

Course Title: Research Methodology in Computer Science

Course Code: A1800

Weighting: 20%

Lecturer: Dr. Mohammed Anbar

Hand-out Date: 20 June 2026

Hand-in Date: 04 July 2026

Hand-in Method: Online Submission via E-Learning (Moodle)

Assignment Report: Literature Review — Machine Learning-Based Approaches for Power Plant Operations and Asset Management

Student ID: 262000880

Student Name: Salem Ahmed Buhluli

Table of Contents

- Summary of the Research Papers
- Critical Analysis
- Connections
- Conflicts between the Findings
- Synthesis Matrix for Literature Review
- References

Summary of the Research Papers

Paper 1: Prediction of Full Load Electrical Power Output of a Base Load Operated Combined Cycle Power Plant Using Machine Learning Methods

This paper investigates whether the full-load electrical power output of a combined cycle power plant (CCPP) can be predicted accurately from ambient conditions using machine learning instead of thermodynamic modelling. The author collected 9,568 hourly records over six years (2006–2011) from a real CCPP operating at full load, using four ambient variables (temperature, ambient pressure, relative humidity, and exhaust vacuum) as inputs. Fifteen regression methods were compared under a repeated cross-validation protocol. The best result was obtained by the Bagging ensemble with REP Tree, which reached a mean absolute error of 2.818 MW and a root mean square error of 3.787 MW, showing that ambient data alone can predict plant output with high precision and support operational planning and efficiency monitoring. Two main drawbacks stand out: 1) the model covers base-load, full-load operation only, so it says nothing about part-load or transient behaviour, which is where prediction is hardest; and 2) the data come from a single plant, so the transferability of the trained models to other plants with different configurations is not demonstrated. (Tüfekci, 2014)

Paper 2: Machine Learning Methods for Wind Turbine Condition Monitoring: A Review

This paper reviews the recent literature on machine learning models applied to condition monitoring of wind turbines, an asset class that is increasingly part of power generation portfolios. The authors organise the reviewed studies along the typical machine learning pipeline: data sources, feature selection and extraction, model selection (classification or regression), validation, and decision-making. Their analysis shows that most studies rely on SCADA or simulated data, that roughly two-thirds of the methods treat the problem as classification while the rest use regression, and that neural networks, support vector machines, and decision trees are the most commonly used models. The review also identifies recurring obstacles: scarcity of labelled fault data, class imbalance, and limited validation on operating wind farms. The main limitations are: 1) as a review, it produces no new empirical evidence and the quality of its conclusions depends on the primary studies, many of which used simulated rather than field data; and 2) the coverage is restricted to wind turbines, so the findings do not automatically carry over to thermal or combined cycle assets. (Stetco et al., 2019)

Paper 3: A Systematic Literature Review of Machine Learning Methods Applied to Predictive Maintenance

This study presents a systematic literature review of machine learning methods used in predictive maintenance (PdM) of industrial equipment, following a structured protocol to identify, filter, and catalogue peer-reviewed papers. The authors classify the selected studies by the machine learning technique applied, the equipment monitored, and the type of data used. The key findings are that Random Forest, artificial neural networks, support vector machines, and k-means are the most frequently applied methods, and that the most promising results come from hybrid approaches in which several methods are combined into a stronger model. The review also highlights a structural weakness of the field: most works use private industrial datasets, which makes replication and fair comparison between papers very difficult. Its main limitations are: 1) the number of primary studies that passed the selection criteria is relatively small, so some equipment types and industries are thinly covered; and 2) the review addresses predictive maintenance in general rather than power plants specifically, so its conclusions need careful interpretation before being applied to the energy sector. (Carvalho et al., 2019)

Paper 4: Real-World Data-Driven Machine-Learning-Based Optimal Sensor Selection Approach for Equipment Fault Detection in a Thermal Power Plant

