LSTM-based Wastewater Treatment Plant Sensor Failure Forecasting

1st Miguel Silva
Dep. of Mathematics (DMat)
University of Aveiro (UA)
Aveiro, Portugal
mig.silva@ua.pt

2nd Eero Jormalainen
Dep. of Economics, Management, Industrial Eng. and Tourism (DEGEIT)
University of Aveiro (UA)
Aveiro, Portugal
eeroj@ua.pt

***Abstract*—In this study, we propose and Long Short Term Memory (LSTM) based model to predict Wastewater Treatment Plants' (WWTP) sensor failure. In particular, we focused on forecasting failure on measuring the quantity of Total Suspended Solids (TSS) based on weather variables over a period of 2 years. We performed feature selection using Random Forest Regression and manual selection by human expertise and developed 72 models to find the best LSTM parameters to the quantity of TSS with the respective sensor failures. After that, we have applied differentiation to the LSTM models' output to capture the abrupt changes of TSS values caused by sensor, encoded the time series with binary system for failure classification and evaluated the best models with adjusted parameters of accuracy, recall, precision and F1-score. The failure prediction performance ended being inadequate, since F1-score of the best models scored 25% while accuracy reached 84%.**

***Index Terms*—long short term memory, wastewater treatment plants, total suspended solids, failure detection**

I. Introduction

In wastewater treatment plants (WWTP), accurate monitoring and control of total suspended solids (TSS) are crucial for ensuring the quality of treated water. The plant employs a series of sensors to measure TSS at various stages: initially in the untreated water, aeration pools differerent treatment lines, post-aeration, and in the effluent water before discharge. Despite this setup, the sensors often experience failures, leading to erroneous spikes in TSS readings. These are caused by high stress in treatment system, where high loads of solids may disturb the sensor by sticking to the measurement platform which can cause the results to spike. These anomalies are particularly prevalent during significant weather changes, such as heavy rainfall or melting snow, which introduce large volumes of water into the system [1].

The frequent sensor malfunctions not only compromise the reliability of TSS measurements but also pose challenges for the operational efficiency of the plant. Understanding and predicting these sensor failures are vital for maintaining accurate water quality assessments and for the continuous improvement of plant operations.

Our primary objective is to develop a forecasting model resorting to Long Short Term Memory (LSTM) recurrent neural networks for TSS readings across these three sensor points over a span of two years, including the periods of sensor failures. With this strategy, we can predict when the sensors are going to fail.

II. Related Work

Very few studies are conducted about attempting to predict WWTP sensor failures using Deep Learning techniques. However Cheng [2] conducted a comprehensive study employing six deep learning models derived from LSTM and GRU architectures, when attempting to build data-driven soft sensors to aid with optimizing treatment performance. Their research included traditional LSTM and GRU, exponentially smoothed LSTM, and adaptive LSTM models. The study demonstrated that GRU-based models converged faster but producing higher RMSE of 160.67, while LSTM-based models still provided very decent accuracy with just a bit smaller RMSE 140.73. Furthermore, the study highlighted the benefits of incorporating exponential smoothing techniques to mitigate the effects of outliers and improve prediction accuracy.

More closer to our work, in 2023 Harrou [3] explored various deep learning models including RNN, LSTM, GRU, BiLSTM, BiGRU to predict the energy consumption of the plant based on sensor and weather data. These models were chosen for their ability to capture temporal dependencies and complex relationships in time-series data. Data cleaning was performed to eliminate outliers and ensure data quality and Cubic spline interpolation was employed to enhance the dataset by generating synthetic data points. This technique allowed for a denser representation of the WWTP variables, capturing finer temporal details and variations. The LSTM model trained on augmented data achieved the best performance with an MAPE of 1.36%, followed closely by the BiGRU model with an MAPE of 1.436%. Models trained on augmented data consistently outperformed those trained on non-augmented data. For instance, the LSTM model showed substantial improvements in RMSE (5.821 MWh) and MAE (3.685 MWh).

Outside of the field of WWTPs, Han et al. [4] have developed a work where they also aim to create a model that predict failures of a system. Their study focused on fault detection with LSTM-based variational autoencoder (LSTM-VAE) for maritime components. In their study, they have used the following features collected from a diesel engine operated on

Norwegian University of Science and Technology's research vessel Gunnerus:

- Boost pressure;
- Engine speed;
- Engine exhaust gas temperature 1;
- Engine exhaust gas temperature 2;
- Fuel rate;
- Lube oil pressure;
- Lube oil temperature;
- Engine power;
- Cooling water temperature.

These features were collected at a rate of 1Hz (one sample per second) in a period of 10 days (6 hours per day) in the month of November of 2019. On the 21st of that month the grup introduced a fault in the system by covering an air filter with cloth. Each day represented 25,000 samples, so they have reduced the daily data into segments of 1000 samples.

