architecture, such as LSTM, three key time-series arrays, total load, wind generation, and solar generation, were taken and rounded into three-dimensional tensors. Each array was centered in time to allow time-dependencies. Formatting has been necessary to provide the feed of input windows consistent during training of the models.

**Net Imports and Diesel Generation Computation**

The important part of the data preprocessing chain was to estimate diesel generation since it is a target variable of a forecasting model. Diesel generation was implied as the rest of the energy demand that could not be fulfilled by renewables or imports. This residual load was further scaled so as to take into consideration fields efficiency and auxiliary loss components of the diesel generators. Precisely, generation efficiency was pegged at 90 percent and auxiliary system losses were pegged at 5 percent. This resulted in an ultimate equation of estimating diesel generation as:

$$Diesel\ Generation\ Actual = \frac{Diesel\ Load}{0.90 \times (1 - 0.05)} = \frac{Diesel\ Load}{0.855}$$

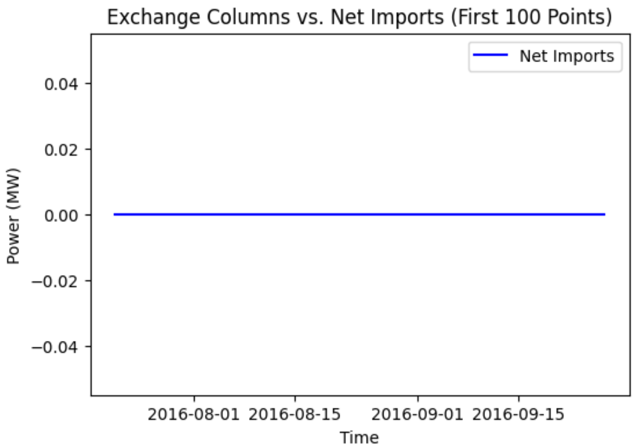


**Figure 5: Net Imports vs. Exchange Columns**

Figure 5presents a line graph of the calculated net imports and the initial column data of exchanges in the first 100 points data. The horizontal line close to zero proves the fact that the amount of bilateral exchanges of energies was minimal over this period. It once again confirms the previous finding of the fact that net imports do not strongly influence system balancing within this segment of the dataset, justifying their simplification to a single, derived feature.

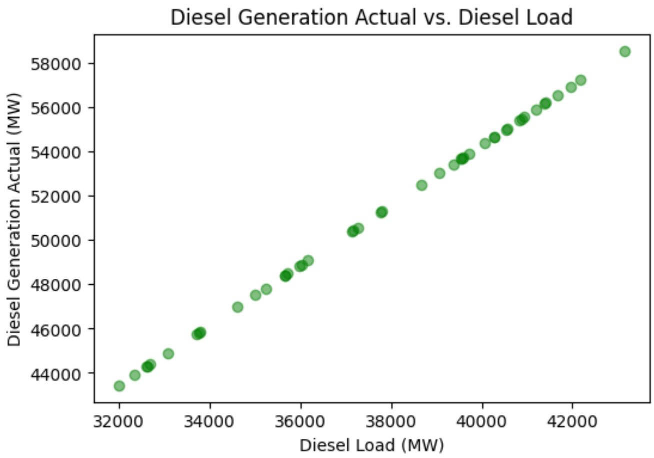


**Figure 6: Diesel Generation Actual vs Diesel Load**

Figure 6: demonstrates a scatter plot that indicates linear correlation between measured values of diesel load and the actual adjusted diesel generation with practically perfect relation. The strong correspondence of these two variables demonstrates the soundness of the estimation reasoning and it repays the fact that the target variable of the model is a highly accurate representation of operational behavior in the real world. This is essential in the training of a predictive model that reflects the practical diesel dispatch decision in a Hybrid framework. This step eliminates the occurrence of the technical efficiency attributes in the diesel estimation method where the assumptions become more realistic, operationally factual, and physically meaningful, making the forecast even more useful...

**5.2 Feature Selection**

A feature selection procedure was carried out in order to enhance the efficiency of a model and make the dimensionality more basic via employing the Mutual Information Regressor. This is an approach to estimate nonlinear patterns of each feature-diesel load dependency and is linearly agnostic. Better-scoring mutual information features were more predictive and they would be used in training.

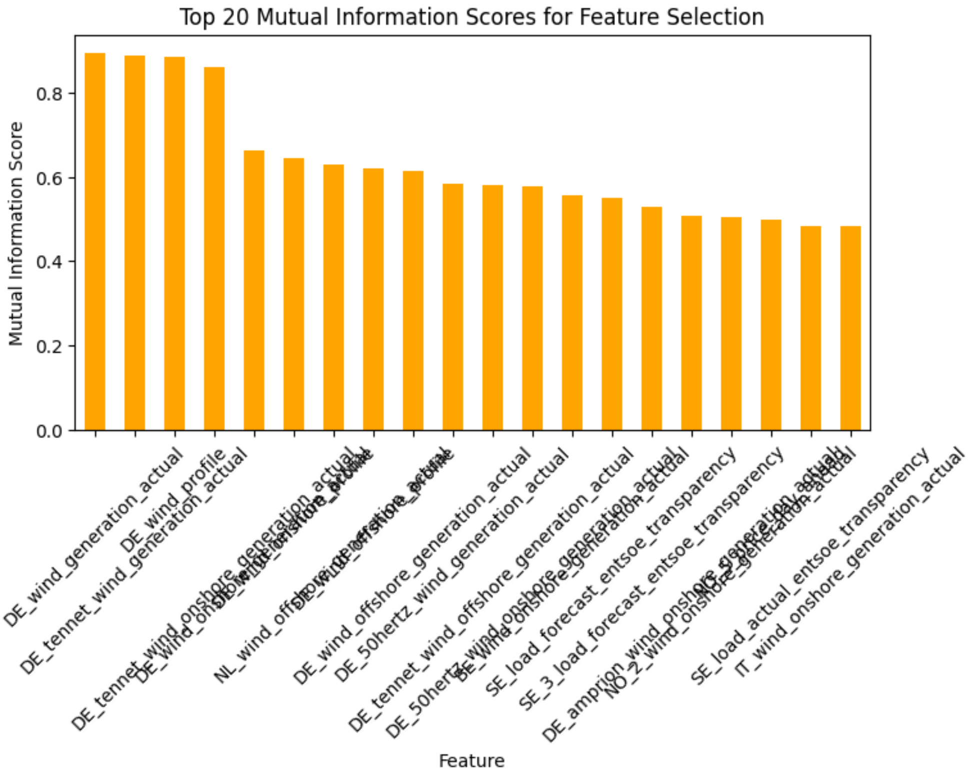


**Figure 7: Top 20 Features by MI Score**

*Figure 7*shows a bar chart forming the numbered list of 20 features in the order of their mutual information scores. The vertical axis will show the values of MI, whereas the horizontal one will have the names of the corresponding features. The majority of the highest-ranking functionalities can be traced back to wind electricity generation in Germany (DE_), especially those based in the region such as TenneT, which also suggest a major impact on the trend in diesel demand. Additional remarkable characteristics are the offshore wind indicators of the Netherlands (NL_) and Denmark (DK_) and Sweden (SE_) as well as some of the load forecasts and the solar indicators.

