T3.2 Digital Twin of the Detection System

Short description of the Early Detection algorithm

The project FELINES aims at designing a protection system capable of sensing electromagnetic fields that are preliminary to a lightning event, and consequently disconnect part (or all) of the electric infrastructure under its protection. These fields are generated by the so-called Preliminary Breakdown (PB) pulses, localized events taking place during the first phases of the lightning inception.

In the project the experimental part was not planned, for this reason the PB pulses are simulated according to [1]. The PB pulse is the "trigger" that starts the algorithm for the Early Detection, which is easily described in the following points:

1. PB pulse start.
2. PB pulse detection by the effect it induces on a Transmission Line (in terms of overvoltages).
3. Evaluation of the consequent Return Stroke (RS) danger level.

Points 2 and 3 are carried out by using Machine Learning (ML) based procedures. More in detail:

A. A Neural Network (NN) is properly designed and trained in order to predict the overvoltage caused by PB pulses. This steps serves as PB pulse detection (point 2) and is the input to point 3.
B. A different NN is properly designed and trained in order to predict the overvoltage peak caused by the consequent RS. The knowledge of these overvoltages and the characteristics of the TL lead to a classification of the event as dangerous or not dangerous.

This deliverable describes the set of results obtained for the specific test case described in the deliverable related to T3.1, and depicted in Fig. 1.

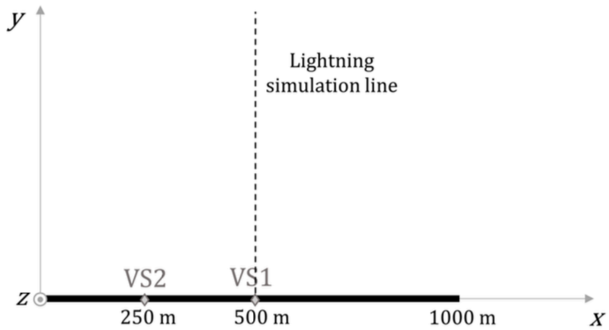

Fig. 1 Line geometry

Numerical results #1

The dataset includes signals with 100 features (nPCA = 100), preserving $\Gamma = 100\%$ of the total variance; with 75% allocated for training and 25% for testing. Each model is optimized using mini-batch gradient descent (batch size = 32) for 275 epochs, providing adequate convergence.

Figure 2 evaluates the model for the regression task of the prediction of RS peak voltage V_{RS} over a wide range of RS peak voltages up to 500 kV.

The scatter plot on the left shows a strong linear correlation between the predicted and true values, with a coefficient of determination $R^2 = 0.90$, indicating that the model captures the underlying relationship with high accuracy. The histogram and boxplot on the right further highlight the distribution of absolute errors, where the majority of predictions have absolute erors lower than 25 kV, and the MAE equals 9.27 kV, which is significantly low compared to the wide range of values.

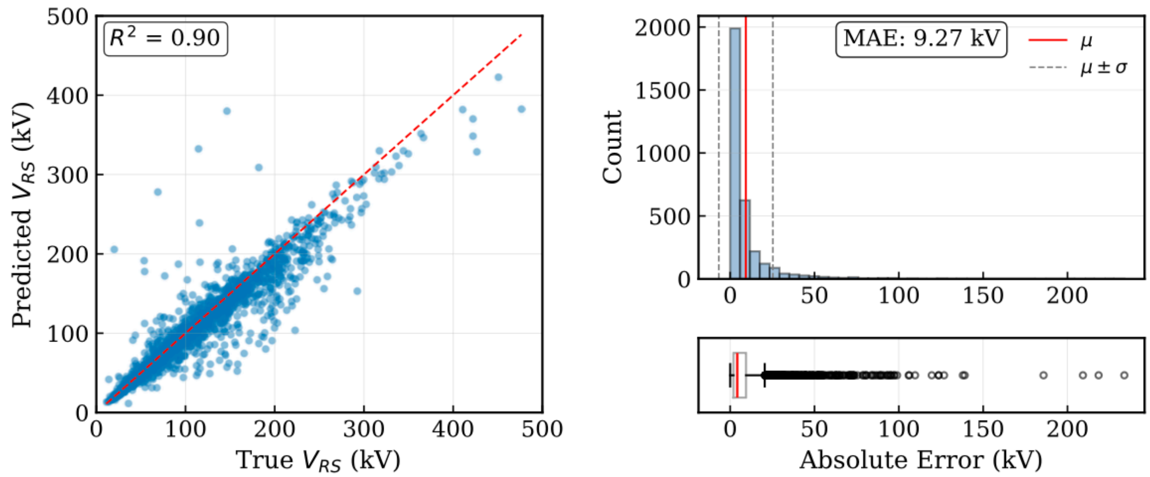

Figure 2: Scatter plots and error analysis for the predicted maximum voltage VRS in Scenario B. (Left) The predicted versus true values. (Right) The histogram and boxplot of absolute errors.