This paper tackles a practical problem in thermal power plants (TPPs): hundreds of sensors feed the historian, but not all of them are informative for fault detection, and redundant inputs degrade the performance of multivariate algorithms. The authors propose a two-step data-driven approach using real operational data from a thermal power plant. First, correlation analysis is applied to select the sensors most sensitive to the target faults; second, machine learning models

are trained on the reduced sensor set to detect boiler waterwall tube leakage and turbine faults, which are among the most frequent causes of TPP shutdowns. The key finding is that the optimal sensor subset maintained fault detection performance while clearly improving the computational efficiency of the models, which matters for online deployment. The main limitations are: 1) the evaluation is based on a single plant and two fault classes, so the generality of the selected sensor sets is unproven; and 2) the study relies on classical multivariate models and does not examine deep temporal architectures that might extract more from the same data. (Khalid et al., 2021)

Paper 5: Dynamic Modeling of NO_x Emission in a 660 MW Coal-Fired Boiler with Long Short-Term Memory

This study addresses emission-side operations: as coal-fired units increasingly participate in peak load regulation, NO_x concentration at the furnace exit varies widely during transients, which complicates the control of selective catalytic reduction (SCR) systems. The authors build a dynamic NO_x prediction model using a long short-term memory (LSTM) network trained on 10,000 samples covering seven days of operation of a real 660 MW tangentially fired boiler, and they carefully tune the learning rate, look-back window, and number of hidden nodes. The results show root mean square errors of only 7.6 mg/Nm³ on training data and 12.2 mg/Nm³ on test data, with mean absolute percentage errors within 3%, and the LSTM clearly outperformed the support vector machine used as a baseline under both steady and transient conditions. The main drawbacks are: 1) the dataset covers only seven days of one boiler, so seasonal effects, coal quality variation, and long-term drift are not captured; and 2) the work stops at prediction and does not close the loop into combustion optimisation or SCR control, which is the ultimate operational goal. (Tan et al., 2019)

Paper 6: Advancements in Gas Turbine Fault Detection: A Machine Learning Approach Based on the Temporal Convolutional Network–Autoencoder Model

This paper proposes a Temporal Convolutional Network (TCN)–Autoencoder with a multi-head attention bottleneck for detecting anomalies in gas turbine vibration data, as a basis for predictive maintenance. The dataset is notable because it is real: around one million data points were collected during December 2021 from four CA 202 accelerometers installed on the bearings of a gas turbine at the Kirkuk power plant, sampled at 1 kHz, with anomalies labelled from maintenance records and confirmed by plant engineers. After noise filtering and linear interpolation of gaps, the data were split 80/20 for training and testing. The proposed model reconstructed the vibration signals more accurately than GRU–Autoencoder, LSTM–Autoencoder, and variational autoencoder baselines, achieving MSE of 1.447, RMSE of 1.193, and MAE of 0.712, and it successfully flagged expert-confirmed events such as blade misalignment and an air-intake obstruction. The main limitations are: 1) the data cover a single month of one turbine, so robustness across seasons, units, and operating regimes is unknown; and 2) performance is reported as reconstruction error rather than detection-oriented metrics such as precision, recall, or false alarm rate, which makes it hard to judge operational usefulness. (Fahmi et al., 2024)

Critical Analysis

Taken together, the six papers cover the two halves of the research title: operations (output prediction by Tüfekci, 2014, and emission modelling by Tan et al., 2019) and asset management (fault detection and condition monitoring by Khalid et al., 2021, Fahmi et al., 2024, and the two reviews). A clear strength of this body of work is its grounding in real operational data — six years of CCPP records, a 660 MW utility boiler, plant historian data from a working thermal plant, and a full month of 1 kHz vibration data from an operating gas turbine — which makes the results far more credible for industry than purely simulated studies. The methodological breadth is also a strength: the papers span classical regression ensembles, sensor selection with multivariate statistics, recurrent networks, and modern attention-based autoencoders, and the two systematic reviews (Stetco et al., 2019; Carvalho et al., 2019) place these individual contributions in the wider landscape and identify Random Forest, SVM, and neural networks as the dominant tools.

