Development of an intelligent load control strategy for Itam distribution injection station using Supervisory Control and Data Acquisition (SCADA) systems with Machine Learning (ML) techniques

Ndifreke M. Uwah[1]

Department of Electrical/Electronics Engineering
University of Uyo, Akwa Ibom, Nigeria
ataubandifreke@gmail.com

U. T. Itaketo[2]

Department of Electrical/Electronics Engineering
University of Uyo, Akwa Ibom, Nigeria
uwanaitaketo@uniuyo.edu.ng,

Kingsley M. Udofia[3]

Department of Electrical /Electronic Engineering,
University of Uyo, Akwa Ibom State, Nigeria
kingsleyudofia@uniuyo.edu.ng

Abstract*— In this study, the development of an intelligent load control strategy for Itam distribution injection station using Supervisory Control and Data Acquisition (SCADA) systems with Machine Learning (ML) techniques is presented. Specifically, the focus is on the development of load control scheme for Itam distribution injection station based on load shedding strategy. The strategy essentially sheds off less priority loads when the total load demand is higher than available power in the substation. The load control strategy is developed by integrating machine learning model and supervisory control techniques called Supervisory Control and Data Acquisition (SCADA) system. The dataset used in the study consisted of time-stamped voltage readings for each phase (R, Y, B) and the corresponding load values (in MW) for the MBAK and IDORO injection stations. From the machine learning prediction results it is noticed that Idoro feeder peaks higher load for a longer period of the day. This load is more stable at least average not below 1.5MW and peak load above 2.5MW. Also, from the graph, the Idoro feeder has higher load within the hour of 12mindnight to 9am and 3pm down suggesting that the feeder feeds the residential area with higher consumption at the close of business. Furthermore, the results show that the Linear Regression model is the best performing model. It has the lowest error values across all three performance metrics used. In any case, although it performs the best, the Linear Regression model is most suited where the relationship between features and load is linear and hence may not be suitable in the situations where there is nonlinear relationship.

Keywords*— *Intelligent Load Control Strategy, Distribution Injection Station, Supervisory Control and Data Acquisition (SCADA) Systems, Linear Regression (LR), Gradient Boosting Regression (GBR) model

1. INTRODUCTION

The power distribution sector is a critical infrastructure for delivering quality and reliable electricity to consumers [1,2,3]. However, conventional distribution systems face several challenges, including high losses, inefficient metering and billing, reliance on analogue data collection, outdated infrastructure, weak and unreliable networks [4,5]. These challenges hinder the sector's ability to meet consumer demands effectively. As electricity demand grows and operational complexities increase, there is a pressing need to integrate advanced technologies and automation into distribution utilities [6,7].

Efficient and reliable power supply is the primary responsibility of injection substations, which act as crucial nodes in the distribution network. Reliability of power system is a crucial factor in the power industry [8,9]. These distribution substations oversee supervision, monitoring, measurement, protection, and control of power delivery to end users. Traditionally, these tasks have been handled manually, but manual approaches are prone to errors, time inefficiencies, and safety risks. To overcome these limitations, it is imperative to leverage modern advances in automation, intelligent systems, and computer control [10]. This system uses a controller for real time monitoring and control of the system. Controllers being the brain of automation are robust electronic device made of integrated chips designed to manage, coordinate and govern the behavior of other devices or systems [11]. They accept input data from sensors, execute pre-programmed algorithms to give a satisfactory output signal to the output devices for desired action. Depending on the type of signal, there are analog and digital controllers but the advent of analog to digital converter and digital to analog converter has further widen the useability of controllers irrespective of type of signal available at the input [12,13].

One of the most effective solutions for achieving optimal operation, enhanced reliability, and intelligent control of power systems is to carefully select a load control strategy that could match performance to specification [14,15]. There are many types of load control strategies such as: rotating load control, direct load control, combination load control, dynamic tariff load control and load shedding strategy [16,17,18].

This research work adopts Load shedding strategy because it sheds off less priority loads when the total load demand is higher than available power in the substation and this is enhanced by the cooperation between machine learning model and supervisory control techniques called Supervisory Control and Data Acquisition (SCADA) systems [19,20]. Supervisory control is a type of control strategy whereby field devices are scanned at regular intervals by a control processor and updates applied to the central control algorithm for updates [31]. SCADA systems enable immediate implication of event as they occur with minimum time lag. This real-time monitoring, data acquisition, and remote control of substations offer capabilities that surpass human effort in terms of coordination, facilitation, and safety. SCADA's telemetry and supervisory functions are particularly valuable for large coverage areas, where they ensure reliable operations and fault management through continuous data logging, alarms, and system protection [22].

