WEATHER FORECASTING FOR POWER PREDICTION

24EEE431 – AI AND EDGE COMPUTING

Report

Submitted by

CB.EN.U4EEE22026	**NEHARIKA K**
CB.EN.U4EEE22032	**PRITHIKA SV**
CB.EN.U4EEE22154	**VEDAVARSHA N T**

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING
AMRITA SCHOOL OF ENGINEERING,
AMRITA VISHWA VIDYAPEETHAM,
COIMBATORE - 641112

MARCH-2025

CONTENTS

AIM OF THE PROJECT

The primary aim of this project is to develop an AI-driven weather forecasting model for power prediction, utilizing machine learning and deep learning techniques. This project focuses on leveraging historical weather data and power consumption records to build accurate predictive models that can assist in optimizing energy management, enhancing renewable energy integration, and improving power grid efficiency.

With the increasing global emphasis on renewable energy sources such as solar and wind, the ability to predict power generation based on weather conditions has become crucial. The variability in weather patterns directly affects the efficiency of renewable energy systems, leading to fluctuations in power generation. Traditional forecasting methods often fail to provide the accuracy required for efficient energy planning. This project aims to overcome these challenges by implementing advanced AI techniques such as Random Forest, Linear Regression, Multilayer Perceptron (MLP), and clustering methods to enhance prediction accuracy.

The project also aims to explore different machine learning algorithms in WEKA to determine the most effective model for power forecasting. By utilizing density-based clustering, classification methods, and deep learning techniques, the project seeks to provide an optimized solution for predicting power generation under various weather conditions. Additionally, this research aims to contribute to smart grid development by enabling data-driven decision-making for energy distribution and consumption.

By achieving these objectives, this project aspires to contribute to the efficient utilization of renewable energy resources, reduce dependency on fossil fuels, and enhance the overall reliability of power systems. The findings from this research can be further extended to real-world applications, including smart cities, industrial power planning, and grid stability enhancement.

PROBLEM STATEMENT AND SOLUTION

Problem Statement:

- Weather forecasting plays a crucial role in predicting the availability and efficiency of renewable energy sources such as wind and solar power
- By accurately forecasting weather conditions, we can predict power generation levels and optimize the operation of energy systems.
- This project aims to use AI techniques for weather forecasting to predict power generation in renewable energy systems.
- The goal is to provide accurate, real-time weather predictions and power output estimates for better grid management and energy optimization.
- The project will use various weather data inputs (e.g., temperature, wind speed, solar radiation, humidity) and apply machine learning models to forecast the expected power output of renewable energy systems.
- By integrating weather forecasting with power prediction, energy companies can improve grid stability, minimize energy waste, and ensure a more reliable power supply.

Solution and Uniqueness/ Differentiator in the Project

1. **Machine Learning-Based Prediction Models**

- Utilizing Random Forest, Linear Regression, and MLP in WEKA for power prediction.
- Implementing clustering techniques to analyze weather patterns for better accuracy.

2. **Deep Learning for Enhanced Accuracy**

- Applying neural networks and deep learning models to capture complex weather-power relationships.
- Using MLP and advanced regression techniques to improve forecasting precision.