Han et al.'s LSTM-VAE architecture was designed to output hidden states from feature inputs, which are then processed through linear modules to estimate the mean and log-variance of the posterior distribution. This model is optimized by minimizing a loss function that combines Kullback-Leibler divergence [5] and mean squared error. For anomaly detection, instead of relying solely on reconstruction error, the methodology proposes using the reconstruction probability as the anomaly score, which is the variable to forecast, leveraging the stochastic nature of VAEs to account for the variability of the latent space. In the end, they have defined a constant threshold to define an instant or period as a failure of the system, which was $\mu + 3\sigma$, where μ and σ are the mean and standard deviation of anomaly score in the validation set. They normalized the data, used mini-batches of size 512 to train the LSTM-VAE, with Adam optimizer with learning rate of 1×10^{-3}, l_2 regularization of coefficient 1×10^{-3}, and 100 epoch. The performance metrics used were the time to detect (TTD), that measures the level of stability of the detection signal, and detection stability factor (DSF), that measures the stability of fault detection, as shown in Equations 1 and 2. In these equations t_{f1} and t_{f2} represent the instants when the failure was introduced and ended, respectively, and t_1, t_2, and t_3 represent instants when the anomaly score crossed the threshold line. When comparing with other models, DSF scored the highest value for the LSTM-VAE with 0.791 with TTD of 66 seconds. The group considered these values relevant and reveal that this model performs better in fault detection than other models.

$$TTD = t_1 - t_{f1} \tag{1}$$

$$DSF = \frac{(t_2 - t_1) + (t_{f2} - t_3)}{t_{f2} - t_{f1}} \tag{2}$$

III. Dataset

The dataset used in this project contains several features measured by sensors in some stages of water cleaning from a WWTP in Finland. This features include TSS (mg/l), ammonia (mg/l), and oxygen (mg/l), which are measured influent pool of the WWTP (incoming water), during the process of aeration, after the process of aeration, and when the water is leaving the WWTP (effluent water). In this project we are focusing on the quantity of TSS in incoming water, in water after aeration, and in effuent water. The data from the WWTP are collected with every hour from May 10_{th} of 2022 at 00:00 to March 21_{st} at 12:00, resulting in a total of 16357 samples. Figure 1 presents the measured values of TSS over time in the three sensors of study.

Along with the data from the WWTP, we have included the meteorological data of the same period from the area the plant is located, also collected every hour. The variables considered for this project, after elimination of irrelevant features, were:

- Temperature;
- Dew point;
- Sensation of temperature;
- Minimum temperature;
- Maximum temperature;
- Pressure;
- Humidity;
- Wind speed;
- Wind direction;
- Wind gusts;
- Rain volume (1 hour);
- Snow volume (1 hour)
- Clouds cover

Adding to the features already given in the dataset, we have created one more feature associated to problems of sensor failure in WWTPs. This feature is related to periods of melting snow and was determined as in Equation 3. It returns the temperature when the temperature is above 0 degrees Celsius and when the snow depth is above a threshold of 0.22 inches (we have defined this threshold because the variable assumed values just below 0.22 inches for summer months when there is no snow.). The feature of snow depth is in the dataset we have used, however we have collected it from another weather dataset of the same location.

$$s_{melt}(t) = \begin{cases} temp(t) & \text{if } temp(t) > 0,\ snow_{depth} > 0.22 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

IV. Methodology

A. Long Short Term Memory Recurrent Neural Network

The architecture of an LSTM network consists of a series of LSTM units stacked together. Each unit contains four main components: input, forget, and output gates, and a cell state. The input gate controls the addition of new information to the cell state, the forget gate determines which information to discard from the cell state, and the output gate decides what information to output to the next layer. The cell state itself is a vector that carries information through the network, acting as a sort of "conveyor belt" that flows through the network

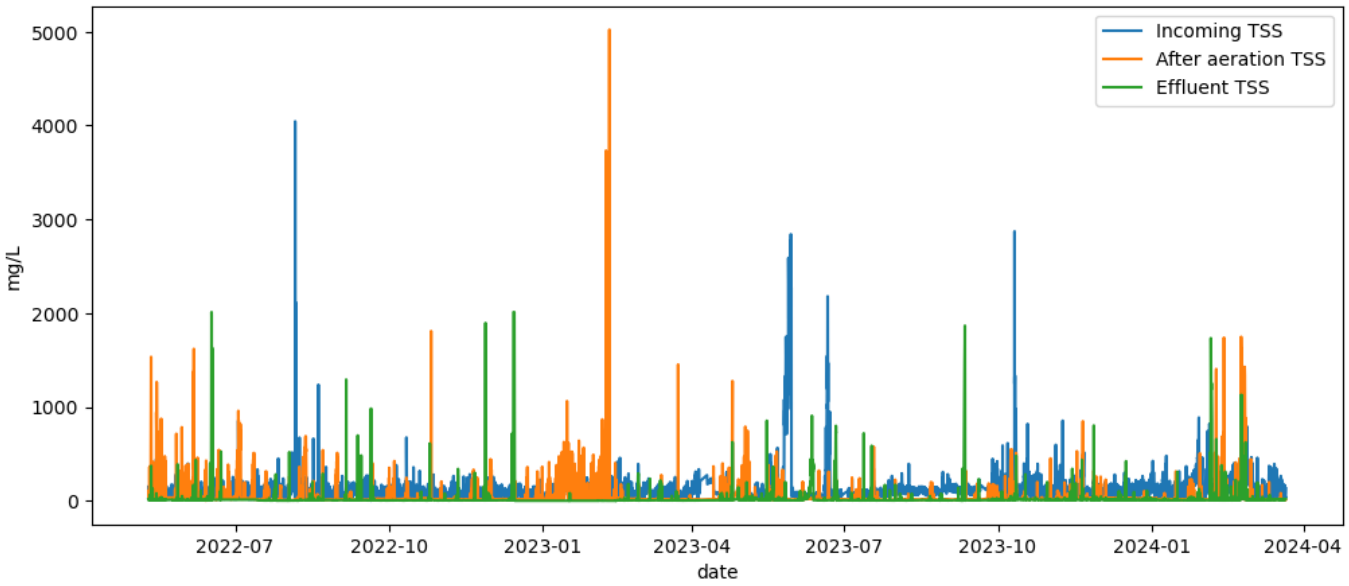

Fig. 1: TSS measurements of incoming water, water after aeration, and effluent water.