2. METHODOLOGY

In this work, the focus is on the development of load control scheme for Itam distribution injection station based on load shedding strategy. The strategy essentially sheds off less priority loads when the total load demand is higher than available power in the substation. The load control strategy is developed by integrating machine learning model and supervisory control techniques called Supervisory Control and Data Acquisition (SCADA) system. The SCADA system scans field devices at regular intervals and updates the system for appropriate control actions [21]. The flowchart for the research procedure is shown in Figure 1.

The Data collection

The research procedure started with data collection. In this case, requisite data is collected from the Itam distribution injection station logbook. The data collection approach ensures that the data quality is guaranteed through validation checks.

The Data Preprocessing

At this stage, the dataset is cleaned by removing duplicates and addressing missing values. Also, derived features such as daily load, logged features example load from previous hours/days are generated. In addition, normalize the features as needed for specific models is carried out.

The Model Development (model training and evaluation)

The following machine learning models were used; Linear Regression (LR), and Gradient Boosting Regression (GBR) model. Each of the machine learning (ML) algorithms was trained and validated using 70% of dataset for training, and 30% for validation. Also, the model performance results were analysed to select the best-performing model. The performance of the models was evaluated using error metrics such as: MAE (Mean Absolute Error), MSE (Mean Squared Error), and RMSE (Root Mean Squared Error).

Specifically, the regression algorithms were employed to predict the load (in MW) of two injection stations in the case study Itam distribution injection station: MBAK and IDORO, based on their three-phase voltage readings (R, Y, B).

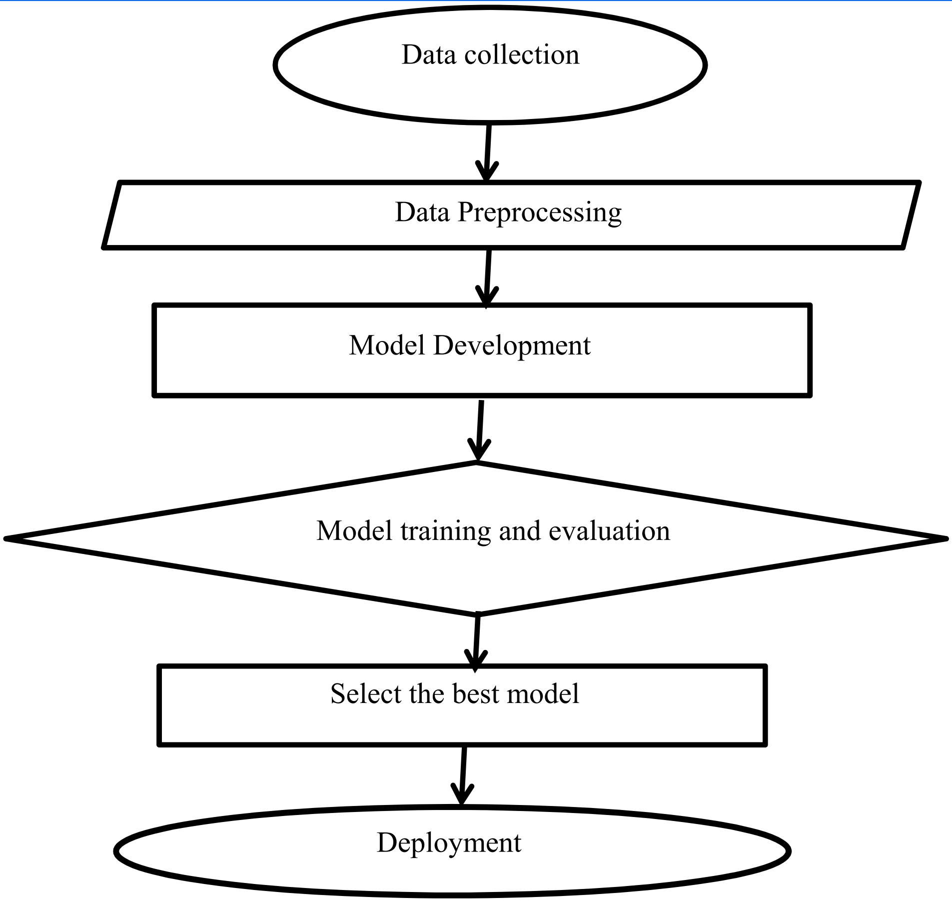

Figure 1: The flow diagram of the research procedure

The Deployment

At this stage the trained model (specifically the best performing ML model) is embedded in the executable scripts for real-time load prediction into the Programmable Logic Controller (PLC). In operation, the ML model is used to compute the maximum load for each feeder and update it in the data sheet.