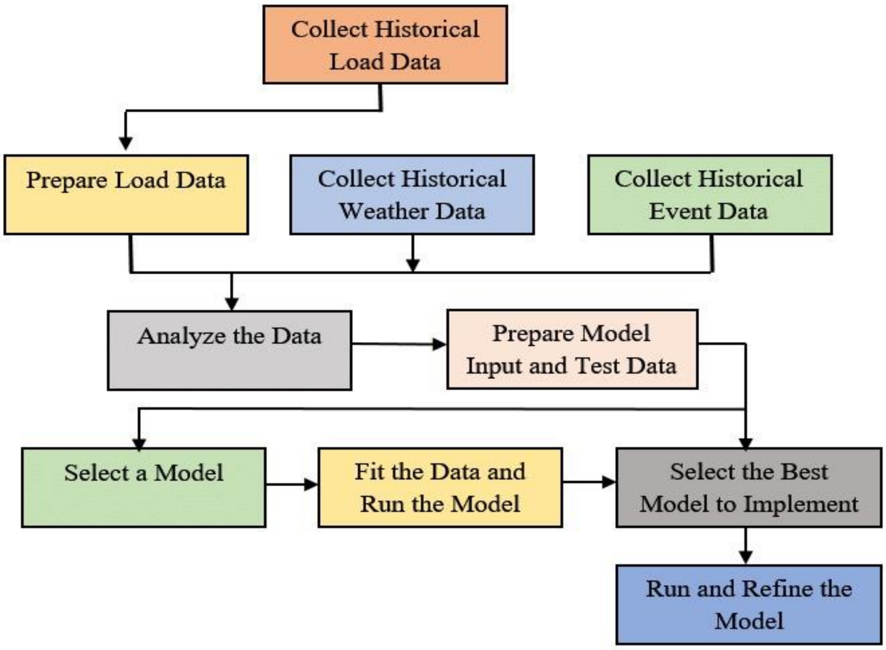
3. **Data-Driven Optimization**

- Training models using historical weather and power generation data for better prediction accuracy.
- Selecting the most efficient algorithm by comparing model performance metrics.

4. **Smart Grid and Renewable Energy Integration**

- Supporting real-time energy management for optimized power distribution.
- Assisting smart grids and energy storage systems in handling demand fluctuations.

FLOW EXPLANATION



The methodology follows a structured flow, as represented in the diagram.

1. Data Collection

The process begins with the collection of three key datasets:

- Historical Load Data: Power consumption patterns over time.
- Historical Weather Data: Temperature, humidity, wind speed, and other meteorological variables.
- Historical Event Data: Major weather events that influenced power demand or generation.

2. Data Preparation and Analysis

- Load Data Preparation: Cleaning and structuring historical load data for further analysis.
- Weather Data Collection: Gathering meteorological data and ensuring consistency.
- Data Analysis: Identifying correlations between power load and weather conditions.

3. Model Development

- Preparing Model Input and Test Data: Transforming raw data into a structured format suitable for machine learning or statistical models.
- Model Selection: Evaluating different forecasting models (such as regression models, neural networks, or time series methods).
- Fitting the Data and Running the Model: Training the selected model using historical datasets.

4. Model Evaluation and Implementation

- Selecting the Best Model: Comparing performance metrics to determine the most effective model.
- Running and Refining the Model: Implementing the model in real-world scenarios and continuously improving accuracy with updated data.

MACHINE LEARNING MODELS USED

A. Power Consumption Prediction:

Implemented Random Forest and XGBoost regressors to predict cumulative power usage using weather data. XGBoost showed higher accuracy in forecasting power consumption trends.

B. Power Consumption Classification:

Utilized Gradient Boosting and K-Nearest Neighbors classifiers to categorize power consumption levels. KMeans clustering was applied to create consumption categories. Gradient Boosting achieved higher accuracy in classification.

C. Power Consumption Prediction Using MLP Classifier:

Implemented a Multi-Layer Perceptron (MLP) classifier to predict power consumption levels based on weather data. The target variable was discretized into three classes (Low, Medium, High) using quantile binning. The trained model achieved a good accuracy score, demonstrating effective classification performance.

D. Power Consumption Prediction and Peak Hour Analysis Using Random Forest Regressor:

Implemented a Random Forest Regressor to predict cumulative power consumption based on weather data. Identified peak hours by classifying the top 10% highest power usage and analyzed the power distribution. Also included a renewable energy optimization step for solar power usage efficiency.

E. Power Consumption Prediction Using CNN-Based Deep Learning Model:

Implemented a CNN-based deep learning model to predict power consumption using weather data. The model includes Conv1D layers for feature extraction, dropout for regularization, and dense layers for regression. Performance was optimized with learning rate tuning, achieving improved accuracy with MAE, MSE, and R^2 metrics.

