Anomaly Forecasting in High Voltage Standing Wave Radio-Frequency Cavities

Abdullah Burkan Bereketoglu
Department of Physics
Middle East Technical University
Ankara, Turkey
burkan.bereketoglu@metu.edu.tr

***Abstract*—In High-Energy Physics (HEP), the calibration of devices requiring high voltage uses small and slow steps as high voltage gradients to prevent damage from non-periodic RF signal buildup, which causes heterogeneous and nonlinear wave behaviors leading to device faults. The systems at various laboratories, even with similar devices requires localized solutions with current-day solution techniques. This paper addresses the calibration of bunching cavities by fundamental parameters of preparation to high voltage levels, focusing on adjusting such parameters, and observing the outputs. To manage the non-isothermal and irreversible energy increases, we propose using a Dual Attention Recurrent Neural Network (DA-RNN), an advanced version of the Nonlinear Auto-Regressive Exogenous Model (NARX). This model captures key temporal and input patterns using LSTM-based attention mechanisms, aiming to achieve a zero-fault, efficient and fast calibration process without supervision. The preliminary results and further experimentation with real closed-loop data and toy RF Cavity model simulated, shows potential to eliminate the localized solutions via a generalized use-model of DA-RNN. Experiments have shown that smaller neuron size with smaller windows are more effective on generalization of the overall structure and prevention of overfitting.**

***Index Terms*—Attention Mechanism, Exogeneous Modeling, LSTM, Radio-Frequency, Anomaly Forecasting, Recurrent Network, Dual-Stage Attention-Based Recurrent Neural Networks, Simulation, Buncher Cavities, High Energy Physics**

I. INTRODUCTION

In High-Energy Physics (HEP), bunching cavities and instruments to be used with alternating current in high voltages are equipped with controllers having strict control limits and small-slow steps in training periods for to calibrate and adapt devices without catastrophic damage both electrically and mechanically to occur to the device itself via field emissions due to RF signal non-homogeneity under small periods. These non-homogeneities lead to heterogeneous, non-stationary, and nonlinear wave behaviors that are highly unpredictable and chaotic, hence hard to control after reaching certain accumulation limits. These waves will then result in various microscopic and macroscopic faults at the interface or surface levels. Moreover, these instruments are designed for each laboratory, hence requires individual engineering at each local laboratory. This leads to local solutions rather than a universal one. However, inner dynamics of bunching cavities and instruments, regardless of their high-voltage level can be characterized and learned by predictive models, meaning that a universal solution model can be made to be trained in simulation to be used for faster, reliable control and zero-fault system making. Furthermore, such optimization of learning the characteristics through machine learning is applicable via suitable nonlinear exogenous models are applied as learners [1].

II. PROBLEM DEFINITION

Tuning/Calibration of High Voltage RF bunching cavities used in accelerators involve adjustment of various parameters, as pulse width, power injected (dBm), and pulse period. Here, we observe temperature and pressure as outputs, where temperature is directly correlated according to ohm's law and wave theory with increased pulse width and decreased pulse period. This temperature increase is a result of irreversible and non-isothermal nature of the pulse's total energy increase at each tuning step. Thermodynamically, more heat dissipation occurs, and the relation

$$Absorption - Emission = Accumulation \tag{1}$$

implies that higher energy levels and energy gradients over time result in higher temperatures. Consequently, higher temperatures and EM-fields, due to higher frequencies and wider waves with higher amplitudes, can lead to a sudden pressure increase due to field emissions [4]. This can cause electrical sparks on the cavity's surface or interface, causing significant damage due to geometrical or electrical failures [3], [6].

Heterogeneous wave behavior can be tolerated, and failures prevented if the time period for updating parameters (pulse width, pulse period, power injected) is extended, allowing equilibrium where the wave adjusts to the geometry and behaves as a standing wave. However, the high time complexity and limited time availability make this impractical, but an optimal time point may be feasible for rapid tuning.