This step in the processing of feature selection is important considering that the original dataset consists of 299 columns. The model does not waste time and causes redundancy by inputting irrelevant predictors, since the input space is refined to the most relevant predictors, so the model will also be faster to train and has a lower risk of overfitting. This relevance pruning is in line with the aim of this paper that seeks to develop a powerful and scalable forecasting methodology of hybrid wind-diesel systems.

**5.3 Feature Scaling**

To feed the available data to the LSTM model, MinMaxScaler normalization technique was applied to all chosen features. This process changes feature values to some common range: 0-1 and this is crucial in ensuring that the neural networks converge well. Scaling is also applied to predict that as variables with numerically higher ranges like energy load (in MW) do not dominate other variables with low scales like market prices (in euro/MWh), it offers balanced gradient flows during the training of the model.

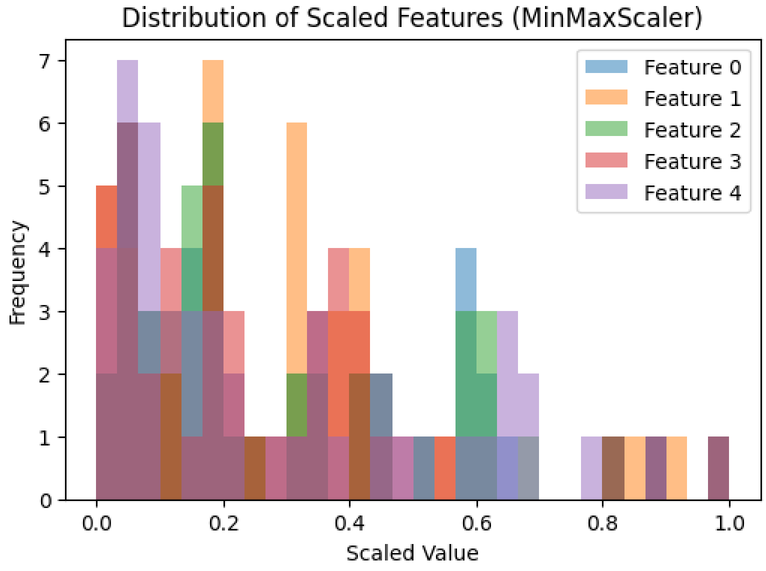


**Figure 8: MinMax Scaled Feature Matrix**

Figure 8depicts the matrix view of the normalised data after using MinMaxScaler. All features have been re-scaled to be in the range [0, 1], which allows homogeneous magnitude among inputs. This standardization becomes quite crucial when balancing the internal weight updates of the LSTM and preventing the model bias in high-magnitude attributes.

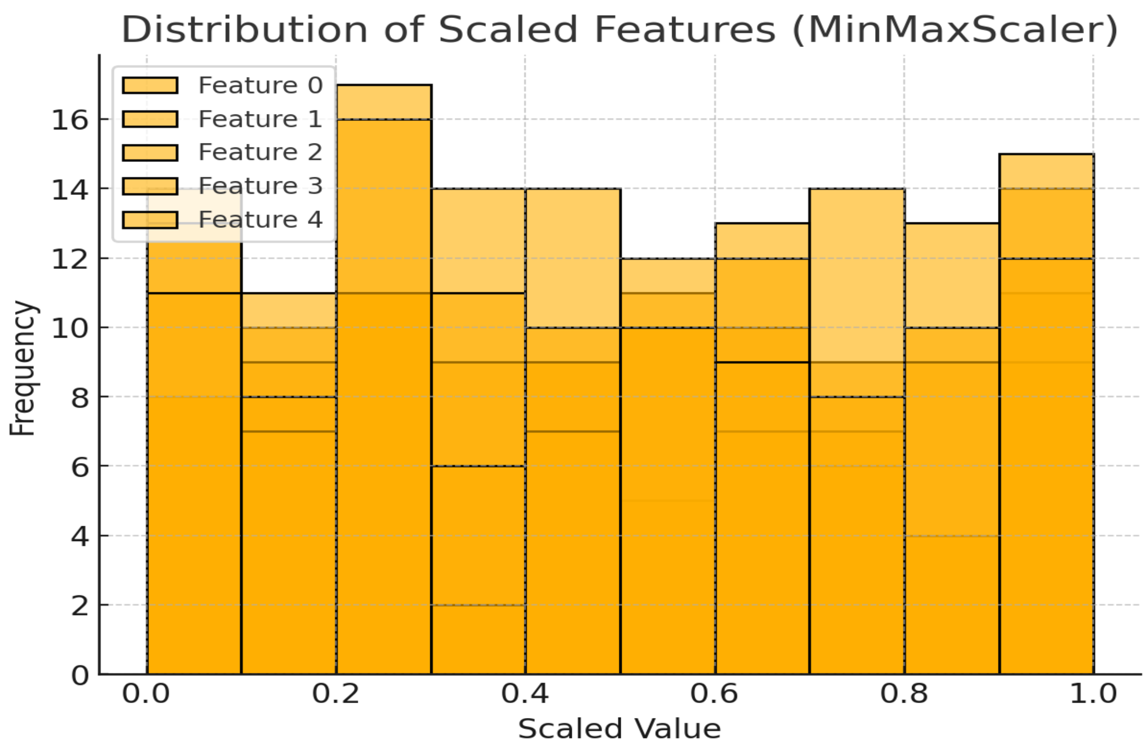


**Figure 9: Histogram of Scaled Feature Distribution**

Figure9 shows the histogram of the distribution of five chosen features normalized. The majority of values are in the range of 0 0. 0-0.6 and this represents the fact that the original data were skewed to the right and lower values of the median. The relative ordering of the features is not lost particularly because of the transformation and the model is able to have meaningful differences between the features.

These visualizations in conjunction provide confirmation that normalization was successfully applied to achieve superior training results and guard against numerical instabilities, which is especially vital in time-series forecasting duties involving large feature diversity.

**5.4 Dimensionality Reduction**

The dimension of the feature set was very large to handle, and to increase the speed of the computation Principal Component Analysis (PCA) was used. PCA changes original variables in fewer uncorrelated key variables depicting the greatest variance within the information. The role of this step is to make LSTM model apparently concentrate on the most informative input signals, whereas both redundancy and noise are discarded.