F. Optimized Electricity Pricing for Peak Load Reduction Using Smoothed Predictions:

Developed a pricing optimization strategy using smoothed load predictions to reduce peak electricity demand. A moving average filter was applied to enhance prediction stability, aligning optimized pricing with demand fluctuations. The approach aims to improve grid efficiency and reduce costs.

G. Peak Hour Prediction Using CatBoost Classifier:

Utilizes the CatBoost Classifier to predict peak electricity usage hours based on weather and power consumption data. The model achieves an impressive accuracy of 99.83% demonstrating its effectiveness in distinguishing between peak and non-peak hours.

RESULT AND DISCUSSION

A. Power Consumption Prediction:

Random Forest Regressor Performance:

MAE: 1971.7670

MSE: 10,911,322.0237

R^2 Score: 0.8807

XGBoost Regressor Performance

MAE: 2526.8329

MSE: 14,803,271.1081

R^2 Score: 0.83832

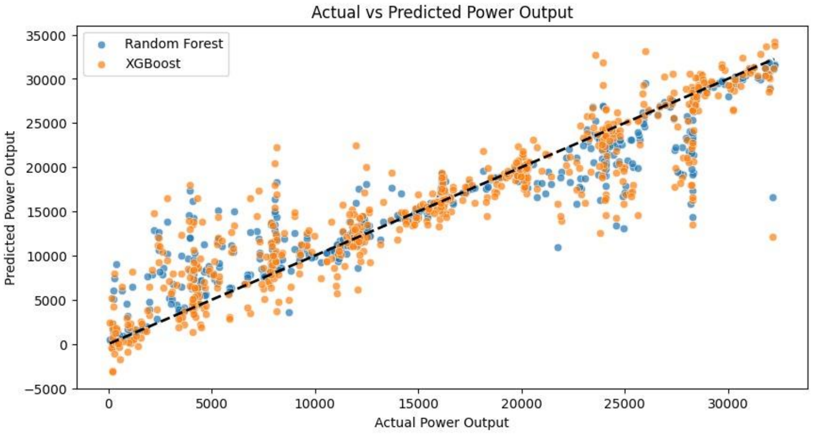


Fig.1. Power Consumption Prediction using Random Forest and XGBoost Regressor

Random Forest Classifier Performance:

- Accuracy: 92.88%
- Precision: 92.90%
- Recall: 92.88%
- F1 Score: 92.88%

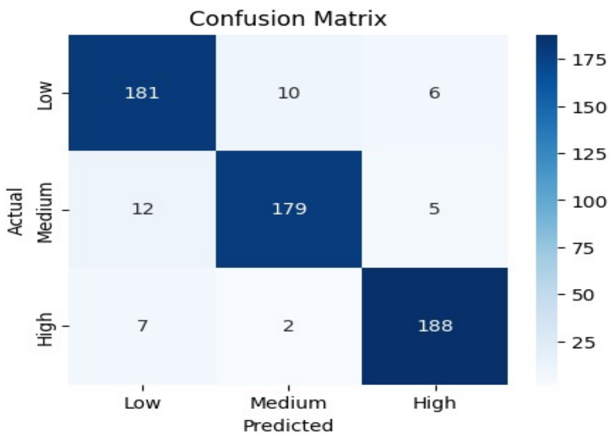


Fig.2. Power Consumption Prediction using Random Forest Classifier

B. Power Consumption Classification:

Gradient Boosting Classifier Performance:

- Accuracy: 87.12%
- Precision: 87.09%
- Recall: 87.12%
- F1-Score: 87.03%

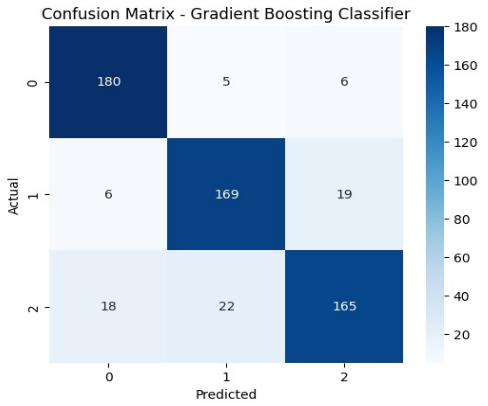


Fig.3. Power Consumption Classification using Gradient Boosting Classifier

K-Neighbors Classifier Performance:

- Accuracy: 88.14%
- Precision: 88.35%
- Recall: 88.14%
- F1-Score: 88.05%

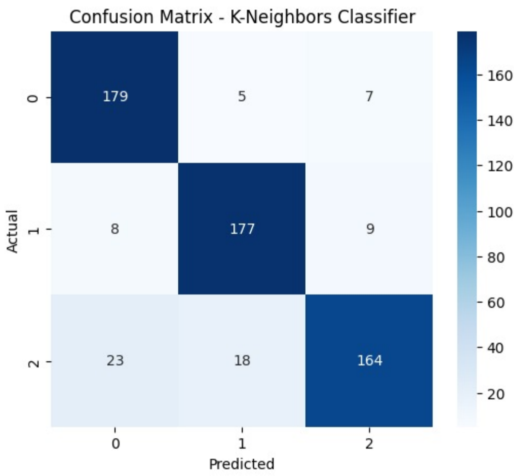


Fig.4. Power Consumption Classification using K-Neighbors Classifier

C. Power Consumption Prediction Using MLP Classifier:

- Accuracy: 0.9288
- Precision: 0.9290
- Recall: 0.9288
- F1 Score: 0.9288

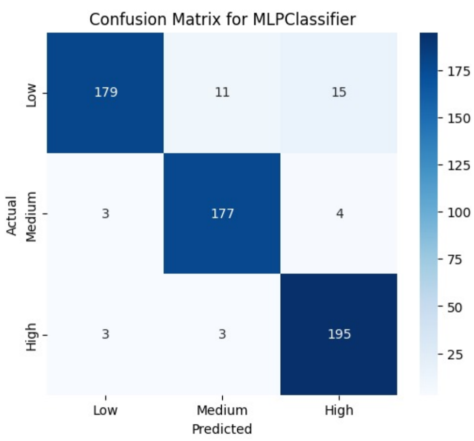


Fig.5. Power Consumption Classification using MLP Classifier

D. Power Consumption Prediction and Peak Hour Analysis Using Random Forest Regressor:

- MAE: 1971.7670
- MSE: 10911322.0237
- R^2 Score: 0.8807

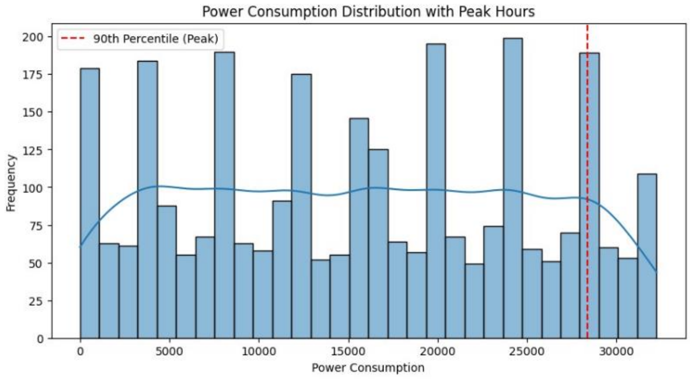


Fig.6. Power Consumption Prediction and Peak Hour Analysis using Random Forest Regressor

