Received: 8 March 2021 | Revised: 14 July 2021 | Accepted: 15 August 2021

IET Smart Grid

DOI: 10.1049/stg2.12051

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Siamese recurrent neural networks for the robust classification of grid disturbances in transmission power systems considering unknown events

André Kummerow | **Cristian Monsalve** | **Peter Bretschneider**

Fraunhofer IOSB, IOSB-AST Ilmenau, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Fraunhofer Center for Machine Learning, Ilmenau, Germany

Correspondence

André Kummerow, Fraunhofer IOSB, IOSB-AST Ilmenau, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Fraunhofer Center for Machine Learning, Ilmenau, Germany.
Email: andre.kummerow@iosb-ast.fraunhofer.de

Funding information

Open Access funding enabled and organized by Projekt DEAL.

Abstract

The automated identification and localisation of grid disturbances is a major research area and key technology for the monitoring and control of future power systems. Current recognition systems rely on sufficient training data and are very error-prone to disturbance events, which are unseen during training. This study introduces a robust Siamese recurrent neural network using attention-based embedding functions to simultaneously identify and locate disturbances from synchrophasor data. Additionally, a novel double-sigmoid classifier is introduced for reliable differentiation between known and unknown disturbance types and locations. Different models are evaluated within an open-set classification problem for a generic power transmission system considering different unknown disturbance events. A detailed analysis of the results is provided and classification results are compared with a state-of-the-art open-set classifier.

KEYWORDS

pattern classification, phasor measurement

1 | INTRODUCTION

1.1 | Disturbance classification and current limitations

The automated processing and analysis of phasor measurement units (PMU) in transmission and distribution power systems has been a main research field for several years [1–4]. PMUs provide time-synchronised frequency, voltage and current phasor measurements with a high time resolution and enable new dynamic control centre functions for the monitoring and control of future power systems. One main issue is the automated classification of grid disturbances (e.g. line trips, load losses and short-circuits) by the use of data-driven machine learning techniques. Precise information about the origin and type of the disturbance enable the automatic activation of countermeasures (e.g. remedial action schemes) to maintain a stable system operation and to prevent severe failures of the power supply. With that, PMU-based recognition systems can overcome drawbacks of existing SCADA systems (low-resolution, spontaneous and not synchronised measurements [5]) and protection systems (limited to short circuits, risk of misoperations [6]).

Several approaches are proposed in [7–17] using different techniques for feature extraction (e.g. wavelet decomposition, Stockwell transformation or time series shapelets) and pattern-based classification techniques (e.g. decision trees, support-vector-machines or artificial neural networks). Other studies investigate different neural network models including convolutional and spiking neural networks [18–20]. All of these approaches address one or more of the three main tasks of the disturbance classification including

- *detection* (recognise deviations from steady state),
- *identification* (recognise the disturbance type),
- *localisation* (recognise the origin of a disturbance).

Typically addressed events include short circuits, outages, switching operations as well as power quality-related

phenomena. The evaluations are based on dynamic grid simulations using reference topologies (e.g. IEEE 39-Bus system) or field measurements from the system operators. Dynamic grid simulations are mainly used to generate appropriate training data to recognise relevant disturbances without the need for extensive long-term measurements. In the application phase, the classification model is confronted with PMU field measurements. Due to the large number of degrees of freedom in today's power systems, only a subset of all possible events can be simulated with moderate numerical effort. This leads to a high risk of misclassification in the application phase due to *unknown events* that are not seen during training. These unknown events can include new disturbance types or locations as well as other events, which cause transient system responses (e.g. capacitor switching, load tap changing and switching operations). Additional performance degradations like measurement errors or unknown steady states are not addressed within this survey. Also, this study focusses on single disturbance events and neglects cascading events or the superposition of multiple events as investigated in [9].

1.2 | Learning approaches to handle unknown classes

To handle unknown events or classes in recognition systems, authors in [21–23] introduced the *open-set* or *open-world recognition* problem. Accordingly, traditional classification systems assume a known and limited number of classes, which is valid during the training and application phase. In this *closed-set* assumption, new instances appear only from the known classes. In the *open-set* assumption, not all classes are known at the training time and new instances of known and unknown classes may appear during the application phase. Appropriate classification systems require the additional capability to reject instances of unknown classes and to correctly assign instances of known classes. The general approach is to reduce the classification error for instances that are far away from the known class regions (*open-space*). Several methods have been presented to cope with this issue including 1-vs-Set machines [21], an OpenMax layer for deep neural networks [24], nearest non-outliers (NNO) [23] and One-Vs-Rest sigmoid classifiers [25].