On the other hand, several weaknesses recur across the studies. Almost every empirical paper is validated on a single plant, a single unit, or a short time window, so external validity is weak: Tüfekci (2014) covers only full-load operation, Tan et al. (2019) only seven days of data, and Fahmi et al. (2024) only one month of one turbine. Evaluation practices are inconsistent — reconstruction errors, regression errors, and classification accuracies are reported in different papers with no common benchmark — which makes it impossible to rank the approaches objectively, a problem both reviews trace to the lack of public industrial datasets. Deployment concerns are also underexplored: interpretability, integration with existing DCS/SCADA systems, model retraining under concept drift, and false-alarm cost are mentioned only in passing, even though they decide whether a model actually reaches the control room. Finally, class imbalance and the scarcity of labelled failures — flagged strongly by Stetco et al. (2019) — are handled differently in every paper, and none of them offers a convincing general solution. Future work would benefit from multi-site validation, shared benchmark datasets, and detection-oriented metrics reported alongside regression errors, so that the trade-off between accuracy and operational cost can be assessed honestly.

Connections

- 1- **Data-driven modelling of real plant data:** all six papers replace or complement physics-based models with machine learning trained on operational sensor data, and five of them use measurements taken from real generating assets (a CCPP, a thermal plant, a 660 MW coal boiler, a gas turbine, and SCADA-monitored wind turbines) rather than purely synthetic data.
- 2- **Predictive maintenance and early fault detection as the common goal:** Stetco et al. (2019), Carvalho et al. (2019), Khalid et al. (2021), and Fahmi et al. (2024) all aim at detecting equipment problems before failure, which is the core of asset management; Tüfekci (2014) and Tan et al. (2019) support the operations side of the same objective by predicting output and emissions.
- 3- **Neural networks and tree-based ensembles dominate:** the methods that recur across the papers are the same families — neural networks (including LSTM, TCN, and autoencoders), support vector machines, and decision-tree ensembles — which both reviews independently identify as the most used techniques in the field.
- 4- **Data preparation is treated as decisive:** several studies emphasise that model performance depends heavily on the input pipeline: Khalid et al. (2021) build an entire method around sensor selection, Tan et al. (2019) tune the temporal window carefully, and Fahmi et al. (2024) apply noise filtering and interpolation before training.
- 5- **Shared limitation of single-site validation:** almost every empirical paper validates on one plant or one unit and acknowledges (explicitly or implicitly) that transferability to other sites is unproven, a gap that both review papers also flag as an open problem in the field.

Conflicts between the Findings

- 6- **Classical ensembles vs. deep learning:** Tüfekci (2014) found that a bagged decision-tree ensemble beat all other methods, including neural networks, for output prediction, whereas Tan et al. (2019) and Fahmi et al. (2024) report that deep temporal models (LSTM, TCN–Autoencoder) clearly outperform shallower baselines. This suggests the best model family depends on whether the task is static regression or dynamic time-series modelling, and the papers do not agree on a single winner.
- 7- **Supervised vs. unsupervised detection:** Khalid et al. (2021) rely on supervised learning with labelled fault data, while Fahmi et al. (2024) use unsupervised reconstruction-based anomaly detection precisely because labelled failures are rare. Stetco et al. (2019) note the same split in the wind literature (about two-thirds classification), so the field has not settled on which paradigm suits scarce-fault industrial data.