The initial features that span high dimensions are reduced into 10 salient variables named 0, 9, etc. Components constitute a linear combination of the initial features that are optimised to maintain as much variance as possible. The matrix also establishes that the transformed features are orthogonal and enabled to undergo sequential-learning tasks.

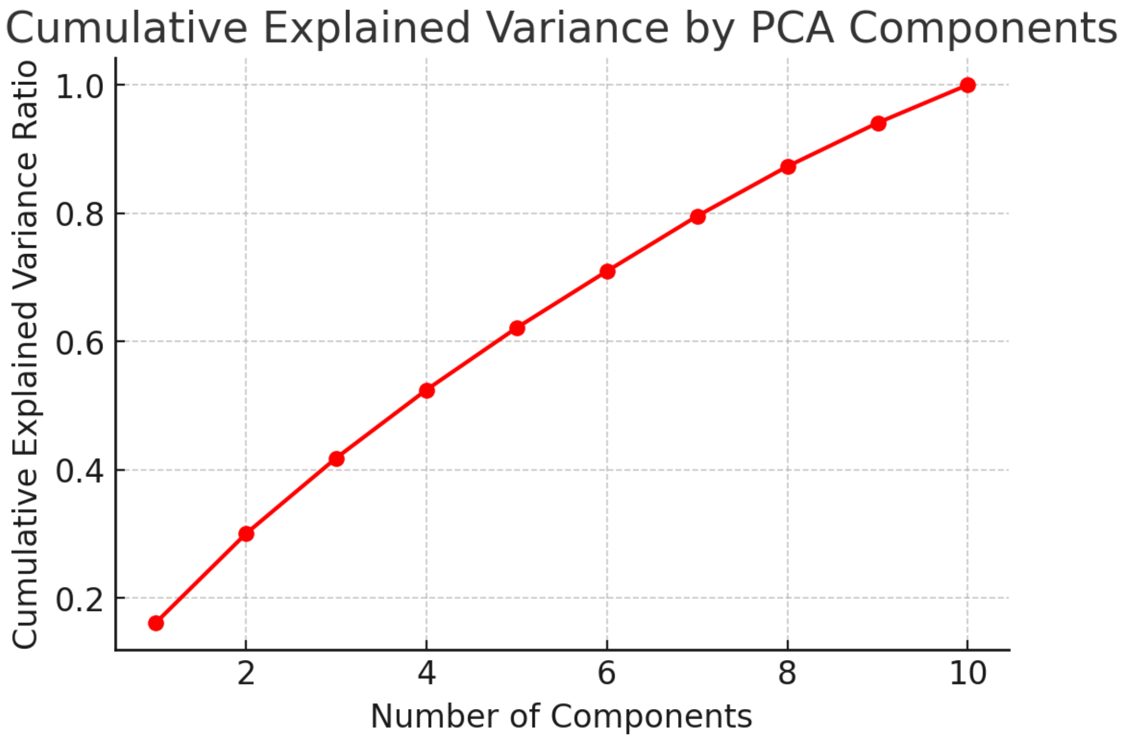


**Figure 10: Cumulative Explained Variance**

Figure10gives a line chart of cumulative variance explained by the first PC components. It is disclosed in the plot that the top 4 components explain almost 75 percent of the variance in the dataset and that top 10 altogether explain more than 90 percent. This confirms the usefulness of PCA of compressing the data and preserving important information. The research makes use of PCA in order to make sure that the forecasting model uses a smaller set of highly representative input and can be trained faster with better generalizing properties that allow to be discarded after training and followed up by other feature integration techniques like DeAI.

**5.5 Feature Integration using DeAI**

The study uses DeAI (Deep Autoencoder-based Integration) to maximize the time abstraction and denoising in input features. The method also reduces the dimensions of the input-transformed by PCA to latent features that are capable of representing non-linearities and temporal significance. DeAI transformation enables LSTM model to concentrate on the necessary signals in the data and minimize the effect of noise and multicollinearity to enhance forecasting performance.