Notably, in order to effectively monitor and control the substation remotely, the work used feedback control approach where field data are collected and communicated to and from PLC. The PLC's CPU stores and processes program data, but input and output modules connect the PLC to the rest of the machine; these I/O modules are what provide information to the CPU and trigger specific results. I/O can be either analog or digital; input devices might include sensors, switches, and meters, while outputs might include relays, lights, valves, and drives. Users can mix and match a PLC's I/O in order to get the right configuration.

When the Programmable Logic Controller (PLC) picks the updated load values it compares it to the maximum available load of 4.5MW. The PLC will then determine which feeder should be closed based on priority and highest load demand. Then the PLC will update the SCADA system and the HMI for prompt and effective monitoring, controlling and data logging.

In order to give a clear methodology of how signal flow from one stage to another, block diagrams are used for these illustrations as shown in different Figure 2, Figure 3 and Figure 4. The results show that the Linear Regression model is the best performing model. It has the lowest error values across all three metrics used. In any case, it assumes a linear relationship which may not hold in complex load behavior data. It is therefore useful as a baseline for comparison. Although it performs the best, it is most suited where the relationship between features and load is linear.

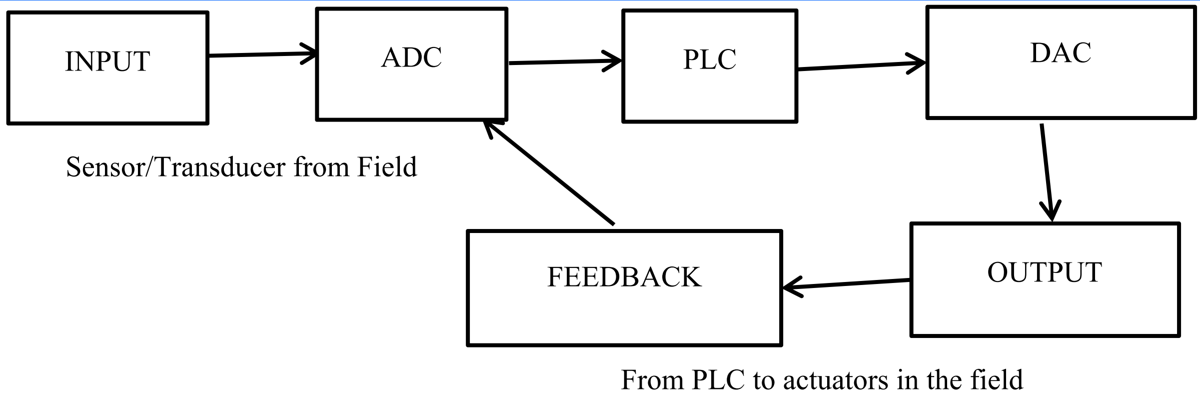

Figure 2: A typical feedback control model used in the study

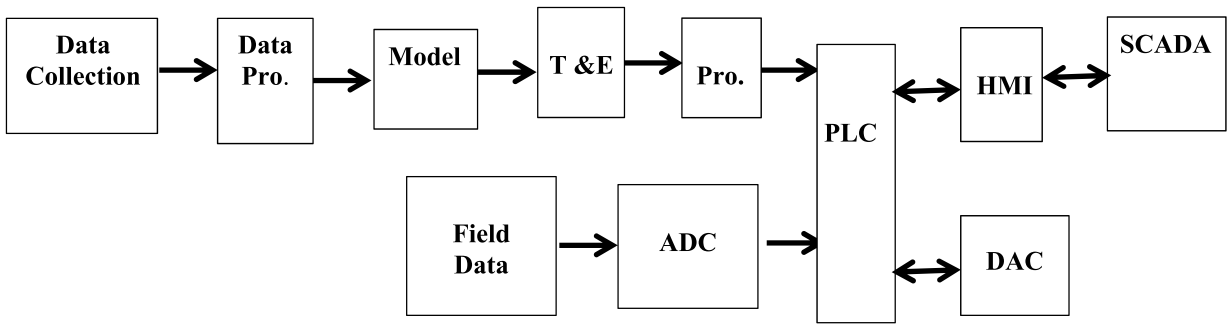

Figure 3: Block diagram illustrating the integration of the machine learning (ML) algorithms **to SCADA system.**