E. Power Consumption Prediction Using CNN-Based Deep Learning Model:

- MAE: 4.9364
- MSE: 41.9432
- R^2 Score: 0.5392

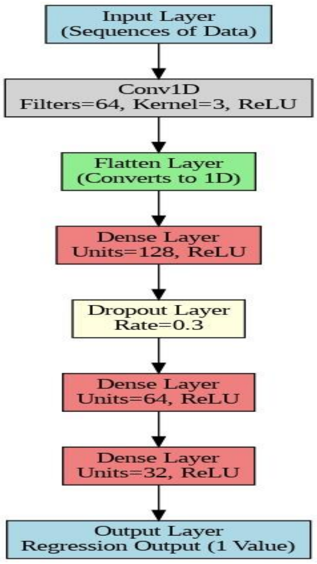


Fig.7. Power Consumption Prediction using CNN Based Deep Learning Model

F. Optimized Electricity Pricing for Peak Load Reduction Using Smoothed Predictions:

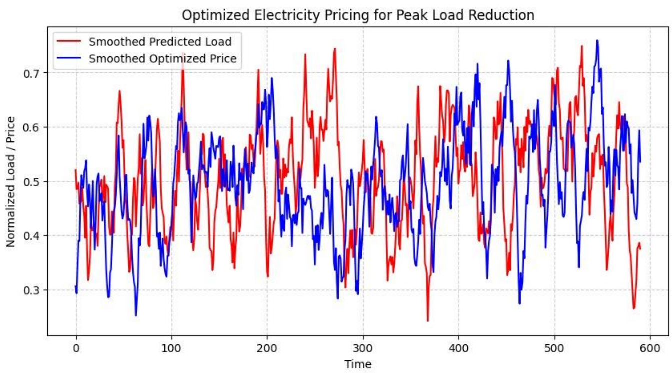


Fig.8. Optimized Electricity Pricing for Peak Load Reduction Using Smoothed Predictions

G. Peak Hour Prediction Using CatBoost Classifier:

- Accuracy: 0.9288
- Precision: 0.9290
- Recall: 0.9288
- F1 Score: 0.9288

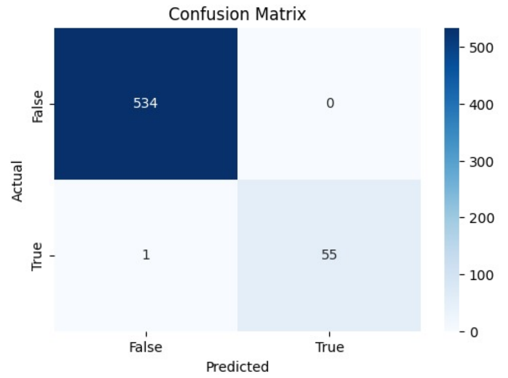


Fig.9. Confusion Matrix of Peak Hour Prediction Using CatBoost Classifier

H. Correlation Heatmap of Weather and Power Data

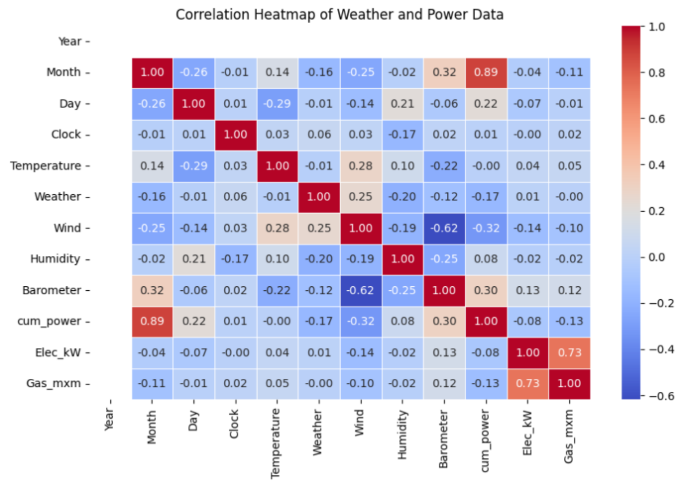


Fig.10. Correlation Heatmap of Weather and Power Data

I. Electric KW Prediction Using Random Forest (Weka)

- Correlation Coefficient: 0.7328
- Mean Absolute Error (MAE): 5.3851
- Root Mean Squared Error (RMSE): 6.7429
- Relative Absolute Error: 64.08%
- Root Relative Squared Error: 68.52%
- Total Instances Tested: 1002
- The correlation coefficient of 0.7328 indicates a moderate positive relationship between predictions and actual values.
- The mean absolute error (5.3851) and root mean squared error (6.7429) suggest room for improvement in accuracy.
- The relative absolute error (64.08%) and root relative squared error (68.52%) indicate that the model has a moderate level of error.