with minimal interference [6]. From one cell to another the other, there is still another parameter to be measured, which is the hidden state. The hidden state in an LSTM network is an encoded representation of the most recent time-step's data, which is neither the final output nor the prediction but can be processed to derive meaningful insights or predictions depending on the task at hand [7]. All the parameters of an LSTM cell are presented in image 2 and are determined as follows:

- Input gate:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{4}$$

- Forget gate:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{5}$$

- Cell state:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

- Candidate cell state:

$$\tilde{c}_t = tanh(W_c x_t + U_c ht - 1 + b_c) \tag{7}$$

- Output gate:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{8}$$

- Hidden state:

$$h_t = o_t * tanh(C_t) \tag{9}$$

In the equations above, x_t stands for the input values, W and U represent the weights (trainable parameters) of each one of the gates, and b stands for the biases of each gate, which are also trainable parameters [8]. Figure 3 shows how the units of cells work together in sequence.

In our models, we have tested several hyperparameters to obtain the optimum TSS forecasts. By applying grid search, we

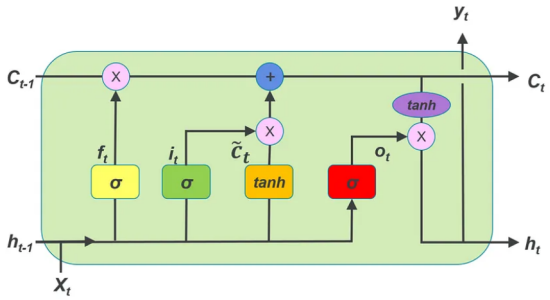

Fig. 2: LSTM cell structure

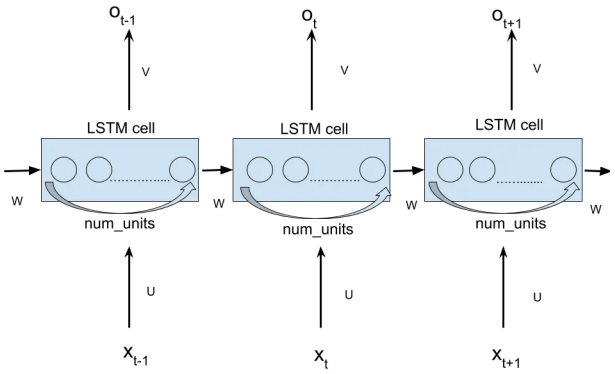

Fig. 3: LSTM basic units sequence

have considered the number of lags in hours (the memory of the model), the number of hidden layers, the number of units per hidden layer, and rate of layer dropout. The parameters considered are presented in Table I. Therefore, knowing that we perform grid search to forecast three time series, the total

TABLE I: LSTM models parameters from grid search.

Parameters	Grid search values
Number of lags	12, 24
Number of layers	1, 2, 3
Number of units	50, 100
Layer dropout rate	0, 0.3

number models trained is 72.

B. Feature selection

To avoid overfitting and remove unnecessary information from the models, we have used random forest regression to score the features compared to the three target variables [9]. First, we have normalized the weather features and the three target variables, and then fitted three models with 200 random estimators each to obtain the features' importance for each one of the measurements of TSS. We have set a threshold of 0.03 to select the features. As we can see from Figures 4 - 6, the random forest regressors classify all features as relatively important for prediction, except for *wind gusts*, *rain (1 hour)*, *snow (1 hours)*, and *snow melt*. From these four features, *rain (1 hour)* and *snow melt* were considered important for TSS of incoming water, *wind gusts* and *snow melt* for TSS of water after aeration, and *snow (1 hour)* for TSS of effluent water. Therefore, for each model have 12 features, except for the model to forecast the effluent TSS which contained 11.

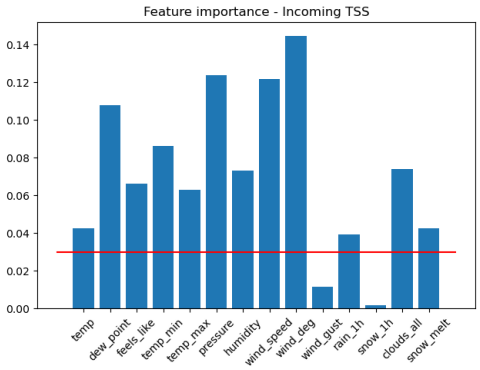

Fig. 4: Feature importance for prediction of TSS in incoming water.