1.3 | Main contributions and paper organisation

Despite being mentioned in previous studies [12], the problem of unknown events is not addressed by current disturbance classification approaches. For the first time, this study investigates the PMU-based disturbance classification task as an *open-set* recognition problem to handle disturbances, which are unknown during the model training. In contrast with existing approaches, the integration of rejection mechanisms allows to distinguish between known and unknown disturbance events and to minimise potential misclassifications during the application phase. This leads to better performance results on new test data and is an important requisite for achieving robust classification models [26]. For that, a Siamese recurrent neural network with a novel *double-sigmoid classifier* is presented and compared to existing *open-set* classification models to correctly assign known disturbances at the training time and to reject unknown disturbances at the test time. For better classification results, different attention-based embedding functions are proposed to extract representative features from the multi-dimensional phasor data. In particular, the study's contributions can be summarised as follows:

- provide a fully trainable PMU classification model to efficiently recognise known and unknown grid disturbance events;
- develop a novel Siamese recurrent neural network architecture with additional attention mechanisms for the simultaneous online identification and localisation of different grid disturbances;
- introduce a novel probabilistic binary classifier based on sigmoid functions to efficiently train and apply Siamese networks in open-set problems.

The study is organised as follows: Section 2 describes the grid topology and the general procedure to derive the PMU signals from dynamic simulations. Paragraph 3.1 introduces the simultaneous disturbance identification and localisation approach based on some preliminary work. Afterwards, the classification models and the evaluation metrics are described in the subsequent paragraphs 3.2, 3.3 and 3.4. Section 4 starts with a description of the different test scenarios, which are examined within this survey. It follows a detailed discussion of the training, validation and test results. Finally, section 5 gives a short summary of this study and describes possible future work. A list of acronyms used in this study is given in Table 1.

1.4 | Basic notations

$\boldsymbol{X}$ - matrix
$\underline{x}$ - vector
x - scalar
$\hat{x}$ - predicted/estimated value
$\bar{X}$ - set

2 | GRID TOPOLOGY AND DYNAMIC SIMULATIONS

The power system model for generating the dynamic simulation results is based on a simplified version of the ENTSO-E European Transmission System. The grid model consists of 33 substations at a 400 kV voltage level. To ensure the N-1 criteria, grid redundancy in terms of multiple interconnections between the nodes is implemented. Some generator units are equipped with automatic voltage regulators (AVR) and power system stabilisers (PSS) to control the system

TABLE 1 List of acronyms

BLPP	Base load power plant outage
DOCSig	DOC-sigmoid classifier
DoubleSig	Double-sigmoid classifier
GRU	Gated recurrent unit
LL	Load loss
LSTM	Long short-term memory
PMU	Phasor measurement unit
PV	PV plant power loss
SC	Short circuit
SCADA	Supervisory control and data acquisition
SNN	Siamese neural network
SingleSig	Single-sigmoid classifier
SPP	Steam power plant outage
STD	Standard deviation