- 8- **Fewer inputs vs. more inputs:** Khalid et al. (2021) argue that reducing the number of sensors improves multivariate fault detection, while Fahmi et al. (2024) feed high-frequency multi-channel vibration data directly into a deep model and let the network handle the dimensionality — two opposing philosophies on feature reduction.
- 9- **Evaluation metrics are inconsistent:** the regression-style papers report MAE/RMSE (Tüfekci, 2014; Tan et al., 2019; Fahmi et al., 2024), while the fault-detection and review papers discuss accuracy, precision, and false alarms. Because no two empirical papers share a metric and a dataset, their reported numbers cannot be compared directly — a conflict both reviews attribute to the absence of public benchmarks.
- 10- **Single-method models vs. hybrid approaches:** Carvalho et al. (2019) conclude that hybrid combinations of several ML methods give the most promising results, yet the empirical papers reviewed here each champion one architecture family; only Fahmi et al. (2024) partially follows the hybrid direction by combining TCN, attention, and autoencoder components.

Synthesis Matrix for Literature Review

Study	Authors	Year	Research Design	Key Findings / Results	Methodology	Limitations
Prediction of full load electrical power output of a base load operated CCPP using ML methods	Tüfekci, P.	2014	Experimental (comparative regression study)	Bagging with REP Tree was the best of 15 methods (MAE 2.818 MW, RMSE 3.787 MW). Four ambient variables are enough to predict full-load output with high accuracy. Ensemble/local learning methods beat single global models. Dataset later published on the UCI repository and became a standard benchmark.	Data: 9,568 hourly records from a real CCPP over 6 years (2006–2011) at full load. Inputs: temperature, ambient pressure, relative humidity, exhaust vacuum; target: net hourly electrical output. 15 ML regression methods compared with repeated 2-fold cross-validation (5x2). Evaluation with MAE and RMSE.	Full-load, base-load operation only; no part-load or transient regimes. Single plant — transferability not tested. Only ambient inputs; internal process variables ignored. Methods predate deep learning; no temporal modelling.
Machine learning methods for wind turbine condition monitoring: a review	Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., Nenadic, G.	2019	Systematic review	Most reviewed models use SCADA or simulated data. About two-thirds of methods are classification; the rest regression. Neural networks, SVMs and decision trees are the most common models. Scarce labelled fault data and class imbalance are the key obstacles; validation on real farms is limited.	Structured survey of recent ML condition-monitoring studies for wind turbines. Studies classified by pipeline stage: data source, feature selection/extraction, model selection, validation, decision-making. Tasks covered include blade fault detection, generator temperature and power-curve monitoring.	No new empirical results; conclusions depend on primary studies. Many primary studies rely on simulated data. Scope limited to wind turbines; findings may not transfer to thermal plants. No quantitative meta-analysis of reported accuracies.
A systematic literature review of ML methods applied to predictive maintenance	Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. P., Basto, J. P., Alcalá, S. G. S.	2019	Systematic literature review	Random Forest, ANN, SVM and k-means are the most frequently applied PdM methods. Hybrid ML methods (combinations of models) give the most promising results. Most studies use private datasets, blocking replication and fair comparison. PdM applications concentrate on rotating machinery and industrial equipment.	Systematic protocol: database search, inclusion/exclusion criteria, cataloguing of selected peer-reviewed papers. Papers classified by ML method, equipment type, and data used. Challenges and research opportunities synthesised across studies.	Relatively small set of primary studies passed the filters; uneven coverage of industries. General PdM focus, not power-plant specific. No standard benchmark proposed despite identifying the need. Fast-moving field: deep learning work after 2019 not covered.
Real-world data-driven ML-based optimal sensor selection for equipment fault detection in a thermal power plant	Khalid, S., Hwang, H., Kim, H. S.	2021	Experimental (real plant data)	Correlation-based sensor selection identified the subset of sensors actually sensitive to faults. Reduced sensor sets kept fault-detection performance while clearly improving computational efficiency. Approach validated for boiler waterwall tube leakage and	Historical operational data from a real thermal power plant historian. Step 1: correlation analysis to rank and select sensors per fault type. Step 2: multivariate ML models trained on the reduced sensor set for boiler and turbine fault detection.	Single plant, two fault classes; generality of the sensor subsets unproven. Classical multivariate models only; no deep temporal architectures. Exact dataset size and fault-event counts not fully detailed.