Besides the features selected, we have selected a second set of features related to sensor failures in WWTPs according to literature and human expertise [1]. The manual selection was based on estimating which weather parameters would influence the stress of the treatment system most. With this second set of features, we have retrained one model for each one of the of the TSS time series according to the best parameters given by the training grid search performed with the features selected by feature importance. The second set of features for each time series forecast was:

- Temperature;
- Humidity;
- Rain (1h);

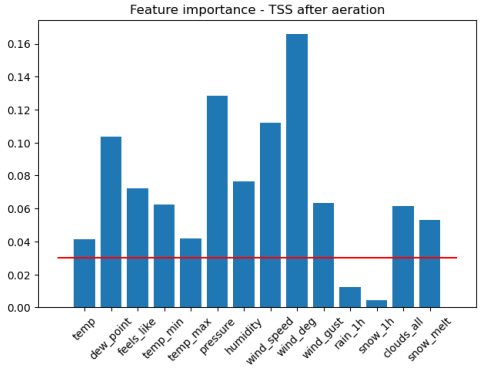

Fig. 5: Feature importance for prediction of TSS in water after aeration.

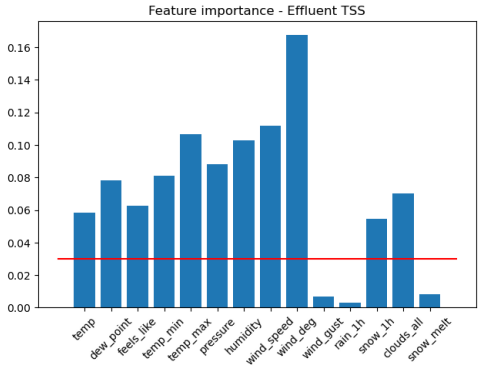

Fig. 6: Feature importance for prediction of TSS in effluent water.

- Snow (1h);
- Snow melt.

C. Processing forecast output for model evaluation

It is unlikely that the LSTM models forecast perfectly sharp and non-stationary time series such as the ones we are modeling. However, due to the nature of this project, as we are mostly interested in the volatility of the time series that represents a shift of the correct measurements, we propose applying differentiation to the forecasted time series in order to emphasize the abrupt variation of the forecast, as validated by Kozionov et al. [10]. Therefore, we are testing if the models are able to predict the changes of the TSS measurements that may represent failures of the sensor. The expression to obtain any differentiated time series is presented in Equation 10.

$$y_{diff}(t) = y(t) - y(t-1) \tag{10}$$

D. Failure prediction system

For each of the three sensors, we developed several models to forecast the failure of the system. Once the forecasts were generated and differentiated as mentioned before, we applied a threshold to transform the continuous output of each model into binary values. Specifically, if a forecasted value exceeds

a predefined threshold for the three target variables, it will be classified as a '1', signifying a predicted failure. Conversely, if the forecasted value remains below this threshold, it will be marked as a '0', indicating no failure. This transformation process allows us to evaluate the performance of our models by comparing the arrays of binary outcomes against actual occurrences of system failures. The way we have defined the thresholds of failure for each series is explained in the following section.

E. Threshold for failure

The models for failure prediction are evaluated by comparing binary samples (failure/no failure), therefore it is important to define the threshold for failure for each sensor, as did Han et al. [4]. As seen in Figure 1, the after aeration TSS and effluent TSS time series are relatively flat, however present major spikes occur when the sensor misreads the quantity of TSS in the water. Therefore, we have defined the failure threshold for the aeration TSS sensor as $2 \times mean(TSS_{aeration})$ and $3 \times mean(TSS_{effluent})$ for the effluent TSS sensor.

The measurements of TSS in the incoming water reveal a higher fluctuation, especially with a period of 24 hours and 12 hours (seen after analyzing the fast Fourier transform of the series). Hence, we have applied Infinite Impulse Response (IIR) Notch Filter to remove the influence of the sinusoidal functions of frequency $\frac{Fs}{24}$ and $\frac{Fs}{12}$, where Fs represents sampling frequency of $\frac{1}{3600}Hz$ [11]. After filtering the time series, we have also extracted the polynomial trend (degree = 20) to obtain an approximation of the normal values of TSS over time. In the end, trough graphical analysis, we have determined that the equation of failure threshold for the time series of TSS of incoming water is as in Equation 11. The visual representation of the failure threshold for this measurement is shown in Figure 7.

$$threshold(t) = trend(t) + 100 \tag{11}$$

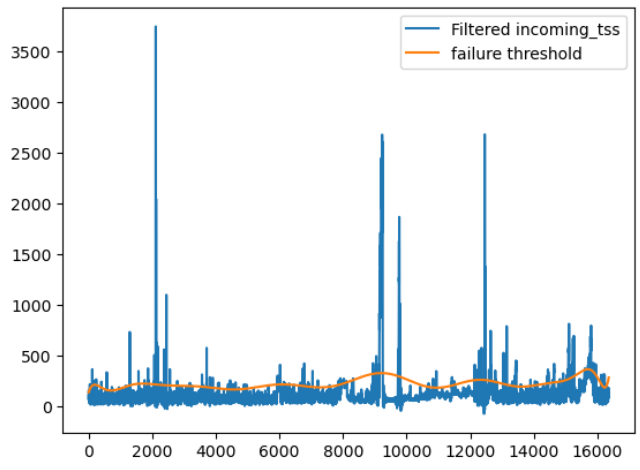

Fig. 7: Failure threshold for TSS measurements of incoming water.