Additionally, not waiting for appropriate periods in one time step changes the required waiting time in subsequent steps, and even may damage or lock in the current state, also necessitates an optimized model to prevent device damage and minimize waiting time, on the contrary a strict controller with periodic manual monitoring can also be used, but always requires human force and prone to human error. Even with all precautions are held appropriately, the nonlinear nature of the case leads to no guarantee, zero-fault progress without strict restrictions, which leads to slower-than-required process,

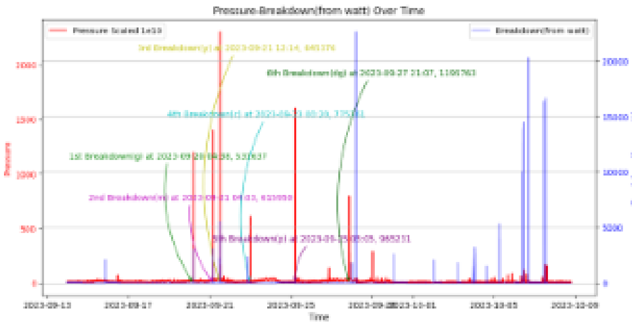

Fig. 1. Experimental Data [5]

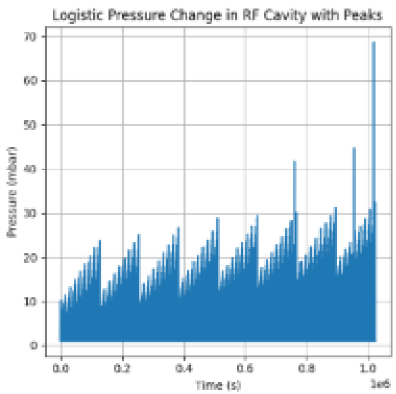

Fig. 2. Simulated Data

if artificial intelligence forecasters are not used for dynamical controlling.

III. Dataset

For this study, we utilized two types of datasets. The first dataset is a preprocessed closed-loop RF cavity data [2], experimentally collected and prepared by the TARLA-FEL beam dynamics group [5] and is shown in figure 1. This data is highly effective in detecting anomalies. However, for forecasting purposes, it lacks sufficient data due to the nature of the closed-loop system, necessitating the creation of a simulated dataset which is shown in figure 2-3.

The second dataset is a toy RF model that mimics the expected behavior of fundamental parameters and variables in the system, and their effects on the output. The simulated data features logistic growth as its excitation state, with updated gradients and tolerances at different voltages, frequencies, and widths. It also includes an exponential relaxation part, completing the toy semi-time independent RF model. This model is referred to as semi-time independent because the gradients are continuously checked for anomaly creation. At shorter periods with higher gradients, more anomalies are observed, whereas, with longer relaxation and excitation periods, fewer peaks are detected. Moreover, a general min-max scaler for each feature is used to scale down data to 0-1 and preserve the differences in the sequence.

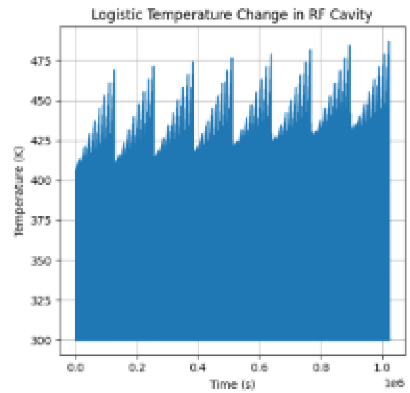

Fig. 3. Simulated Data

IV. Model & Training

For model training, we employed the Dual Stage Attention Based Recurrent Neural Network (DA-RNN) as proposed by Qin et al. [1]. This model includes two attention layers: one at the encoder state, which creates a context vector relating the encoder hidden state to select input features at each time step and is shown in Figure 4 .The second attention mechanism which is temporal attention relates the decoder hidden state with the encoder hidden state to select relevant encoder hidden states across all time steps, which is selection of temporal preferences from the temporal attention context vector using the decoder hidden state and encoder hidden state generated new inputs, is shown if Figure 5. Also two different python based libraries used for to find the best training approach, PyTorch for the desired data training strategy, which is initializing each file with the previous weights from the last epoch of the last file trained, to focus better to the valued parts of the problem in the training.