Study	Authors	Year	Research Design	Key Findings / Results	Methodology	Limitations
				turbine faults — the leading causes of TPP shutdowns.	Performance compared against models using the full expert-chosen sensor set.	No online/real-time deployment evaluation.
Dynamic modeling of NOx emission in a 660 MW coal-fired boiler with long short-term memory	Tan, P., Xia, J., Zhang, C., Fang, Q., Chen, G.	2019	Experimental (real boiler data)	LSTM predicted NOx dynamics with RMSE of 7.6 mg/Nm³ (training) and 12.2 mg/Nm³ (test). Mean absolute percentage errors within 3%. LSTM outperformed the SVM baseline under both steady and transient load conditions. Accurate NOx prediction supports SCR control during peak-load regulation.	10,000 samples covering 7 days of operation of a 660 MW tangentially coal-fired boiler. Operational parameters used as inputs; NOx at furnace exit as target. Hyperparameters (learning rate, look-back steps, hidden nodes) systematically optimised. Comparison against support vector machine on the same data.	Only 7 days of data from one boiler; seasonal and coal-quality variation not captured. Prediction only — no closed-loop combustion optimisation. Generalisation across load ranges beyond the sampled week untested. No interpretability analysis of the learned dynamics.
Advancements in gas turbine fault detection: TCN–Autoencoder model	Fahmi, A.-T. W. K., Reza Kashyzadeh, K., Ghorbani, S.	2024	Experimental (real plant data)	TCN–Autoencoder with multi-head attention beat GRU–AE, LSTM–AE and VAE baselines. MSE 1.447, RMSE 1.193, MAE 0.712 on reconstruction. Detected expert-confirmed anomalies incl. blade misalignment and air-intake obstruction. Multi-scale dilated convolutions captured short- and long-range temporal patterns.	~1,000,000 vibration data points from four CA 202 accelerometers on turbine bearings at the Kirkuk power plant (Dec 2021, 1 kHz). Preprocessing: moving-average noise filtering, linear interpolation of gaps. Anomalies labelled from maintenance records, confirmed by engineers; 80/20 train/test split. Multi-branch TCN encoder (dilation 1/2/4) + multi-head attention bottleneck + TCN decoder; Adam, MSE loss.	One month of data from a single turbine/plant. Reconstruction-error metrics only; no precision/recall or false-alarm rate. Anomaly labels depend on maintenance-record quality. No cross-season or cross-unit validation; deployment aspects not assessed.

References

- Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 106024. <https://doi.org/10.1016/j.cie.2019.106024>
- Fahmi, A.-T. W. K., Reza Kashyzadeh, K., & Ghorbani, S. (2024). Advancements in gas turbine fault detection: A machine learning approach based on the temporal convolutional network–autoencoder model. *Applied Sciences*, 14(11), 4551. <https://doi.org/10.3390/app14114551>
- Khalid, S., Hwang, H., & Kim, H. S. (2021). Real-world data-driven machine-learning-based optimal sensor selection approach for equipment fault detection in a thermal power plant. *Mathematics*, 9(21), 2814. <https://doi.org/10.3390/math9212814>
- Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., & Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy*, 133, 620–635. <https://doi.org/10.1016/j.renene.2018.10.047>
- Tan, P., Xia, J., Zhang, C., Fang, Q., & Chen, G. (2019). Dynamic modeling of NO_x emission in a 660 MW coal-fired boiler with long short-term memory. *Energy*, 176, 429–436. <https://doi.org/10.1016/j.energy.2019.04.020>
- Tüfekci, P. (2014). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *International Journal of Electrical Power & Energy Systems*, 60, 126–140. <https://doi.org/10.1016/j.ijepes.2014.02.027>