TABLE II: Parameters of the best model for each time series forecast

Forecasted parameter	MAE	Number of lags (h)	Number of layers	Number of units	Dropout rate
Incoming TSS	0.0193	24	2	50	0.0
Aeration TSS	0.0071	12	3	50	0.0
Effluent TSS	0.0093	24	1	100	0.3

F. Performance metrics

The performance metrics we have used to assess the capability of the models to classify instances of sensor failure were accuracy, recall, and F1-score. These metrics are displayed in Equations 12 - 13. In the context of this project, the true positives (TP) represent the number of correct failure prediction, true negatives are the number of correct non-failure predictions (TN), false positives (FP) represent the number of wrong failure predictions, and false negatives (FN) are the number of wrong non-failure predictions [12].

$$Accuracy = \frac{TP + TN}{FP + FN} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{15}$$

V. Results and Discussion

In this section we present the result of the implemented methodology and the discussion of these results.

A. Grid search results

After training the 72 models above mentioned and before differentiating the forecasted series to classify failure instances, for each one of the three TSS time series we have selected the forecast the showed the lowest value of Mean Absolute Error (MAE). The values presented in Table II show that MAE is very low. However, due to the trend of each time series being very flat and the rarity of failure occurrences, the low values of MAE do not mean that forecast matches, predicting correctly all the spikes shown before.

B. Loss function analysis

Upon reviewing the loss functions of the three LSTM models, it became evident that despite our efforts to mitigate overfitting, such as through feature selection and the application of dropout layers, a common issue emerged. Specifically, the training loss consistently fell below the validation loss across all models [13], especially for the model to forecast TSS of incoming water and TSS of effluent water, as seen in annexed Figures 13 - 15. This phenomenon is indicative of overfitting, where the model performs exceptionally well on the training data but struggles to generalize its learning to unseen validation data.

The inability to align the validation loss curve with the training loss curve highlights the necessity for alternative approaches in handling forecast data, especially when the goal is to predict sensor failures based on historical data trends. Also, the presence of large spikes in the TSS time series complicates the task of developing accurate predictive models, since the model tends to adjust to the abrupt variances, which increases the change of overfitting. Therefore, we have applied differentiation after the LSTM stage of the model.

C. Differentiation and failure prediction performance

In the pursuit of forecasting TSS over time for the three sensors using a LSTM models, we aimed to capture the nuanced variations in these values, including the significant spikes that are critical for monitoring and managing water quality. However, upon evaluating the initial predictions, it became apparent that while the model successfully identified some nuances, it struggled to accurately represent the spikes in TSS levels. To address this shortfall and enhance the model's ability to capture the desired peaks, we employed differentiation techniques on the forecasted values. This approach was applied uniformly across all datasets, as evidenced by Figures 8 - 10, which represent the real TSS time series, the differentiated forecast using the features selected by feature importance, and the threshold to consider a TSS reading as sensors failure.

Figure 8, in particular, illustrates the TSS levels of incoming water. Although the differentiated model managed to catch some nuances of the spikes, it also introduced a noticeable shift in the timing of these peaks. This observation was mirrored in the representations of TSS after aeration and in effluent water, suggesting that while the differentiation technique improved the model's ability to detect fluctuations, it did not entirely resolve the issue of accurately capturing spike timings.

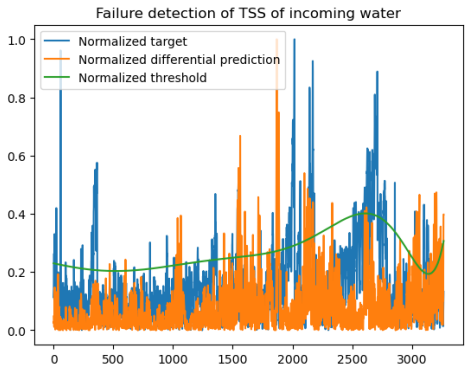

Fig. 8: Differentiated forecast compared to TSS of incoming water.

Despite these limitations, it is clear that the differentiated model did succeed in identifying certain fluctuations across the three time series, particularly in the post-aeration phase and in the affluent water. This indicates that the model has the capability to catch more spikes than initially anticipated,

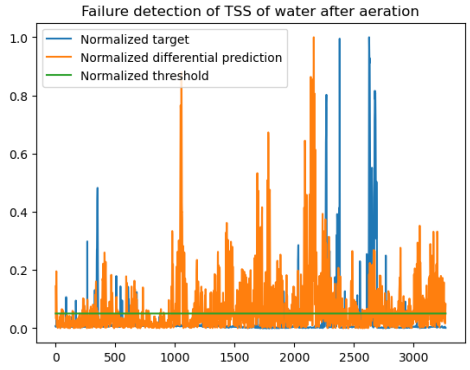

Fig. 9: Differentiated forecast compared to TSS of water after aeration.