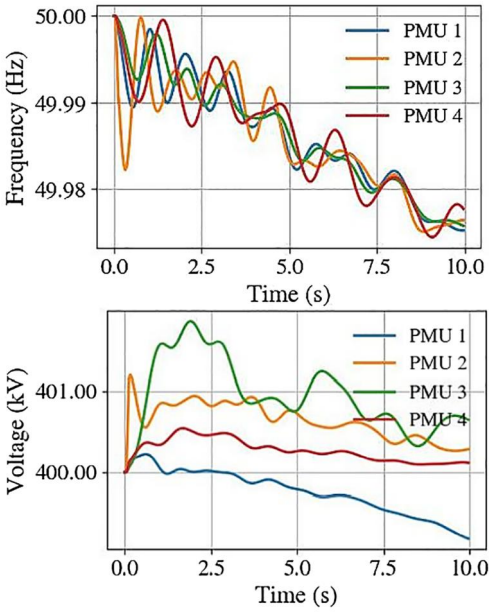

FIGURE 1 Exemplary phasor measurement unit signal patterns for a generator trip

stability. The parameters of these units have been optimised in order to ensure the system stability for the simulated contingencies. A DIgSILENT® Programming Language (DPL) script controls the event's creation, the stability assessment as well as the data export to the simulation database. Based on a selection of suitable operational points at the steady-state, multiple transient symmetric simulations are performed with a step size of 1 ms for different contingencies (e.g. generator or line outages). The simulations comprise 440 contingencies for seven different operational points and serve as the training and test database for the classification models. The PMU signals are extracted from the simulation results as moving averages over a fixed time interval. Based on prior investigations and literature research, frequency and voltage magnitude signals are used as relevant predictors with a reporting rate of 25 f.p.s. for the classification tasks. Some exemplary simulation results for four different PMU sensors are given in Figure 1.

In the case of no severe stability violations, the simulations are stopped about 140 s after the event triggering. In a post-processing procedure, some contingencies are discarded from the database including stability violations as well as events with no significant signal changes compared to the steady-state conditions (e.g. trips of weakly loaded generators).

3 | CLASSIFICATION MODELS CONSIDERING UNKNOWN DISTURBANCE EVENTS

3.1 | Basic problem formulation

Based on previous work in [27], a single prediction model approach is used to identify and locate disturbances by analysing measurement signals from several PMUs, which are distributed over the grid. The analysis of the spatiotemporal relationships between different PMU signals allows to classify disturbance events, which are not directly observable by PMUs and therefore make the classification approach somewhat independent from a specific PMU placement. With that, the predictors or input matrix $\boldsymbol{X} \in \mathbb{R}^{T \times M}$ for the classification model consists of M measurements from multiple PMU sensors with a fixed number of T timesteps. The targets or class labels y contain the type y_{Type} and location y_{Loc} of the disturbance such that $y = [y_{\text{Loc}}; y_{\text{Type}}]$. All K training classes C are summarised in the class configuration $\overline{C} = [C_1, \ldots, C_K]$.

3.2 | Siamese recurrent neuronal networks with double-sigmoid classifiers

Siamese neural networks (SNN) are widely used for face recognition, person re-identification or object tracking tasks to compute discriminative embeddings [28, 29] or to directly perform classifications [30, 31]. SNNs use a predefined set of known class examples (support set) and compute the distance or similarity between these support instances and a new instance (query instance) to estimate its class affiliation. In contrast with traditional classification approaches (e.g. Softmax-Regression), SNN-based feature embeddings can lead to a better generalisation performance due to a better representation of the intra-class and inter-class variances [28, 32]. Additionally, class rejection mechanisms can be implemented in a straightforward way. Within this study, recurrent neural networks are used to efficiently process and analyse the high-dimensional PMU data records and to learn discriminative feature embeddings. The use of gating mechanisms in gated recurrent units (GRU) or long short-term memory (LSTM) cells enables backpropagation over long time periods and

minimises the vanishing gradient problem. GRU or LSTM cells can be easily integrated into SNN architectures, which allows an end-to-end learning of all model components. In this study, GRUs are preferred over LSTMs to minimise overfitting problems. The computation time for recurrent neural network cells is very low compared to other feature extraction techniques (e.g. dynamic time warping (DTW) or time-frequency transformations) due to the simple matrix calculations, which enables a fast online processing and classification of the PMU data streams [27].

The general model architecture of the proposed Siamese recurrent neural network-based classifier is given in Figure 2. The support $\boldsymbol{X}_S$ and query instances $\boldsymbol{X}_Q$ are fed into a GRU layer to compute the corresponding hidden state matrices $\boldsymbol{H}_S$ and $\boldsymbol{H}_Q$.

A subsequent embedding layer extracts the representative feature vectors $\underline{f}_S$ and $\underline{f}_Q$ by the use of attention mechanisms. This is done by analysing the hidden state information of all time steps and not just the last hidden state of the input sequence. For this, attention weights α^t are calculated with a scoring function, which computes a score value s^t for each hidden state $\underline{h}^t$. The score values are squashed into the range [0,1] with a softmax function to obtain the attention weights as follows:

$$\alpha^t = \frac{exp(s^t)}{\sum_t exp(s^t)}. \tag{1}$$

As given in (2), the final context vector $\underline{c}_\alpha$ serves as a feature vector and is the weighted sum of the attention weights and the hidden states

$$\underline{f} = \underline{c}_\alpha = \sum_t \alpha^t \cdot \underline{h}^t. \tag{2}$$