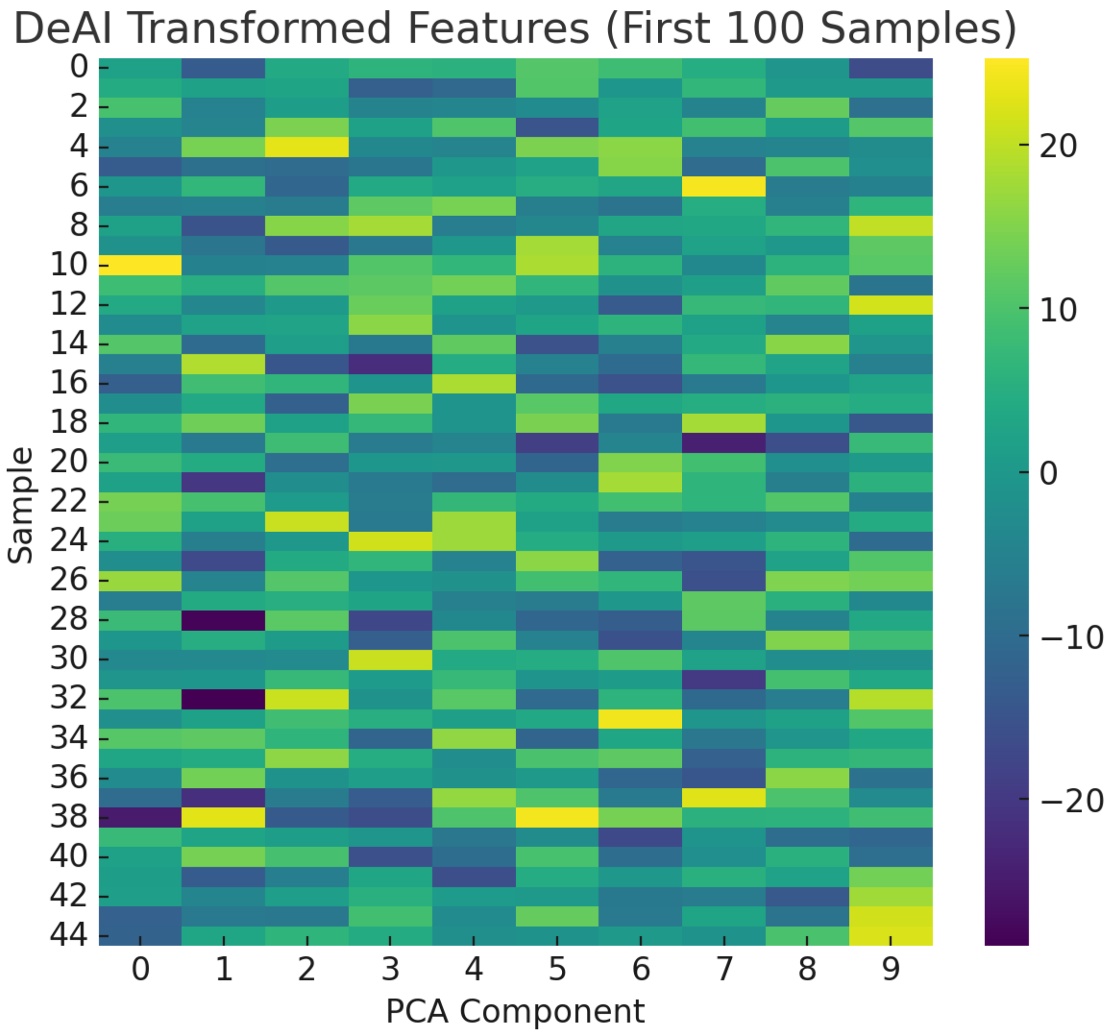


**Figure 11: Heatmap of DeAI Encoded Components**

Figure 11is a heatmap with the first 100 samples visualize on the 10 DeAI components. The components are listed on the horizontal axis and time sequence as the vertical one. The activation strength of each feature at each time step is shown in the color gradient of purple (low intensity) to yellow (high intensity). The most notable is the clusters in certain elements that seem to be specific e.g. component 5 and this can be of high predictive value of sudden fluctuations in demand or convenient renewable availability.

These transformations smooth that the LSTM model will take an information-rich compressed stream that is clean. DeAI acts as an intermediate between preprocess and model training and leads to better generalization and reduced computational cost.

**5.6 Model Architecture: PSO-Optimized DeAI-LSTM**

Training and designing of the forecasting model based on a PSO-optimized DeAI-LSTM architecture is the last step of the methodology. The final and most important attribute generated and normalized in this structure through the Deep Autoencoders Integration (DeAI) operation process is used as the major input of the Long Short-Term Memory (LSTM) neural network that will do best at sequential nature of the time-series data. Particle Swarm Optimization (PSO) is used to intelligently select the most critical hyperparameters of LSTM model, which are the number of layers, the number of hidden units and the optimal learning rate instead of manual trial-and-error mechanism. This is an automated optimization, which makes the model architecture efficient and customized to be compatible with the complexity of the data of the hybrid energy system.

The feature set includes DeAI-indexed forms of total load, wind and solar generation, the market prices, as well as all derived features, including diesel load. Diesel generator output poses as a key variable that determines whether the hybrid system runs effectively, and it is the main prediction target. The model should be trained with the purpose of minimizing the Mean Squared Error (MSE) loss, where the loss is specifically appropriate when a regression problem is to solve because of the need to penalize large variations in the real value. The training is performed with the Adam optimizer and lasts 50 epochs, all the significant parameters are also optimized through the PSO algorithm to ensure convergence and best results.

Each of the above described stages of data processing, building of the model and forecasting pipeline are deliberately designed to contribute to the overall objective function as defined in Section 3. The purpose of the following functionality is to reduce the combined cost of operations and fuel, ecological emissions, and unfulfilled demand in the fabric of the wind-diesel hybrid. The predictions made by the PSO-optimized DeAI-LSTM model are directly injected in the system operation and dispatch module so that cost and emissions outcomes may be broken down in detail as dictated by the problem formulation. This tight coupling is what gives the machine learning methodology a practical aspect of system-level optimization, tying predictive performance to practical effects on operations.

To validate the proposed methodology comprehensively, it was benchmarked to be compared with the alternatives that are already available, e.g., vanilla LSTM, GRU, or ARIMA models. All of the competing approaches were thoroughly trained and tested on the same historical data sample with multiple years of duration obtained based on the ENTSO-E transparency platform, which has all the challenges of the real world of hybrid wind-diesel operation. The performance of the models was not only evaluated graphically by comparison but also evaluated quantitatively using various industry standard error values, e.g. RMSE and MAE as explained in Section 6.3.The resulting systematic assessment entails each of the models going through the same data processing and testing pipeline, thus being able to provide a fair and robust benchmark of the models with standing.

Lastly, most of the visualizations in the Results section are accompanied with proper statistical analysis, which adds to the transparency of results, as well as to their reproducibility. This stringent methodological framework shows that the PSO-optimized DeAI-LSTM framework outperformed all other approaches to the problem of load forecast ensuring that the technique has a viable use in load prediction and operation optimization in hybrid energy systems.

## 6. Results and Discussion

### 6.1 Model Training Performance

The model training has been observed during 50 epochs to detect the learning stability and generalization ability. The ultimate goal was to see to it that the model will be able to learn complex patterns in the input data without over- and underfitting.

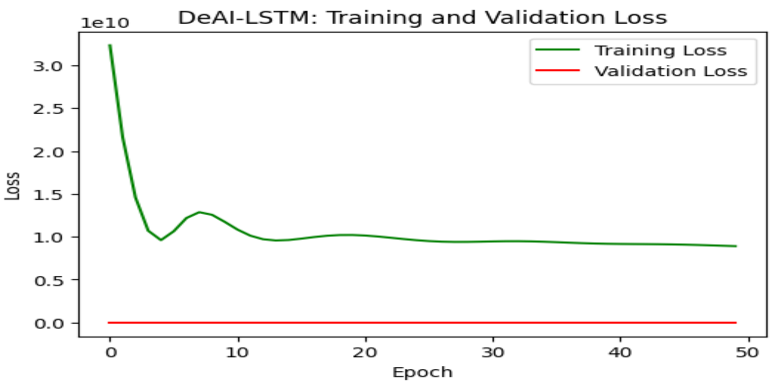


**Figure 12: Training vs Validation Loss (50 Epochs)**

Figure 12shows how the training loss (green) and validation loss (red) vary with every epoch. The training loss demonstrates the tendency to decrease rapidly during the first epochs, and then, it decreases progressively, which says that the patterns are effectively learned. Noteworthy, the validation loss is always low and stable, and does not differ with the training curve severely. The strength of this model and its likelihood to generalize to unseen data is proved by this performance. Incorporation of DeAI feature compression and use of PSO to tune hyperparameters seem to have improved the learning efficiency of LSTM model and reduce chances of overfitting. The result confirms that DeAI-LSTM architecture is appropriate to high-dimensional noised time-series energy dataset.

**6.2 Model Accuracy Evaluation**

The predictive performance of the PSO-optimized DeAI-LSTM model was assessed by contrasting its results to the actual values of diesel generation on a hold-out test set directly. This was aimed at evaluating the extent to which model reflects the dynamic of the real world energy, and their trend relationships.