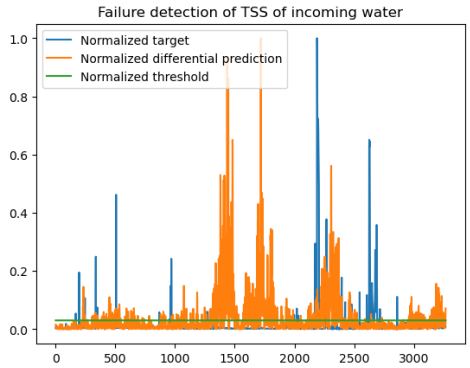

Fig. 10: Differentiated forecast compared to TSS of effluent water.

although the exact representation of these spikes remains a challenge, especially because we seen that it forecasts many more spikes than it should. Besides, it is important to note that we applied a Butterworth low-pass filter to the predictions of TSS after aeration and effluent TSS to mitigate the effects of high-frequency noise [14]. The cut frequency was $Fc = 0.00005$, which was chosen arbitrarily be visual inspection. This filtering process contributed to the reduce the number of spikes visible, however, it did not boost significantly the performance.

As mentioned before, we have selected a second set of features for each one of the three TSS forecasting models, which included only variables that tend to affect the normal functioning of WWTPs more directly. We have used this dataset with the best LSTM parameters presented in Table I, however, we did not better results, as the differentiated forecast did not present evident changes from the ones presented in Figures 8 - 10. The forecasts graphs of the models trained with this second set of features are presented in Figures 16 - 18.

The final stage of the model consisted in transforming the forecasts into binary classification based on the position of the TSS values compared to the failure thresholds. As mentioned before, the every time TSS would be above the failure

TABLE III: Performance metrics for failure prediction models.

Models	Accuracy	Precision	Recall	F1-score
Features selected by feature importance (M1)				
Incoming TSS	0.8384	0.1989	0.0889	0.1229
Aeration TSS	0.5289	0.0748	0.5571	0.1319
Effluent TSS	0.6443	0.0390	0.2905	0.0689
Features selected by manual selection (M2)				
Incoming TSS	0.8231	0.1786	0.1081	0.1347
Aeration TSS	0.5366	0.0798	0.5905	0.1407
Effluent TSS	0.4790	0.0498	0.5810	0.0918

threshold the instance would be classified as 1, and would be classified as 0 otherwise. After converting the results to binary classification, we have determined the the models capability of predicting failure of the sensors based on accuracy, recall, precision, and F1-score. The results for both the models with the features selected (M1) and by feature importance scoring and by manual selection (M2) are presented in Table III.

As we can see in Table III, the only metric that seems to return an optimal value is accuracy for both Incoming TSS models, where it scores 83.84% for M1 and 82.31% for M2. For other models of type M1 and M2 accuracy was lower, but considerably higher than other metrics. However, since the number of failure instances is much lower than not, the classification system is highly unbalanced, which makes accuracy not valid for this evaluation. The second feature to consider is recall, which measures the rate of positives (failures) predicted correctly out of all real positives (true positives and false negatives). For this metric we can see that three models can detect slightly more than half of all failure instances, and these models are M1 for TSS after aeration, M2 for TSS after aeration and effluent TSS, where the recall values are 55.71%, 59.05%, and 58.10%, respectively. However, when we take precision into consideration, it is visible how the model struggles not classify most instances as a failure. As precision indicates the number of true positives (real failure instances) out of all instances where the model predicted as positive positive (sum of true positives and false positives), we see that all models struggled not to classify the majority of instances as failure, as we can especially confirm from the forecasts graphs for TSS after aeration (Figure 9) and incoming TSS (Figure 8).

As mentioned before, Han et al. [4] developed a model where they have used TTD and DSF as evaluation metrics in comparison with the failure threshold. However, in the context of your study we could not use these metrics, since the nature of the faults in TSS sensors is more abrupt and the frequent variations do not allow to define t_{f1} and t_{f2}. As for the results for others studies, we could not compare the performance we have obtained because we have not found any other paper that used a similar approach as ours.

D. Assertion of failure prediction performance

Due to the nature of the time series, it is very likely that both the forecasted values and the differentiated forecasted series present peaks that do not coincide with the peaks of the of the target time series. We have observed that this was due to the poor performance of the model, however, there were some peaks that were present in both the differentiated time series and the target series that corresponded to the same failure but with a small time shift. Therefore, in these instances it did not mean that the prediction failed to detect the failure, but failed to detect the exact instant when the failure would occur, as it is illustrated in Figure 11. The figure presents two failure periods of the TSS sensor after aeration that were predicted with correctly with a time shift, which in turn classified several instances as false positives.

That being said, we propose a solution to evaluate these instances where false positive instances should be counted has true positive instances. After transforming the differentiated forecasted series into an array of 0s and 1s, we have verified whether at every false positive instance there would be a true failure instance 3 time stamps (3 hours) before or 3 time stamps after the false failure signaling. While checking the failure classification one by one over time, every time that occurred a false positive occurred, we subtracted 1 to the number of false positives and added 1 to the number of true positives considering the existence of true failures in the time gap mentioned before.

Beside reevaluating false positives as actual failures, we have also detected other instances where the models classified the instance as 0 (non-failure) when the target array indicated 1 (failure). In this case, we observed that the peak in the target series was isolated above the threshold and all the neighboring values were below the failure threshold, as shown in Figure 12, indicating that it is referring to a period of time when the sensor practically did not fail. In such instances, the differentiated forecast was below the threshold, indicating a false negative, however, we have analyzed both the target and prediction binary arrays to change such instances from false negatives to true positives.