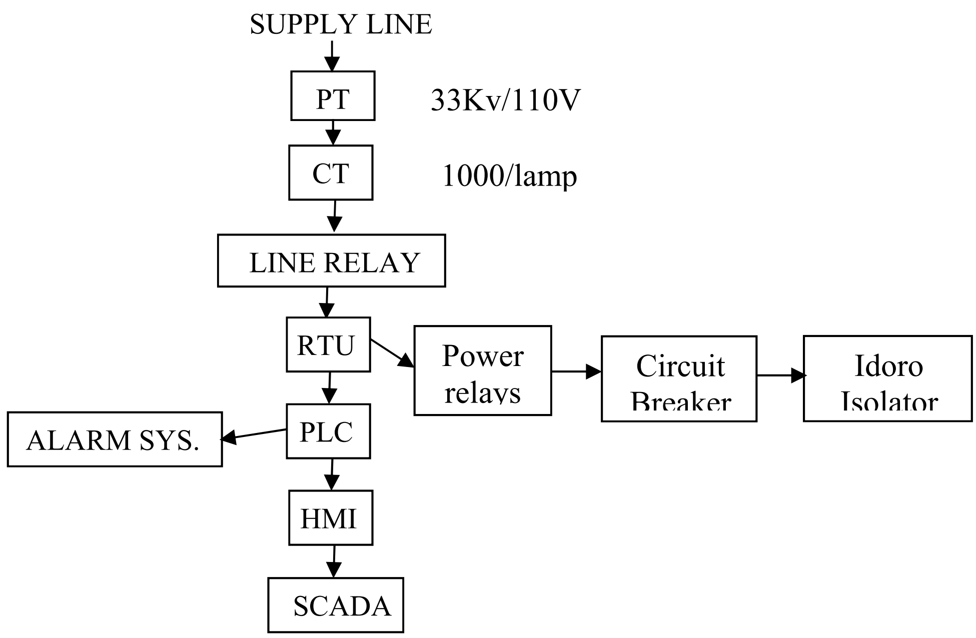

Figure 4: A block diagram illustrating the connection of component from the field devices to the central control room.

3. RESULTS AND DISCUSSION

The dataset used in the study consisted of time-stamped voltage readings for each phase (R, Y, B) and the corresponding load values (in MW) for the MBAK and IDORO injection stations. The graphs of actual and predicted Idoro feeder load for the Gradient Boost Regressor model is presented in Figure 5 while the one for Mbak feeder is presented in Figure 6. The graphs of actual and predicted Idoro feeder load for the Gradient Boost Regressor model is presented in Figure 5 while the one for Mbak feeder is presented in Figure 6. In addition, The performance evaluation results for the two machine learning load prediction models on the MBAK feeder or station are presented in Table 1 while the results for the Idoro feeder