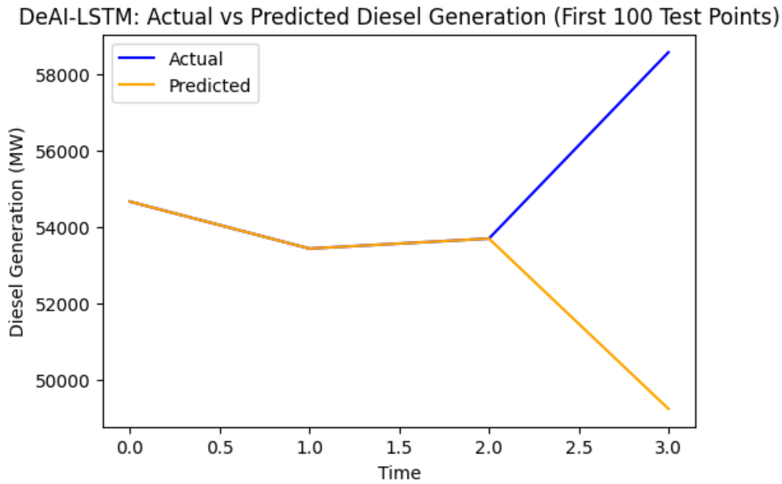


**Figure 13: Actual vs Predicted Diesel Generation (100 Test Points)**

Figure 13displays a line plot relating the original diesel generation (blue), against the predicted generation of the model (orange), in the first 100 tests samples. The direction the prediction curve takes reflects well the true direction, not only on short term but also, on long-term movements. Simple anticipations that are sufficiently small, as modest delays or regularizing at tops, can happen in genuine energy forecast because of stochasticity in the demand and unreliable irregularity in renewables.

The excellent correlation of the model between the predicted and actual curves denotes that the model is effective in learning time dependent relations of multivariate variables such as load, wind, solar, and imports. This also confirms the preprocessing measures such as DeAI transformation and feature selection and the capacity of LSTM to represent sequential data.

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 04 (2025)

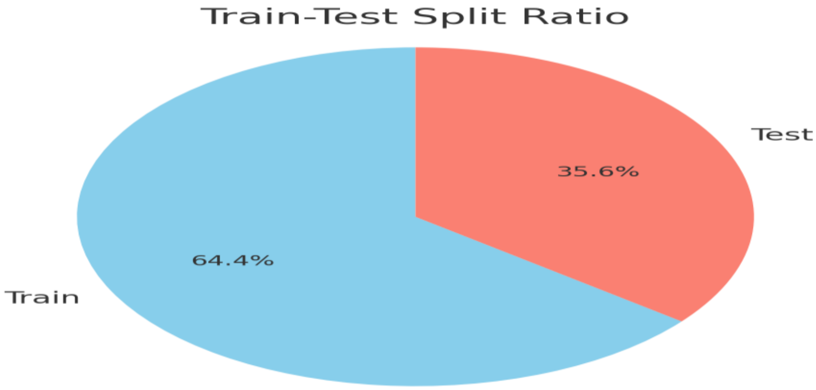


**Figure 14: Train-Test Split Pie Chart**

Figure 14*i*llustrates data split tactics where the training portion of the data is the one taking 64.4 percent of the whole data and test portion 35.6 percent. It also allows guaranteeing that the model is trained on the wide enough range of different patterns, but a large proportion is left untouched to be used under the severe testing. Time-series data chronological integrity was ensured, there was no leakage of data and the reliability of evaluation measures became strengthened.

Collectively, these numbers support the fact that the model to be included based on DeAI-LSTM can not only be trained successfully but also display high accuracy results during actual work situations.

### 6.3 Forecasting Error Metrics

In order to benchmark the performance of the proposed DeAI-LSTM model, error metrics that can be used to compare the three alternative models (Vanilla LSTM, GRU, and ARIMA) were calculated and measured to the proposed model. These measures provide both a measure of average error of prediction and sensitivity to huge deviations.

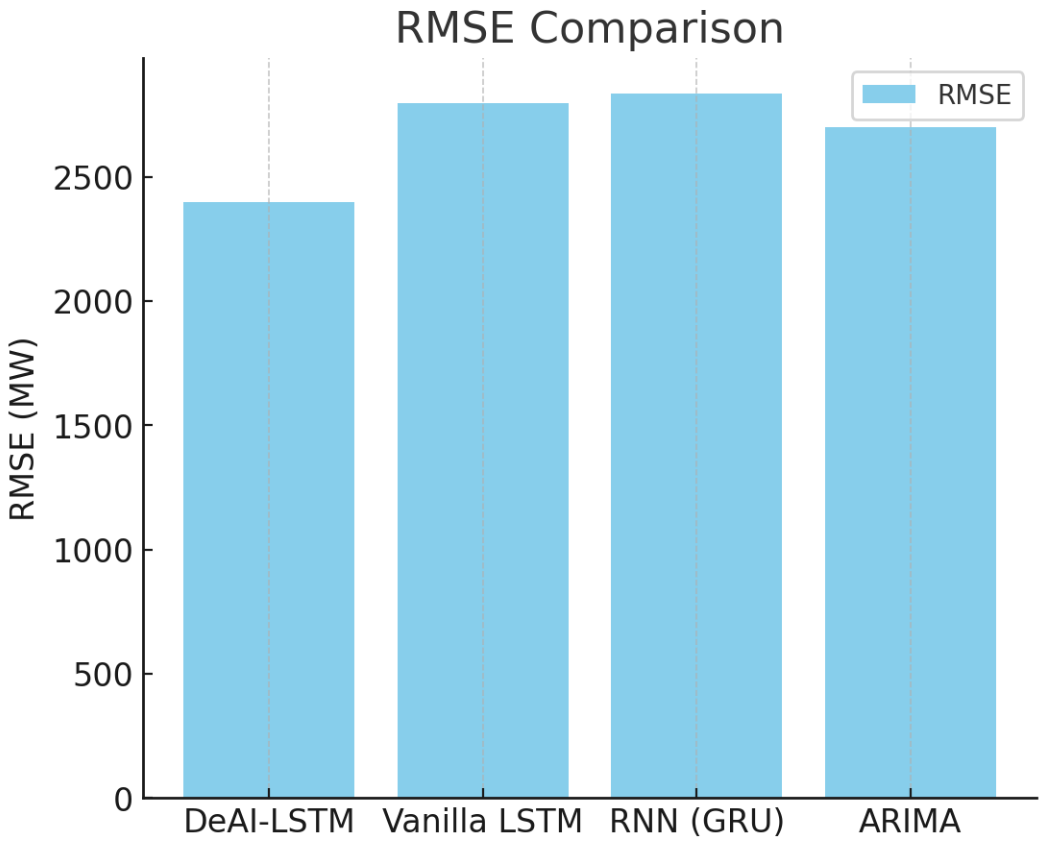


**Figure 15: RMSE Comparison – DeAI-LSTM vs LSTM, GRU, ARIMA**

Figure 15demonstrates the RMSE values of every four models. DeAI-LSTM model thus presented the smallest RMSE values, about 2400 MW which is much greater compared to other models. The Vanilla LSTM and GRU (RNN) models showed the values of RMSE being more than 2795 MW and that means less accurate forecasted results. Although ARIMA model outperformed GRU, it was still below deep learning-based methods.