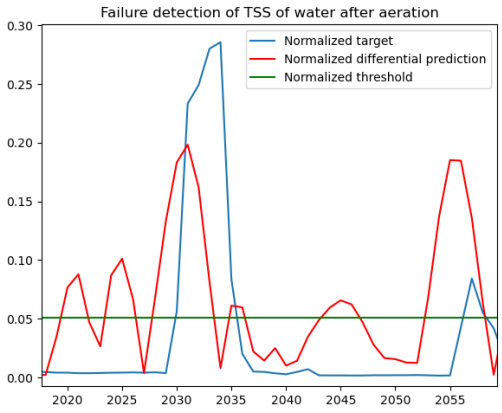

Fig. 11: Two instances when the failure of the sensor was predicted with a time shift.

After that, we have recalculated all the performance metrics, as presented in Table IV. As we can see, all metrics improved, especially recall in the M2 models of TSS after aeration and

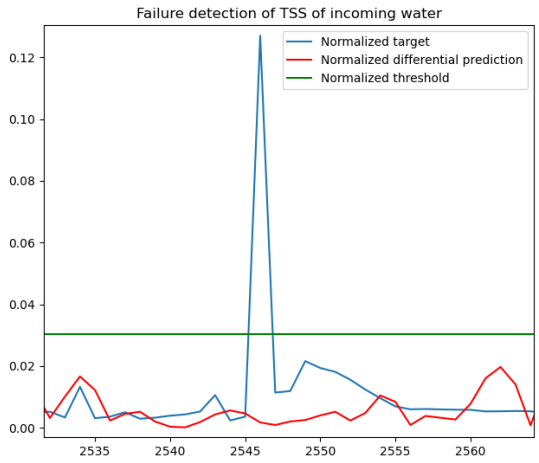

Fig. 12: Instances when a non-failure of the sensor was predicted while only an isolated peak of the target series was above the failure threshold.

TABLE IV: Adjusted performance metrics for failure prediction models.

Models	Accuracy	Precision	Recall	F1-score
Features selected by feature importance (M1)				
Incoming TSS	0.8470	0.3495	0.1464	0.2063
Aeration TSS	0.5623	0.1440	0.7075	0.2394
Effluent TSS	0.6596	0.0845	0.4952	0.1575
Features selected by manual selection (M2)				
Incoming TSS	0.8427	0.3254	0.1925	0.2419
Aeration TSS	0.5706	0.1476	0.7387	0.2461
Effluent TSS	0.5185	0.1176	0.8024	0.2052

effluent TSS. This reiterates the fact that the model forecasted moments of failure with high performance. However, precision and recall present values that are far from optimal, indicating that all the models have the tendency to forecast too many instances as failure where they are actually not. Therefore, even though we have implemented some measures to adjust the performance metrics, we did not get a desirable model to forecast instances of sensor failure.

In the end, even though the differences were not very significant, the models that scored a better balance between accuracy and F1-score to predict sensor failure when measuring TSS were the M2 models. In these models, the values of F1-score were higher due to the higher recall, showing that the features that were manually selected according to their known impact on WWTP sensor measurements can translate better results in this type of forecasting. This corroborates what is stated by Wang et al. [1].

VI. Conclusions

The performance of TSS sensors is affected by a multitude of variables, including the characteristics of the wastewater, the operational conditions of the treatment process, and environmental factors. This introduces significant nonlinearity and complexity into the data, making it challenging for LSTM models to accurately predict failures. The research also highlights the importance of feature engineering and professional knowledge when finding use cases outside the data science industry. Features selected a

In this study, we have developed a LSTM model to forecast the values of TSS measured by WWTP sensors of incoming water, water after aeration, and effluent water, and consequently predict when the sensors would fail reading these values. The performance of the models was not optimal, even though accuracy and recall scored high values.

VII. Future work

As future work, we propose adpating the variables to build an LSTM-based variational autoencoder to forecast sensor failure as Han et al.'s study [4]. We also propose more research to be done to find correlation between stress of the treatment system, influent quality and sensor failure to improve on manual feature selection .

References

[1] D. Wang, Y. Chen, M. Jarin, and X. Xie, "Increasingly frequent extreme weather events urge the development of point-of-use water treatment systems," *npj Clean Water*, vol. 5, p. 36, 8 2022.
[2] F. Hernández-del Olmo, E. Gaudioso, N. Duro, and R. Dormido, "Machine learning weather soft-sensor for advanced control of wastewater treatment plants," *Sensors*, 2019.
[3] "Energy consumption prediction in water treatment plants using deep learning with data augmentation," *Results in Engineering*, vol. 20, p. 101428, 2023.
[4] P. Han, A. L. Ellefsen, G. Li, F. T. Holmeset, and H. Zhang, "Fault detection with lstm-based variational autoencoder for maritime components," *IEEE Sensors Journal*, vol. 21, pp. 21903–21912, 10 2021.
[5] A. Dhinakaran, "Understanding kl divergence," *Towards Data Science*, 2 2023.
[6] R. T.J.J., "Lstms explained: A complete, technically accurate, conceptual guide with keras," *Analytics Vidhya*, 9 2020.
[7] D. Thakur, "Lstm and its equations," *Medium*, 7 2018.
[8] A. Saxena, "Introduction to long short-term memory (lstm)," *Analytics Vidhya*, 1 2023.
[9] J. Brownlee, "Feature selection for time series forecasting with python," *Machine Learning Mastery*, 9 2020.
[10] A. Kozionov, M. Kalinkin, A. Natekin, and A. Loginov, "Wavelet-based sensor validation: Differentiating abrupt sensor faults from system dynamics," pp. 1–5, IEEE, 9 2011.
[11] Z. Zhang, Z. Li, and Z. Li, "An improved real-time r-wave detection efficient algorithm in exercise ecg signal analysis," *Journal of Healthcare Engineering*, vol. 2020, pp. 1–7, 07 2020.
[12] K. Nighania, "Various ways to evaluate a machine learning model's performance," *Towards Data Science*, 12 2018.
[13] V. Saikiran, "How to diagnose common machine learning problems using learning curves," *SoapBoxLabs*, 7 2022.
[14] I. S. Sayed and N. S. M. Nasrudin, "Effect of cut-off frequency of butterworth filter on detectability and contrast of hot and cold regions in tc-99m spect," *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology*, vol. 05, pp. 100–109, 2016.