After the DL model completes the regression task of predicting the RS pulse maximum voltage, to classify pulses as safe or dangerous, a threshold voltage of $V_{th} = 75\,kV$ is applied. Those exceeding the threshold are labeled as dangerous, while those below it are considered safe.

The voltage value exceeding the threshold are labeled as harmful, while those below it are considered safe. To interpret the classification results, the confusion, recall, and precision metrics are defined. The confusion matrix shows how many samples are classified for each class. The recall matrix is obtained by normalizing each row of the confusion matrix to quantify the model's ability to detect events in a given class. Conversely, the precision matrix is obtained by normalizing each column and reflecting the reliability of each predicted label.

As shown in Figure 3, the classification performance of the proposed approach based on count-based, recall-normalized, and precision-normalized confusion matrices is evaluated.

In panel (a), it is evident that for the majority of samples, the predicted labels and actual labels are the same, with only a few misclassifications between the 'S' and 'D' classes.

The recall matrix in panel (b) demonstrates that the model successfully identifies 96.6% of class 'S' and 92.8% of class 'D' samples, which means it can detect the majority of the dangerous pulses. Also, the precision matrix in panel (c) confirms that 91.9% of the samples predicted as 'S' and 97.0% of those predicted as 'D' are correct, reflecting the model's high reliability, especially in the prediction of dangerous cases.

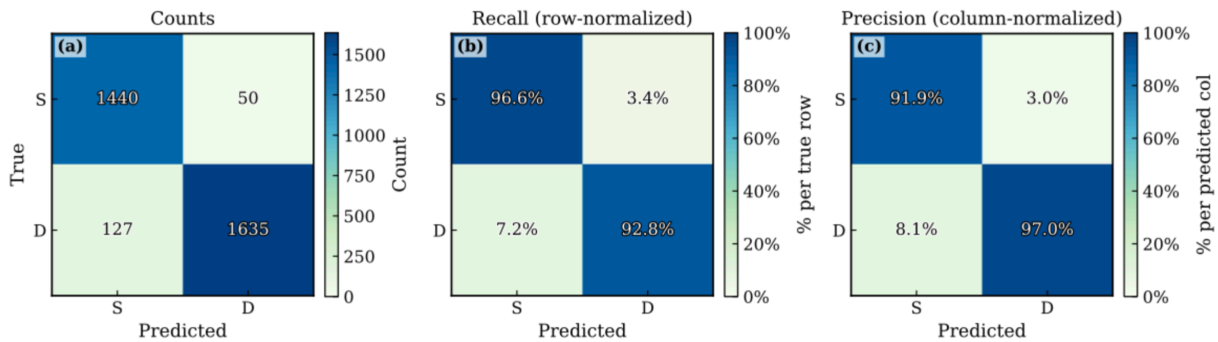

Figure 3: (a) Count-based confusion matrix. (b) Recall matrix (percentage per true class). (c) Precision matrix (percentage per predicted class).

Numerical results #2

In order to show the efficiency of the proposed algorithm, the results relative to a different transmission line (with the same geometry) are shown here. The new test case, again represented by Fig. 1, is characterized by single phase conductor overhead distribution line with a 50-kV Critical FlashOver voltage (CFO) is considered, and the threshold to assess if a indirect lightning strike results into a dangerous flashover is 1.5 CFO, as recommended in [2].

As for the line geometry: length 1 km, conductor height 10 m, conductor radius 1 cm. Matching impedance are set at both the line ends to minimize reflections. A PEC soil is assumed. The line insulators are modeled by means of their parasitic capacitance of 1 pF.

The performance of the regression model in this new scenario is shown in Figure 4. A strong agreement between the predicted and true values with a high coefficient of determination $(R^2 = 0.92)$ is observed in the scatter plot.

The histogram and boxplot of the absolute errors further corroborate this performance, showing a low mean absolute error (MAE) of approximately 7.40 kV. Although a small number of outliers with larger errors are present, the overall error distribution remains tightly concentrated at low values, confirming the robustness of the model in test case #2.