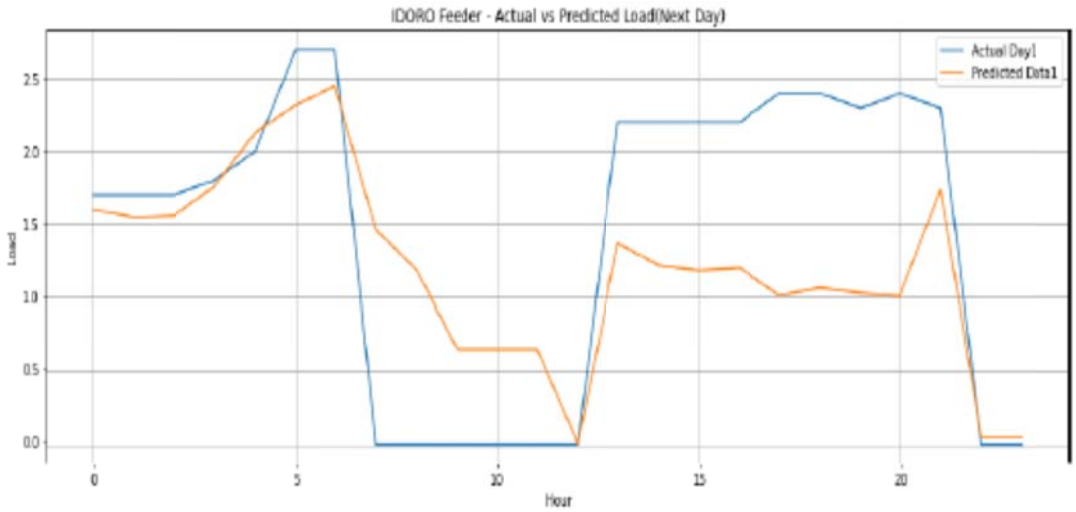

Figure 5: Graphs of actual and predicted Idoro feeder load for the gradient boost regressor model

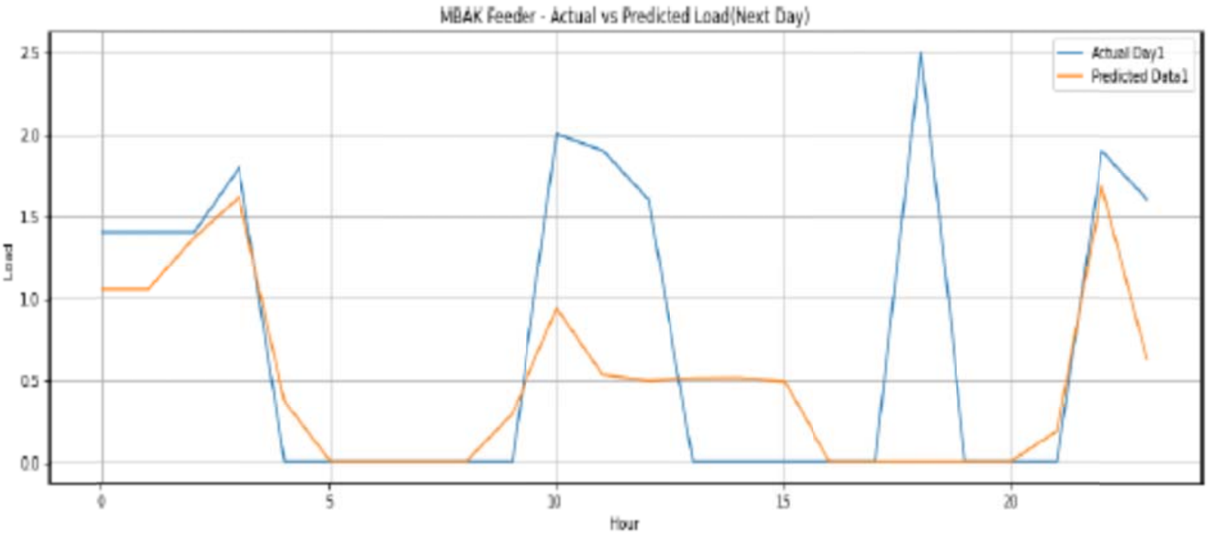

Figure 6: Graphs of actual and predicted Mbak feeder load for the gradient boost regressor model