APPENDIX

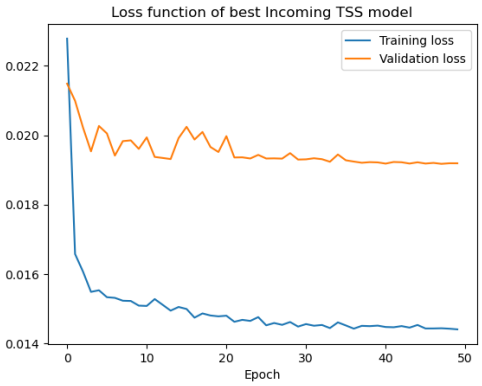

Fig. 13: Loss function of the best LSTM model for incoming TSS.

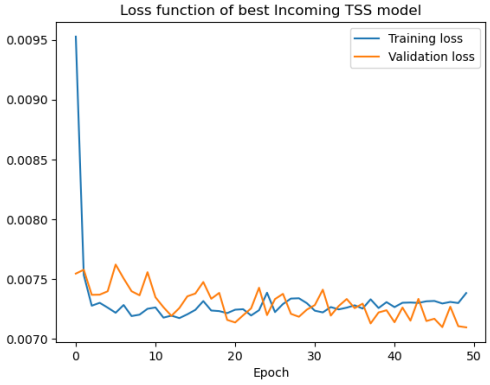

Fig. 14: Loss function of the best LSTM model for TSS after aeration.

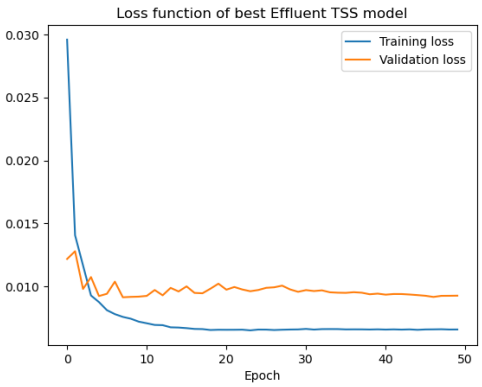

Fig. 15: Loss function of the best LSTM model for for TSS.

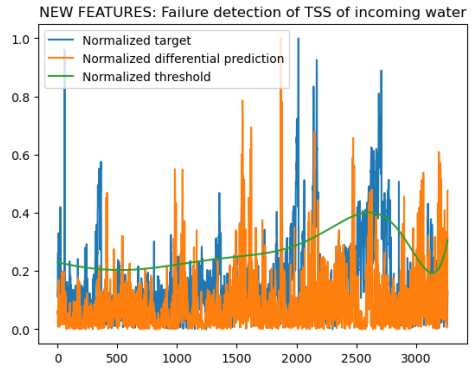

Fig. 16: Differentiated forecast compared to TSS of incoming water with manually selected features.

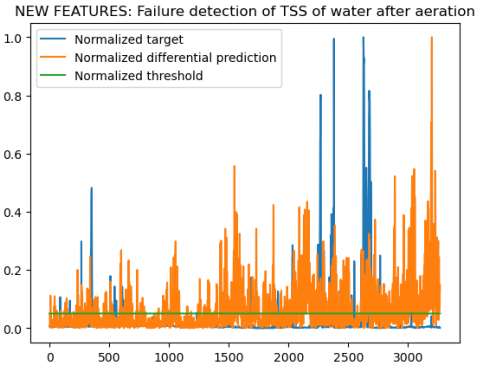

Fig. 17: Differentiated forecast compared to TSS of water after aeration with manually selected features.

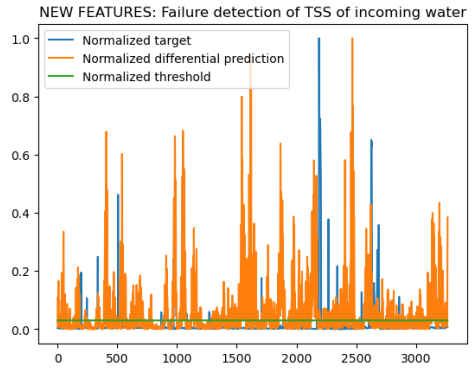

Fig. 18: Differentiated forecast compared to TSS of effluent water with manually selected features.