The DA-RNN model can rapidly overfit, but its efficiency in generalizing with fewer trainable parameters makes it a highly effective sequential model. During training, different methodologies were tested. Initially, excitation and relaxation phases were given independently, but this approach failed to

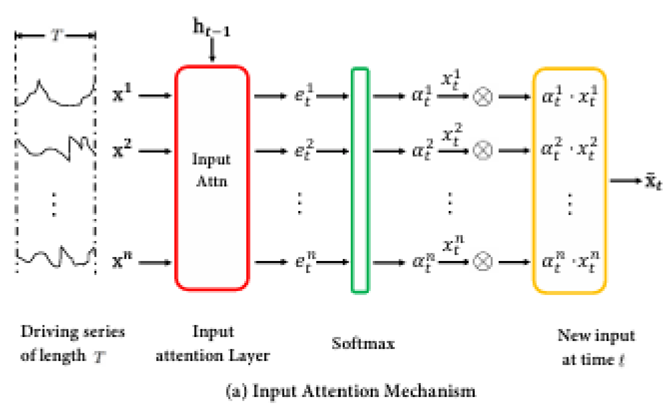

Fig. 4. Input Attention Mechanism [1]

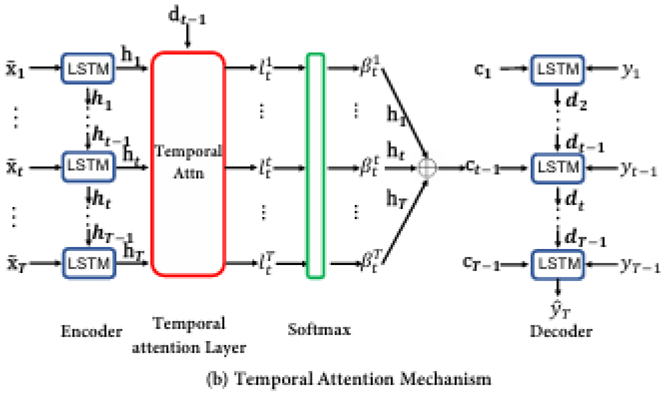

Fig. 5. Temporal Attention Mechanism [1]