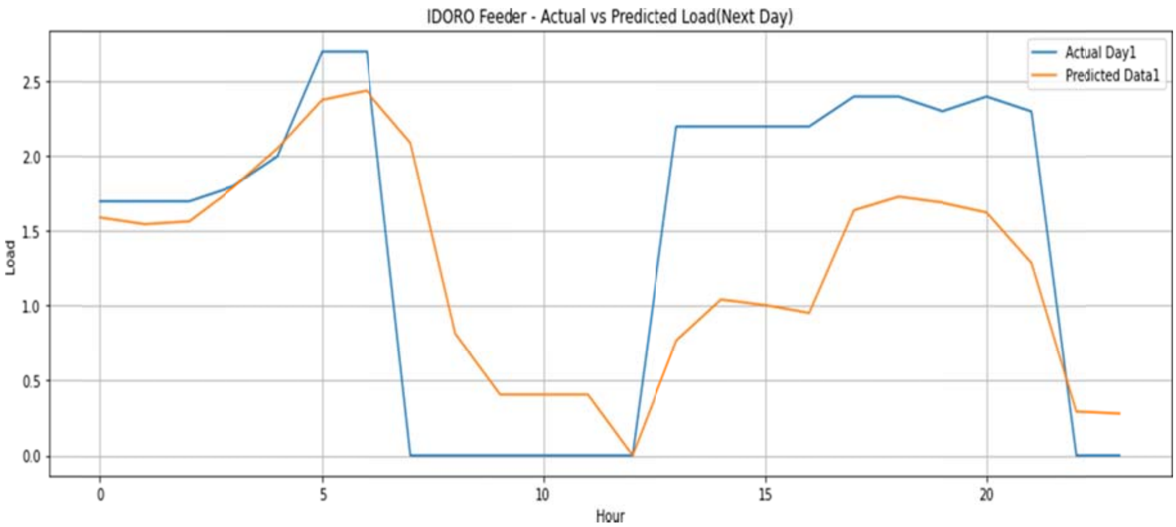

Figure 7: Graphs of actual and predicted Idoro feeder load for the linear regression model

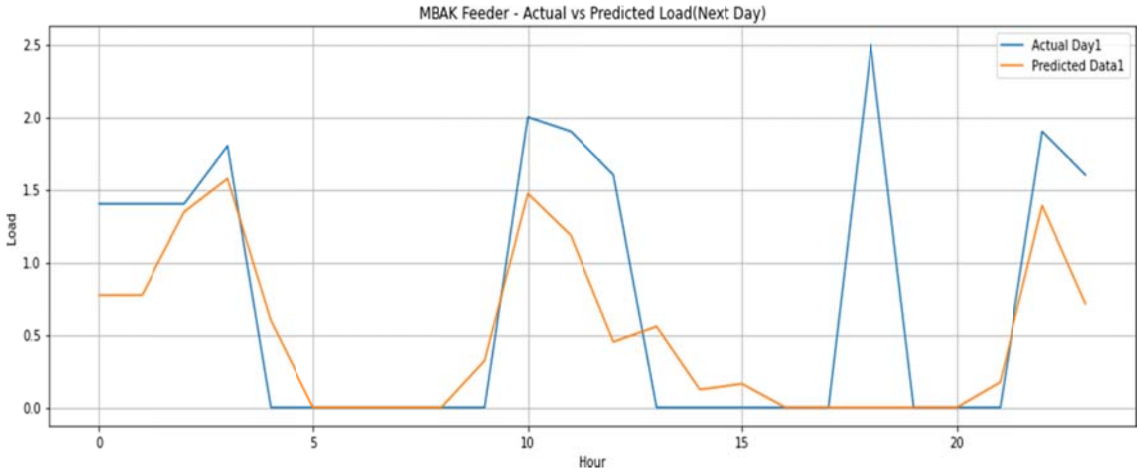

Figure 8: Graphs of actual and predicted Mbak feeder load for the linear regression model

Table 1 The performance of the machine learning load prediction models on the MBAK Station

Model	MAE	MSE	RMSE
Linear Regression	0.4068	0.4623	0.2137
Gradient Boosting Regr.	0.4372	0.5319	0.2829

Table 2 The performance of the machine learning load prediction models on the Idoro Station

Model	MAE	MSE	RMSE
Linear Regression	1.2315	0.4692	0.2201
Gradient Boosting Regr.	1.2301	0.6832	0.04668

From the machine learning prediction results it is noticed that Idoro feeder peaks higher load for a longer period of the day. This load is more stable at least average not below 1.5MW and peak load above 2.5MW. Also, from the graph, the Idoro feeder has higher load within the hour of 12mindnight to 9am and 3pm down suggesting that the feeder feeds the residential with higher consumption at the close of business. These analyses suggest that Idoro feeder is the priority load for the next day. Comparatively, Mbak feeder load is less stable with lower load demand suggesting that it will be less prioritized for next day loading schedule.

The results obtained from the simulation and analysis demonstrate the significant advantages of incorporating machine learning (ML) techniques into the control strategy of a SCADA-powered injection station. The performance metrics—including response time, control accuracy, and energy consumption—clearly show that the proposed ML-based model outperforms conventional control strategies such as Proportional-Integral-Derivative (PID) and rule-based control methods. One of the most notable improvements observed was in response time. The ML-based controller adjusted to varying process conditions faster than the traditional PID controller. This improvement is attributed to the predictive nature of the ML algorithm, which anticipates changes in process variables (such as inlet pressure and flow rate) and adjusts the control signals proactively rather than reactively. In contrast, PID controllers depend solely on error feedback, which introduces time lags before corrective actions are implemented. This demonstrates the ability of ML-based systems to operate in predictive mode, enhancing process stability and minimizing oscillations.