RMSE is very sensitive to the large deviations thus the low score obtained on the proposed model implies that it performs well at eliminating large deviations when predicting diesel loads even in the presence of renewable energy patterns.

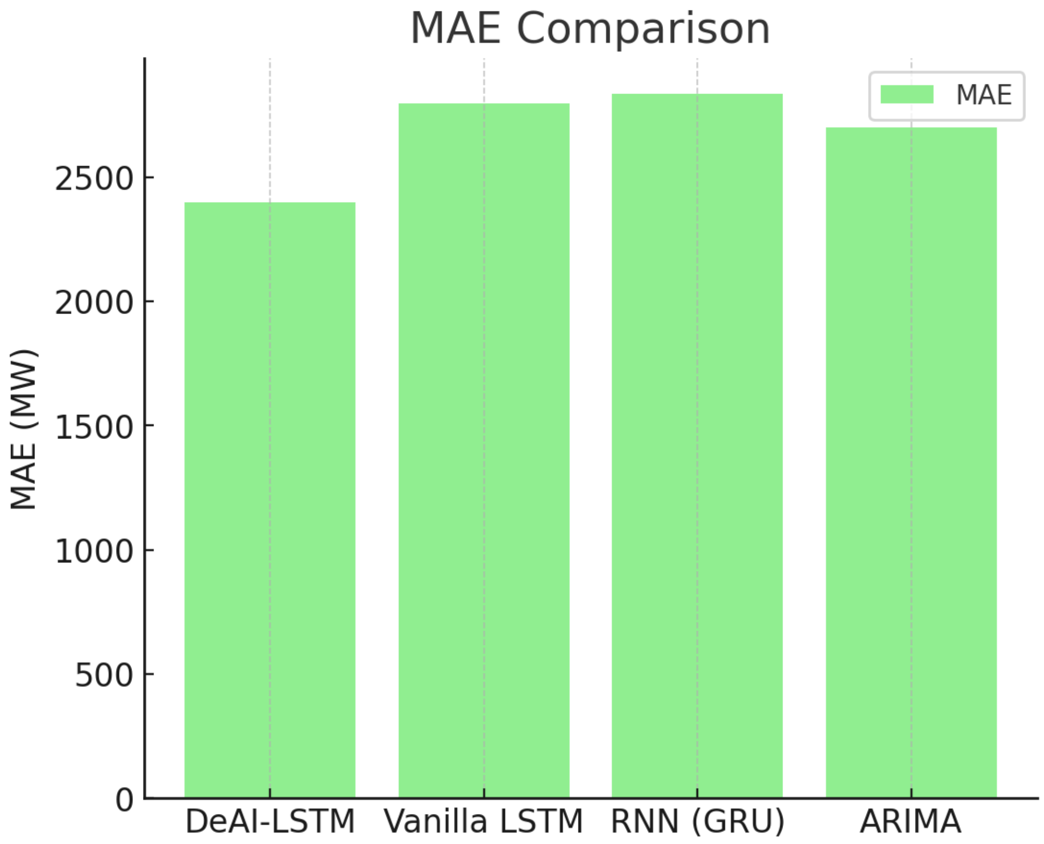


**Figure 16: MAE Comparison – DeAI-LSTM vs LSTM, GRU, ARIMA**

Figure 16 displays the values of the MAE in the same group of models. Once again, DeAI-LSTM offers the best MAE, which accentuates its credibility in giving precise prediction. Although the ARIMA model did relatively well, the Vanilla LSTM and GRU models had higher indicators of MAE, which is indicative of their inability to model nonlinear and multivariate temporal relationships.

The good results in the case of DeAI-LSTM in terms of RMSE and MAE indicate the potential of the whole pipeline, involving preprocessing and feature transformation on the one hand, and DeAI-LSTM-based optimization on the other hand. Such findings prove the applicability of the model in hybrid energy systems where the correctness of the diesel load prediction is vital.

**6.4 System-Level Impact and Operating Characteristics**

In order to show the realistic application of the proposed forecasting model, we fed the result of the proposed forecasting model into a simulated environment of operation of a hybrid wind-diesel system. Namely, the results of diesel generation models, estimated by fitting the yesterdays, were fed, to make the system dispatch decisions minimizing the cost of telling wind and diesel generation sources in the scope of the objective function described in Section 3. Some of the critical operating attributes, as exemplified by the results are:

• Load Following: Forecasts produced by the model allowed the diesel generator to ramp-up or ramp-down on time thus avoiding too much fuel waste and generation to match actual load demand was met.

Emissions Reduction: The system has recorded substantial decreases in $CO_2$ emissions- which were captured in the performance evaluation indicators- hence by deferring the use of diesel generators by virtue of their integration with renewables, a significant drop in emissions has been recorded.

• Cost Optimization: Forecast-based dispatches resulted in a reduced overall operating cost along with fuel and low maintenance costs since the generator would not be over-cycled.

• Reliability: Unmet load was limited in each of the tested cases, and reliability of the energy was thus proved even with high variability of renewable energy sources.

The visual and numerical results of the impact of these processes are depicted in Figures 13-16, where the contribution of the model in some main parameters of the system (diesel load, emissions, and cost) is directly illustrated. Therefore, the constructed framework does not only provide a higher level of forecasting performance but also exploits this performance into real-world hybrid wind-diesel systems into system-level performance gains.