J. Power Consumption Prediction (cum_power) Using Random Forest (Weka)

- Correlation Coefficient: 0.9958 (Indicating a very strong relationship between predictions and actual values)
- Mean Absolute Error (MAE): 884.6019
- Root Mean Squared Error (RMSE): 1133.7993
- Relative Absolute Error: 10.63% (Low error, good accuracy)
- Root Relative Squared Error: 11.85% (Indicating high model performance)
- Total Instances Tested: 1002
- High correlation coefficient (0.9958) suggests excellent prediction accuracy.
- Low relative errors (10.63% and 11.85%) indicate the model performed well.
- The model was very fast to train (0.35 sec) and test (0.03 sec), making it computationally efficient.

K. Power Consumption Prediction (Elec_kW) Using MLP (Weka)

- Correlation Coefficient: 0.6122 (Moderate correlation)
- MAE: 8.1537
- RMSE: 10.53
- Relative Absolute Error: 97.03%
- Root Relative Squared Error: 107.00%
- Moderate correlation but high errors indicate poor prediction accuracy.
- Long training time (17.13 sec) compared to Random Forest.
- MLP struggled with predicting Elec_kW, requiring tuning or better features.

L. Power Consumption Prediction (Elec_kW) Using Linear Regression (Weka)

- Correlation Coefficient: 0.729 (Moderate correlation)
- MAE: 5.4397
- RMSE: 6.737
- Relative Absolute Error: 64.73%
- Root Relative Squared Error: 68.46%
- Faster training & testing compared to MLP and Random Forest.
- Moderate prediction accuracy, better than MLP but lower than Random Forest.

M. Clustering Power Consumption Data Using K-Means (Weka)

- Cluster 0: 48% (1409 instances)
- Higher weather index, wind speed, and humidity
- Lower barometric pressure and cumulative power consumption
- Cluster 1: 52% (1539 instances)
- Lower weather index and wind speed
- Higher barometric pressure and cumulative power consumption
- Within-Cluster Sum of Squared Errors (WCSS): 4076.50
- Missing Values: Replaced with mean/mode
- Cluster 0 represents lower power consumption with higher weather impact.
- Cluster 1 has higher power usage with lower wind speed & weather variations.

N. Density-Based Clustering Results (WEKA)

1. Cluster 0 (Lower Cumulative Power):
    - Mostly January (1.01), mid-month (15.57), around 12:50. Cooler (~6.07°C) with extreme weather (72.15), stronger winds (16.67 km/h), higher humidity (80.5%), and lower pressure (1020.62). Low power consumption (7440.13), moderate electric usage (4.48 kW), and higher gas consumption (8.72).
2. Cluster 1 (Higher Cumulative Power):

- Mostly February (2.11), around the 13th, at 07:20. Slightly warmer (~6.52°C) with milder weather (55.87), lower wind speeds (12.84 km/h), slightly lower humidity (78.54%), and higher pressure (1028.16). Power consumption is significantly higher (23129.93), with similar electric usage (4.55 kW) and lower gas consumption (8.16).

CONCLUSION

This project successfully demonstrated the effectiveness of various machine learning and deep learning techniques in predicting and classifying power consumption based on weather data. The results showed that Random Forest and XGBoost were highly effective in regression tasks, while Gradient Boosting and K-Nearest Neighbors performed well in classification. The Multi-Layer Perceptron (MLP) model achieved high accuracy in power consumption classification. Additionally, CNN-based deep learning models provided promising results in forecasting power consumption trends. The integration of clustering techniques, such as K-Means and density-based clustering, helped analyze consumption patterns. The findings from this project can contribute to optimizing power usage, improving grid efficiency, and developing energy management strategies for sustainable power consumption.

REFERENCE

[1] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[2] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD)*, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[5] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 281–305, Feb. 2012.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[7] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Stat.*, vol. 28, no. 2, pp. 337–407, Apr. 2000.

[8] L. Rokach and O. Maimon, "Clustering methods," in *Data Mining and Knowledge Discovery Handbook*. Boston, MA, USA: Springer, 2005, pp. 321–352.

[9] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, 1967, pp. 281–297.

[10] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.