Another key observation is the significant increase in control accuracy. The ML model, trained on historical process data, was able to capture nonlinear and complex relationships between system variables—something that traditional controllers struggle with. Industrial processes such as fluid injection often exhibit nonlinearities due to temperature variations, valve dynamics, and pressure fluctuations. The machine learning model, especially the neural network used in this study, successfully modeled these nonlinearities, leading to improved accuracy and smoother control transitions. This capability makes ML particularly advantageous for applications where process conditions are not constant or easily modeled by analytical equations.

Energy efficiency also improved markedly under ML-based control. The reduction in pump energy consumption and over-injection events indicates that the system operated closer to optimal conditions. Because the ML model learns to predict the exact amount of injection needed to maintain target pressures or concentrations, it minimizes waste and excessive control actions. This contributes not only to operational cost savings but also to sustainability by reducing unnecessary energy usage. Furthermore, the ML-based control system exhibited robustness under varying operational disturbances, such as sudden fluctuations in inlet pressure or changes in fluid viscosity. During simulation, the controller maintained steady operation with minimal deviation from setpoints, demonstrating its adaptability. The continuous learning or retraining capability of ML models means that system performance can improve over time as more data becomes available. This adaptability is a distinct advantage over static control logic that requires manual tuning or reconfiguration when process parameters change.

From a systems integration perspective, the incorporation of the ML model within the SCADA environment proved feasible and efficient. The model was able to receive real-time data from the SCADA server, process it, and transmit control signals to actuators with minimal latency. This demonstrates the potential for real-time deployment of ML models within industrial SCADA frameworks. However, the computational demand of the ML model must be considered—particularly when deploying on low-power or edge devices. In the study, model optimization techniques such as model pruning and feature selection were used to ensure the model ran efficiently without compromising performance. In terms of system reliability and safety, the integration of ML introduces both benefits and challenges. The benefit lies in improved fault tolerance; the ML model can detect anomalies or deviations that traditional systems may overlook. However, challenges include ensuring that the model's predictions remain accurate in the presence of unseen data or sensor faults. Thus, implementing fallback mechanisms (such as reverting to PID control during ML failure) is recommended for industrial deployment to ensure safety and continuity of operations.

Comparatively, the rule-based SCADA logic performed adequately under steady-state conditions but showed limitations during transient scenarios. It lacked the flexibility to adjust to process variations, resulting in less efficient control. This further supports the argument that purely rule-based or linear control systems are insufficient for modern, data-intensive industrial environments. Overall, the discussion confirms that the proposed ML-based control model significantly enhances the performance of SCADA-powered injection stations. The integration of predictive analytics into the control loop allows for adaptive, data-driven decision-making, improving both operational efficiency and process reliability. The findings align with previous studies in intelligent process control, reinforcing the growing consensus that machine learning will play a central role in the next generation of industrial automation systems.

Nevertheless, practical implementation requires attention to several factors. These include the availability and quality of training data, real-time computational capability, cybersecurity considerations, and integration with existing PLC and SCADA protocols. Addressing these factors will ensure that ML-based control systems are not only technically sound but also safe, secure, and compliant with industrial standards. In summary, the results and discussion underscore the transformative potential of machine learning in industrial control applications. The proposed approach advances the state of SCADA automation by moving from static, reactive control strategies toward intelligent, adaptive, and self-optimizing systems—laying the groundwork for the realization of Industry 4.0 and autonomous industrial control networks.

4. CONCLUSION

The research work presented in this dissertation has explored the modelling and implementation of a load control strategy within a SCADA-powered distribution injection station, using machine learning techniques to enhance monitoring, prediction, and decision-making capabilities in the power distribution network. A comprehensive analysis of the existing challenges in load management and power distribution inefficiencies informed the motivation for this study. By integrating machine learning algorithms with SCADA system architecture, a robust framework was developed for real-time load prediction, fault detection, and intelligent control of distributed energy flows.