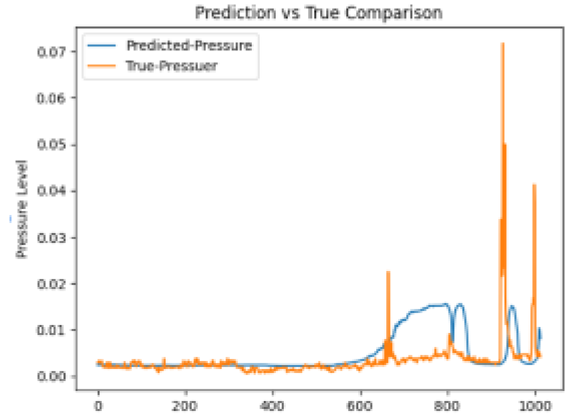

Fig. 6. Prediction vs True Comparison for Experimental Data

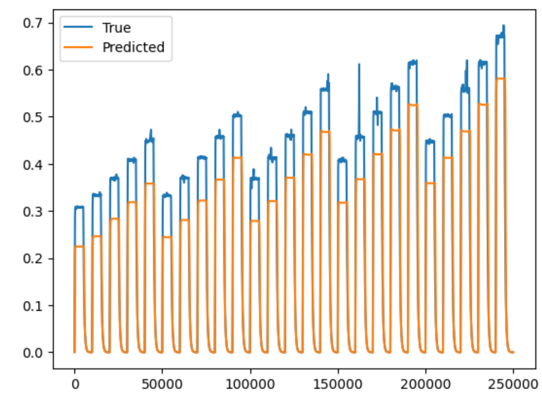

Fig. 7. Prediction vs True Comparison for Simulation Model Data, hidden size 200

capture the behavior. Subsequently, the model was trained on the full dataset, which also failed to explain the behavior. Finally, by passing a single full parameter set, the DA-RNN model performed satisfactorily. It was observed that increasing the number of neurons led to overfitting on the training set, while a smaller window size, and hence less introduction to the training set patters increases the overall generalization and reduces dead neuron counts due to continued learning periods.

V. Results

In our experiments, we analyzed the model's performance in forecasting anomalies in the test data of experiments and also made a toy RF model to make simulation data and analyzed the model in the simulation data as well. The evaluation of the model was based on Mean Absolute Percent Error (MAPE) due to its sensitivity to anomalies.

A. Experimental Data

The first set of results is based on experimental data. As depicted in Figure 6, the model struggles to predict anomalies accurately, resulting in a high MAPE. The predictions emphasize the presence of discrepancies but do not capture the peaks effectively. Here it also indicates that in closed-loop systems, due to lack of representative data, it is hard do detect the patterns to forecast.

The hidden layer sizes for the experimental model were set to 200 units and 100 units in the attention layers. Despite the model's complexity, it overfit rapidly and failed to learn beyond the first epoch.

B. Simulation Model Data

To mitigate the overfitting issue, we reduced the hidden layer sizes and attention units in the simulation model. The revised model used hidden sizes of 100, 50, 30, and 20 units, respectively. First, to be able to compare with the experimental dataset predictions, same inputs are used as in 7, which shows overfit to the training itself. As shown in Figure 8, this adjustment led to improved generalization, with both validation and test losses decreasing while the training loss remained stable.

Moreover, reducing the window size (temporal size) in training data conjunction with smaller hidden and attention units further enhanced generalization, though overfitting issues persisted. However, we have not yet determined the optimal point where the MAPE loss stabilizes without overfitting.

C. Summary of Findings

- **High MAPE in Experimental Data**: The model's inability to predict anomalies, due to lack of non-misrepresented data, resulted in a high MAPE.
- **Overfitting in Larger Models**: Models with larger hidden sizes overfit quickly for the structure and did not show improvement after the first epoch.
- **Better Performance - Smaller Models**: Reducing hidden sizes and window sizes in the simulation model led to better generalization and reduced validation and test losses, and learning continued after the first epoch if made smaller.

VI. Discussion

The experimental and simulation results presented in this study highlight several critical observations and insights into the tuning of bunching cavities using the Dual Attention Recurrent Neural Network (DA-RNN) model.

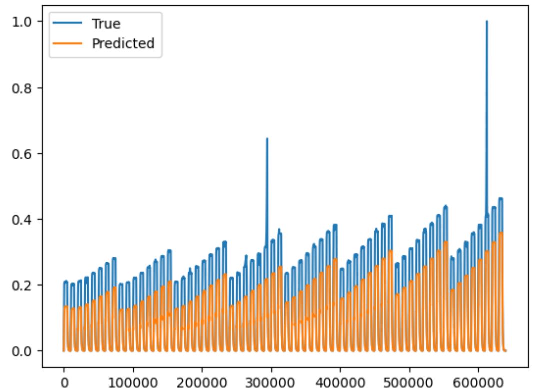

Fig. 8. Prediction vs True Comparison for Simulation Model Data, Hidden Size 20

A. Analysis of Results

The high MAPE observed in the experimental data underscores the challenges associated with predicting anomalies in real-world closed-loop systems, due to a currently running feedback system already working inside some pre-defined control boundaries to prevent incidents. These systems often lack sufficient representative data, due to producing white noise for the problem as it stabilizes the outputs, making it difficult for models to learn and forecast patterns accurately. Despite the complexity of the model with larger hidden sizes, it overfit rapidly, indicating that the model was too complex for the available data, which also comes from the misrepresentation of the system due to controlled environment.