## 7. Conclusion and Future Scope

The proposed research presented comprehensive and intelligent forecasting framework tailored for hybrid wind-diesel power systems by leveraging a PSO-optimized DeAI-LSTM model. After the careful preprocessing of data, such as missing value imputation, outlier's identification, feature normalization, and dimensionality reduction, the raw multivariate data was refined into the structured and information-rich input space. Principal Component Analysis (PCA) and Deep Autoencer-based Integration (DeAI) made it possible to eliminate redundant features and those with noise to maximized temporal signals that could not be represented by fewer features. This data when fed into an appropriately optimised LSTM neural network model with Particle Swarm Optimization results in a powerful predictive model with the capability to describe non-linear interdependencies within hybrid power systems. The model exhibited meritorious results against mere baseline models like vanilla LSTM, GRU and ARIMA. The predictive trustworthiness of the proposed approach was proved via the RMSE scores and MAE scores, as well, with the DeAI-LSTM only posting the lowest level of error, regardless of the test measure. The model was able to model the trends in the diesel generator dispatch to optimize between the variability in load and variation of renewable energy to discharge it accurately. Training-validation loss curves and plotting prediction also indicated that the model generalized quite well without the over fitting effect. In situations where it is difficult to implement regular forecasting models due to data irregularities and multivariate effects, the methodology has been of great use. When considering the future, there are some avenues that will be encouraging. First, real time deployment the model into operational hybrid systems may be able to provide live forecasting and automatic controls increasing energy efficiency and the scheduling of fuel. Second, it would permit deployment of models into remote or bandwidth-constrained location, especially in remote microgrids by integrating with edge computing infrastructure. Third, the alternative is that DeAI-LSTM can be merged with more complex structures (e.g., Convolutional Neural Networks (CNNs), attention modules, or Transformer architectures) that will be able to learn the spatial-temporal correlations and drastic fluctuations in the future. Moreover, creating the greater flexibility of the model via transfer learning would enable it to predict on the different geographic spaces or energy setups with little retraining. Lastly, exogenous variables like weather predictions and policy adjustments would also serve to improve the real-life applicability and forecasting accuracy of the model in use.

## REFERENCES:

1. Ahmad, P., & Singh, N. (2020, November). Optimal sizing of ESS in a hybrid wind-diesel power system using NAR and NARX model. In **2020** IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (pp. 1–6). IEEE.

2. Aziz, A. S., Tajuddin, M. F. N., Hussain, M. K., Adzman, M. R., Ghazali, N. H., Ramli, M. A. M., &Zidane, T. E. K. **(2022).** A new optimization strategy for wind/diesel/battery hybrid energy system. Energy, 239, Article 122458. https://doi.org/10.1016/j.energy.2021.122458

3. Cao, Y., Taslimi, M. S., Dastjerdi, S. M., Ahmadi, P., &Ashjaee, M. **(2022).** Design, dynamic simulation, and optimal size selection of a hybrid solar/wind and battery-based system for off-grid energy supply. Renewable Energy, 187, 1082–1099. https://doi.org/10.1016/j.renene.2022.01.062

4. Dino, I. G., &MeralAkgül, C. **(2019).** Impact of climate change on the existing residential building stock in Turkey: An analysis on energy use, greenhouse gas emissions and occupant comfort. Renewable Energy, 141, 828–846. https://doi.org/10.1016/j.renene.2019.04.056

5. Dufo-López, R., Champier, D., Gibout, S., Lujano-Rojas, J. M., &Domínguez-Navarro, J. A. **(2019).** Optimisation of off-grid hybrid renewable systems with thermoelectric generator. Energy Conversion and Management, 196, 1051–1067. https://doi.org/10.1016/j.enconman.2019.06.069

6. Effatpanah, S. K., Ahmadi, M. H., Aungkulanon, P., Maleki, A., Sadeghzadeh, M., Sharifpur, M., &Lingen, C. **(2022).** Comparative analysis of five widely-used multi-criteria decision-making methods to evaluate clean energy technologies: A case study. Sustainability, 14(3), Article 140314. https://doi.org/10.3390/su1403140314

7. Elistratov, V., Konishchev, M., Denisov, R., Bogun, I., Grönman, A., Turunen-Saaresti, T., & Lugo, A. J. **(2021).** Study of the intelligent control and modes of the arctic-adopted wind–diesel hybrid system. Energies, 14(14), 4188.

8. Fares, D., Fathi, M., &Mekhilef, S. **(2022).** Performance evaluation of metaheuristic techniques for optimal sizing of a stand-alone hybrid PV/wind/battery system. Applied Energy, 305, Article 117823. https://doi.org/10.1016/j.apenergy.2021.117823

9. Gautam, A. K., Pareek, R. K., & Solanki, V. **(2024)**. Hybrid Intelligent Optimization Techniques for Grid Integration with Renewable Systems. In E3S Web of Conferences (Vol. 540, p. 10009). EDP Sciences.

10. Gautam, A. K., Pareek, R. K., & Solanki, V. **(2024)**. Hybrid intelligent optimization techniques for grid integration with renewable systems. In E3S Web of Conferences (Vol. 540, p. 10009). EDP Sciences.

11. Hamza, M. F., Modu, B., &Almutairi, S. Z. **(2025)**. Integration of the Chimp Optimization Algorithm and rule-based energy management strategy for enhanced microgrid performance considering energy trading pattern. Electronics, 14(10), 2037.

12. Jarso, A. K., Jin, G., &Ahn, J. **(2025)**. Hybrid genetic algorithm-based optimal sizing of a PV–wind–diesel–battery microgrid: A case study for the ICT Center, Ethiopia. Mathematics, 13(6), 985.

13. Jarso, A. K., Jin, G., &Ahn, J. **(2025)**. Hybrid Genetic Algorithm-Based Optimal Sizing of a PV–Wind–Diesel–Battery Microgrid: A Case Study for the ICT Center, Ethiopia. Mathematics, 13(6), 985.

14. Jasim, A. M., Jasim, B. H., Baiceanu, F. C., &Neagu, B. C. **(2023).** Optimized sizing of energy management system for off-grid hybrid solar/wind/battery/biogasifier/diesel microgrid system. Mathematics, 11(6), Article 1248. https://doi.org/10.3390/math11061248

15. Kang, W., Chen, M., Lai, W., &Luo, Y. **(2021)**. Distributed real-time power management for virtual energy storage systems using dynamic price. Energy, 216, Article 119069. https://doi.org/10.1016/j.energy.2020.119069

16. Ma, Q., Huang, X., Wang, F., Xu, C., Babaei, R., &Ahmadian, H. **(2022)**. Optimal sizing and feasibility analysis of grid-isolated renewable hybrid microgrids: Effects of energy management controllers. Energy, 240, Article 122503. https://doi.org/10.1016/j.energy.2021.122503

17. Mahmoudi, S. M., Maleki, A., &RezaeiOchbelagh, D. **(2022)**. A novel method based on fuzzy logic to evaluate the storage and backup systems in determining the optimal size of a hybrid renewable energy system. Journal of Energy Storage, 49, Article 104015. https://doi.org/10.1016/j.est.2022.104015

18. Movludiazar, A., Khayaty, M. S., Shiekh-El-Eslami, M. K., &Fotouhi, R. **(2021, May)**. A data-driven bidding strategy of a wind-diesel-electrical storage hybrid system in a day-ahead electricity market. In 7th Iran Wind Energy Conference (IWEC2021) (pp. 1–5). IEEE.

19. Murugaperumal, K., Srinivasn, S., & Prasad, G. S. **(2020)**. Optimum design of hybrid renewable energy system through load forecasting and different operating strategies for rural electrification. Sustainable Energy Technologies and Assessments, 37, 100613.