REFERENCES

1. Gönen, T., Ten, C. W., & Mehrizi-Sani, A. (2024). *Electric power distribution engineering*. CRC press.
2. Azmi, K. H. M., Radzi, N. A. M., Azhar, N. A., Samidi, F. S., Zulkifli, I. T., & Zainal, A. M. (2022). Active electric distribution network: applications, challenges, and opportunities. *Ieee Access*, *10*, 134655-134689.
3. Prettico, G., Marinopoulos, A., & Vitiello, S. (2022). Guiding electricity distribution system investments to improve service quality: A European study. *Utilities Policy*, *77*, 101381.

4. Aghahadi, M., Bosisio, A., Merlo, M., Berizzi, A., Pegoiani, A., & Forciniti, S. (2024). Digitalization processes in distribution grids: a comprehensive review of strategies and challenges. *Applied Sciences*, *14*(11), 4528.
5. Gallegos, J., Arévalo, P., Montaleza, C., & Jurado, F. (2024). Sustainable electrification—advances and challenges in electrical-distribution networks: a review. *Sustainability*, *16*(2), 698.
6. Arévalo, P., & Jurado, F. (2024). Impact of artificial intelligence on the planning and operation of distributed energy systems in smart grids. *Energies*, *17*(17), 4501.
7. Arent, D. J., Green, P., Abdullah, Z., Barnes, T., Bauer, S., Bernstein, A., ... & Turchi, C. (2022). Challenges and opportunities in decarbonizing the US energy system. *Renewable and Sustainable Energy Reviews*, *169*, 112939.
8. Khismatullin, A. S., & Bashirov, M. G. (2021, June). Methods of improving the power supply reliability of industrial site. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1155, No. 1, p. 012067). IOP Publishing.
9. Gönen, T., Ten, C. W., & Mehrizi-Sani, A. (2024). *Electric power distribution engineering*. CRC press.
10. Atieh, A. A., Abu Hussein, A., Al-Jaghoub, S., Alheet, A. F., & Attiany, M. (2025). The impact of digital technology, automation, and data integration on supply chain performance: exploring the moderating role of digital transformation. *Logistics*, *9*(1), 11.
11. Cirstea, M., Benkrid, K., Dinu, A., Ghiriti, R., & Petreus, D. (2024). Digital electronic system-on-chip design: Methodologies, tools, evolution, and trends. *Micromachines*, *15*(2), 247.
12. Waho, T. (2022). *Introduction to Analog-to-Digital Converters*. River Publishers.
13. Athur, M., Joan, A., Emmanuel, M., & Buhari, M. D. (2024). A Review in Advanced Digital Signal Processing Systems. *Digital Signal Processing (DSP)*.
14. Alhamrouni, I., Abdul Kahar, N. H., Salem, M., Swadi, M., Zahroui, Y., Kadhim, D. J., ... & Alhuyi Nazari, M. (2024). A comprehensive review on the role of artificial intelligence in power system stability, control, and protection: Insights and future directions. *Applied Sciences*, *14*(14), 6214.
15. Behara, R. K., & Saha, A. K. (2022). artificial intelligence control system applied in smart grid integrated doubly fed induction generator-based wind turbine: A review. *Energies*, *15*(17), 6488.
16. Anderson, B., Reilly, J., & Krishnan, V. (2022). Load control for frequency response-a literature review.
17. Ozkop, E. (2024). A Survey on Direct Load Control Technologies in the Smart Grid. *IEEE Access*, *12*, 4997-5053.
18. Zhang, J., Wang, J., Cao, Y., Li, B., & Li, C. (2024). A coordinated control strategy of multi-type flexible resources and under-frequency load shedding for active power balance. *Symmetry*, *16*(4), 479.
19. Mortaji, H., Ow, S. H., Moghavvemi, M., & Almurib, H. A. F. (2017). Load shedding and smart-direct load control using internet of things in smart grid demand response management. *IEEE Transactions on Industry Applications*, *53*(6), 5155-5163.
20. Jindal, A., Singh, M., & Kumar, N. (2018). Consumption-aware data analytical demand response scheme for peak load reduction in smart grid. *IEEE Transactions on Industrial Electronics*, *65*(11), 8993-9004.
21. Itaketo, Umama T. (2016). Process control Engineering (with worked examples). (first Ed.) Scholar's Press.
22. Raghunandan, K. (2022). Supervisory control and data acquisition (SCADA). In *Introduction to Wireless Communications and Networks: A Practical Perspective* (pp. 321-337). Cham: Springer International Publishing.