In contrast, the simulation model data demonstrated that reducing the hidden layer sizes and attention units significantly improved the model's generalization ability. This reduction somewhat decreased overfitting levels and resulted in decreased validation and test losses, while having similar training losses. This finding suggests that simpler models may be more effective in capturing the underlying dynamics of the system without overfitting. Hence, feasible and easier model to be used which is much more dynamic to use.

The adjustment of window sizes (temporal size) on par with smaller hidden and attention units further enhanced generalization. However, overfitting issues continued, indicating that there is still room for optimization in model configuration. The results suggest that finding the optimal balance between model complexity and training data representation is crucial for effective anomaly forecasting.

B. Practical Implications

The findings of this study have practical implications for the calibration of high-energy physics instruments under High Voltage with RF systems. By using DA-RNN model [1] with appropriately sized hidden layers and attention units, it is possible to achieve more reliable and efficient calibration-training processes. This approach can potentially eliminate the need for localized, uniquely designed, engineered, solutions by providing a generalized model that can be trained on simulation data and applied across different laboratories.

The ability to forecast anomalies accurately and prevent device faults is particularly valuable in high-energy physics, where equipment damage can be costly and time-consuming to repair. The proposed model offers a promising solution for achieving zero-fault calibration and efficient device management.

C. Limitations

One limitation of this study is the reliance on simulated data to supplement the experimental data, even though it is defined as a mere toy RF model, although promising. While the simulated data provides valuable insights, it may not capture all the complexities and variabilities of real-world systems, such as time dependent relaxation and gradient differences, which require much more complex simulation design. Additionally, the current model configuration is still prone to overfitting, and further optimization is needed to find the optimal balance between model complexity and generalization, where this will be a focus of the future work to be done.

VII. Future Work

Future work will focus on refining the model to effectively predict anomalies without overfitting. This will involve exploring different model configurations, including further reductions in hidden layer sizes and adjustments to the temporal window sizes. Additionally, the development of a time-dependent version of the toy RF model, based on Cheng's Encyclopedia of RF and Microwave Engineering, will be pursued to enhance the model's accuracy and applicability.

By continuing to optimize the model and incorporating more representative data, we aim to achieve a robust and generalized solution for the calibration of bunching cavities and other high-energy physics instruments.

References

[1] Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., and Cottrell, G., "A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction", ¡i¿arXiv e-prints¡/i¿, 2017. doi:10.48550/arXiv.1704.02971

[2] Fabio Nelli, Python Data Analytics : Data Analysis and Science Using Pandas, matplotlib, and the Python Programming Language. Berkeley, Ca: Apress, 2015.

[3] F. Valerio-Lopez and F. Valerio-López, "Electron gun high voltage conditioning data analysis."

[4] "TARLA - Redmine," Tarla-fel.org, 2022. https://redmine.tarla-fel.org/attachments/download/8703/Faz-2/20Rapor.pdf (accessed Jun. 12, 2024).

[5] "Parçacık Hızlandırıcıları ve Algıçları Yerel Altyapı ve Ar-Ge Çalıştayı," Indico. https://indico.cern.ch/event/1330502/contributions/5689933/ (accessed Jun. 12, 2024).

[6] A. Hassanein, Z. Insepov, J. Norem, A. Moretti, Z. Qian, A. Bross, Y. Torun, R. Rimmer, D. Li, M. Zisman, D. N. Seidman, and K. E. Yoon, "Effects of surface damage on rf cavity operation," Phys. Rev. ST Accel. Beams, vol. 9, no. 6, pp. 062001, Jun. 2006. doi: 10.1103/PhysRevSTAB.9.062001.