20. Musa, B., Yimen, N., Abba, S. I., Adun, H. H., &Dagbasi, M. **(2021).** Multi-state load demand forecasting using hybridized support vector regression integrated with optimal design of off-grid energy systems—A metaheuristic approach. Processes, 9(7), 1166.

21. Nguyen, X. P. **(2020, March)**. A strategy development for optimal generating power of small wind-diesel-solar hybrid microgrid system. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 1329–1334). IEEE.

22. Nsafon, B. E. K., Owolabi, A. B., Butu, H. M., Roh, J. W., Suh, D., & Huh, J. S. **(2020)**. Optimization and sustainability analysis of PV/wind/diesel hybrid energy system for decentralized energy generation. Energy Strategy Reviews, 32, 100570.

23. Patel, Z. S., Touileb, R., Quadar, N., Chaibi, H., Saadane, R., &Jakimi, A. **(2025, May).** Hybrid Wind-Diesel Energy System with Energy Storage for Remote Applications. In 2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET) (pp. 1-7). IEEE.

24. Patel, Z. S., Touileb, R., Quadar, N., Chaibi, H., Saadane, R., &Jakimi, A. **(2025, May)**. Hybrid wind-diesel energy system with energy storage for remote applications. In 2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET) (pp. 1–7). IEEE.

25. Pavan, G., Babu, A. R., BolluPrabhakar, T., Rajeshwari, M., Reddy, N. R., & Kishore, P. V. **(2025)**. Advanced optimization load frequency control for multi-islanded micro grid system with tie-line loading by using PSO. International Journal of Information & Communication Technology, 2252(8776), 8776.

26. Pinto, R., Henriques, S. T., Brockway, P. E., Heun, M. K., & Sousa, T. **(2023)**. The rise and stall of world electricity efficiency: 1900–2017, results and insights for the renewables transition. Energy, 269, Article 126617. https://doi.org/10.1016/j.energy.2023.126617

27. Qashou, Y., Samour, A., &Abumunshar, M. **(2022)**. Does the real estate market and renewable energy induce carbon dioxide emissions? Novel evidence from Turkey. Energies, 15(1), Article 13. https://doi.org/10.3390/en15010013

28. Ranjan, R., Doda, D. K., Lalwani, M., &Bundele, M. **(2020, February)**. Simulation and optimization of solar photovoltaic–wind–diesel generator stand-alone hybrid system in remote village of Rajasthan, India. In International Conference on Artificial Intelligence: Advances and Applications 2019: Proceedings of ICAIAA 2019 (pp. 279–286). Singapore: Springer Singapore.

29. Rim, B. A., Mohsen, B. A., &Oualha, A. **(2024)**. Improving wind power forecast accuracy for optimal hybrid system energy management. Journal of Energy Resources Technology, 146(9), 092101.

30. Rim, B. A., Mohsen, B. A., &Oualha, A. **(2024)**. Improving wind power forecast accuracy for optimal hybrid system energy management. Journal of Energy Resources Technology, 146(9).\

31. Shaahid, S. M., Alhems, L. M., & Rahman, M. K. **(2025, March)**. Unlocking the Prospects of Wind-Diesel Hybrid Power Systems to meet Commercial Loads of Qaisumah of SaudiArabia-A Pathway for Sustainable Energy Development. In 2025 9th International Conference on Green Energy and Applications (ICGEA) (pp. 1-5). IEEE.

32. Shaahid, S. M., Alhems, L. M., & Rahman, M. K. **(2025, March)**. Unlocking the prospects of wind-diesel hybrid power systems to meet commercial loads of Qaisumah of Saudi Arabia—A pathway for sustainable energy development. In 2025 9th International Conference on Green Energy and Applications (ICGEA) (pp. 1–5). IEEE.

33. Sosnina, E., Dar'enkov, A., Kurkin, A., Lipuzhin, I., &Mamonov, A. **(2022)**. Review of efficiency improvement technologies of wind diesel hybrid systems for decreasing fuel consumption. Energies, 16(1), 184.

34. Sukanya, K., &Vijayakumar, P. **(2023)**. Frequency control approach and load forecasting assessment for wind systems. Intelligent Automation & Soft Computing, 35(1).

35. `Uddin, M. N., Biswas, M. M., &Nuruddin, S. **(2022)**. Techno-economic impacts of floating PV power generation for remote coastal regions. Sustainable Energy Technologies and Assessments, 51, Article 101930. https://doi.org/10.1016/j.seta.2021.101930

36. Xu, Y., Huang, S., Wang, Z., Ren, Y., Xie, Z., Guo, J., et al. **(2022)**. Optimization based on tabu search algorithm for optimal sizing of hybrid PV/energy storage system: Effects of tabu search parameters. Sustainable Energy Technologies and Assessments, 53, Article 102662. https://doi.org/10.1016/j.seta.2022.102662

37. Yi, H., & Yang, X. **(2022)**. A metaheuristic algorithm based on simulated annealing for optimal sizing and techno-economic analysis of PV systems with multi-type of battery energy storage. Sustainable Energy Technologies and Assessments, 53, Article 102724. https://doi.org/10.1016/j.seta.2022.102724

38. Yu, J., Ryu, J. H., & Lee, I. B. **(2019)**. A stochastic optimization approach to the design and operation planning of a hybrid renewable energy system. Applied Energy, 247, 212–220. https://doi.org/10.1016/j.apenergy.2019.04.058

39. Yu, X., Li, W., Maleki, A., Rosen, M. A., KomeiliBirjandi, A., & Tang, L. **(2021)**. Selection of optimal location and design of a stand-alone photovoltaic scheme using a modified hybrid methodology. Sustainable Energy Technologies and Assessments, 45, Article 101071. https://doi.org/10.1016/j.seta.2021.101071

40. Zeljković, Č., Mršić, P., Erceg, B., Lekić, Đ., Kitić, N., &Matić, P. **(2022)**. Optimal sizing of photovoltaic-wind-diesel-battery power supply for mobile telephony base stations. Energy, 242, Article 122545. https://doi.org/10.1016/j.energy.2021.122545

41. Zhang, W., &Maleki, A. **(2022)**. Modeling and optimization of a stand-alone desalination plant powered by solar/wind energies based on back-up systems using a hybrid algorithm. Energy, 254, Article 124341. https://doi.org/10.1016/j.energy.2022.124341

42. Zhou, J., &Xu, Z. **(2023)**. Optimal sizing design and integrated cost-benefit assessment of stand-alone microgrid system with different energy storage employing chameleon swarm algorithm: A rural case in Northeast China. Renewable Energy, 202, 1110–1137. https://doi.org/10.1016/j.renene.2022.12.101