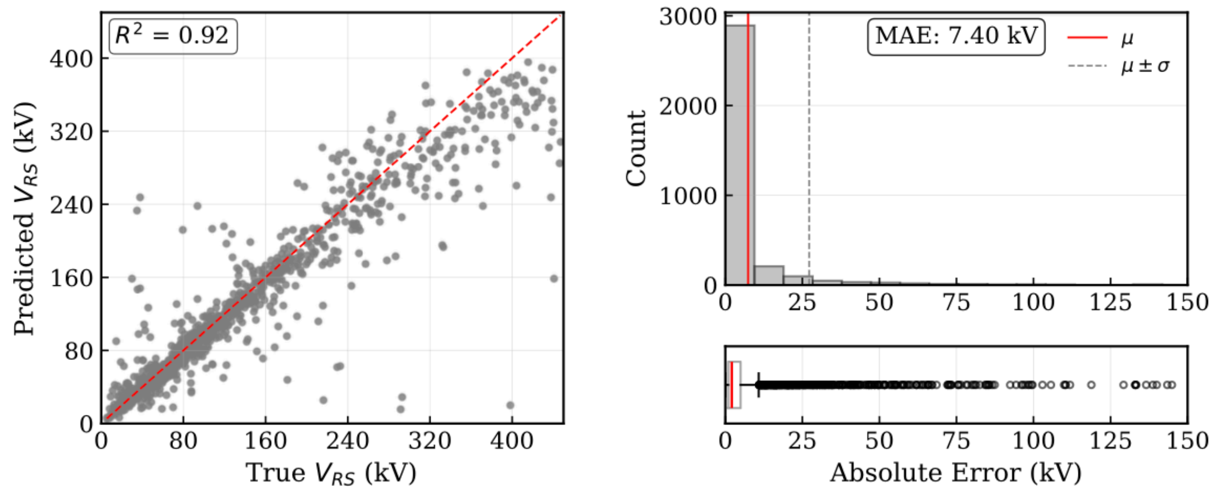

Figure 4: Scatter plots and error analysis for the predicted maximum voltage VRS in test case #2 (Left). The predicted versus true values (Right). The histogram and boxplot of absolute errors.

The results of the classification are summarized in Fig. 5. As shown in panel (b), the model achieves a very high recall for the Safe ('S') class (99.0%), indicating that nearly all safe events are correctly identified. The recall for the Dangerous ('D') class is also high (91.3%), demonstrating a strong capability to detect hazardous events. Panel (c) further shows that the precision is well balanced between the two classes, with values of 97.0

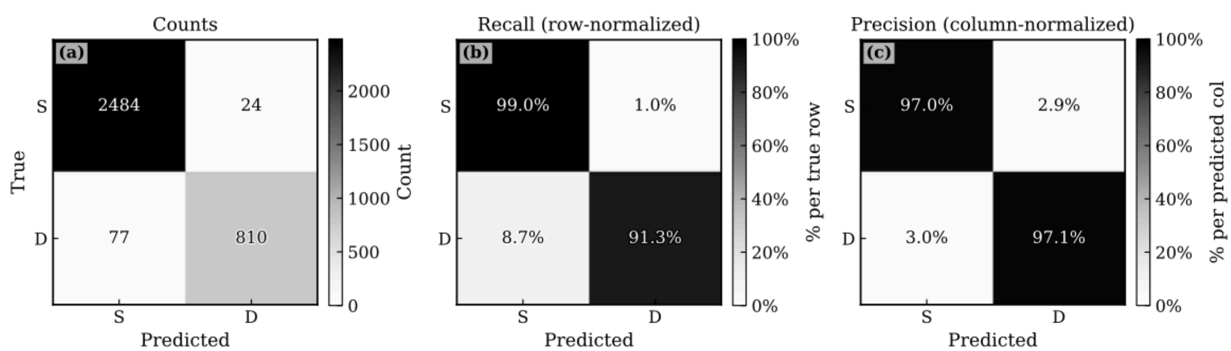

Figure 5: (a) Count-based confusion matrix. (b) Recall matrix (percentage per true class). (c) Precision matrix (percentage per predicted class) in test case #2.

The test case #2 is relative to a transmission line model that is simplified with respect to test case #1, so a less realistic scenario. However, the Early Detection algorithm performs extremely well in both cases with extremely high perfomances in temrs of recall and precision.

Final considerations

The Machine Learning based “pipeline” consisting in the two Neural Networks recalled abroad and described in the deliverable relative to T3.1 is, as a matter of fact, a Digital Twin of the detection system: the model is capable of modelling the voltage induced by the PB pulses (i.e. the role of the sensors) to estimate the overvoltage produced by the corresponding RS (i.e. the model of the transmission line) and decide whether the lighgtning event is dangerous or not.

The NNs are implemented on a Matlab – Python framework, hence very close to a real online implementation.

References

[1] D. Mestriner, M. Nicora, M. Brignone, R. Procopio, Y. Zhu and V. A. Rakov, "Modeling and Statistical Characterization of Preliminary Breakdown Pulses in Negative Cloud-to-Ground Lightning Flashes," in *IEEE Transactions on Electromagnetic Compatibility*, doi: 10.1109/TEMC.2026.3666274.

[2] IEEE, Guide for improving the lightning performance of electric power overhead distribution lines, IEEE Std. 1410-2010 (Revision of IEEE Std 1410-2004) (2011) 1–73doi:10.1109/IEEESTD.2011.5706451.