Within this study, different trainable and non-trainable scoring functions are investigated. The trainable scoring model (derived from [33]) uses a small feedforward neural network to compute the score values. The network can be shared over time steps (*global embedding*—see Equation (3)) or learnt for each time step separately (*local embedding*—see Equation (4)):

$$s^t_{\text{Global}} = \tanh\left(\boldsymbol{W}_S \underline{h}^t + \underline{b}_S\right) \quad \text{and} \tag{3}$$

$$s^t_{\text{Local}} = \tanh\left(\boldsymbol{W}^t_S \underline{h}^t + \underline{b}^t_S\right). \tag{4}$$

The non-trainable scoring model (*cosine embedding*) uses a cosine similarity to compare each hidden state $\underline{h}^t$ with the last hidden state $\underline{h}^T$ given by

$$s^t_{\text{Cosine}} = \frac{\langle \underline{h}^t, \underline{h}^T \rangle}{\left\|\underline{h}^t\right\| \cdot \left\|\underline{h}^T\right\|}. \tag{5}$$

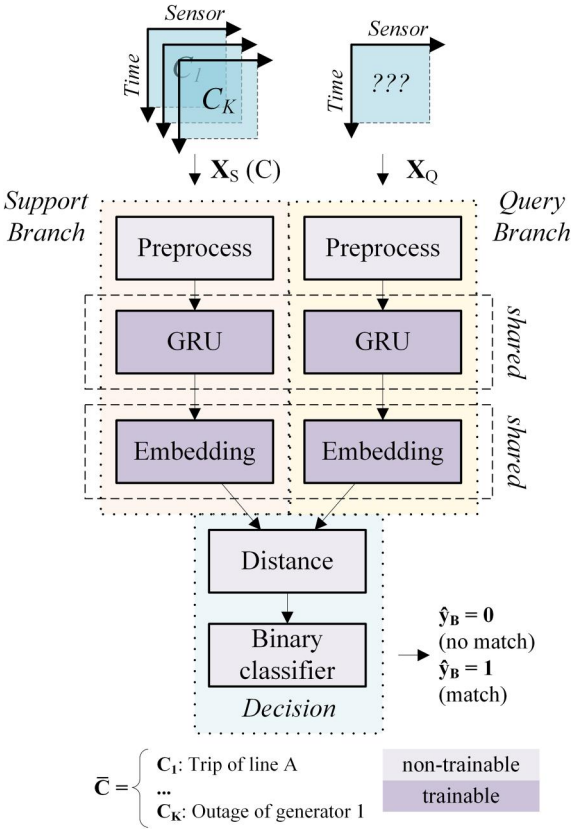

FIGURE 2 Basic concept of a Siamese gated recurrent units based disturbance classification model

Figure 3 shows the model building blocks of the attention-based embedding functions. The GRU and embedding weights $\underline{\theta}_F$ and $\underline{\theta}_E$ are shared between the support and query branch. A comparator computes the Manhattan distance d between the support and query feature vector.

The distance value is passed to a *Double-Sigmoid classifier*, which combines two contrary sigmoid functions. The general idea is illustrated in Figure 4. In the case of a *Single-Sigmoid classifier* (**SingleSig**), high probabilities are computed for low distances such that below the threshold distance d_{Thr} observations of $\boldsymbol{X}_Q$ are assigned to a class of $\boldsymbol{X}_S$ ($p_B \to 1$). If the distance value exceeds d_{Thr}, the observation belongs to another class ($p_B \to 0$). The separation between positive predictions ($y_B = 1$) and negative predictions ($y_B = 0$) mainly depends on this threshold value. Similar to the contrastive or triplet loss formulations [29], the *Double-Sigmoid classifier* (**DoubleSig**) introduces an additional margin to increase the space between the positive and negative predictions and to sharpen the corresponding decision boundaries. This is done by computing the probability for positive predictions p_{Pos} and negative predictions p_{Neg} with two separate sigmoid functions:

$$p_{\text{Pos}}(d) = 1 - \frac{1}{1 + \exp\left(-\frac{d - d_{\text{Thr}}}{\sigma}\right)} \text{ and} \tag{6}$$

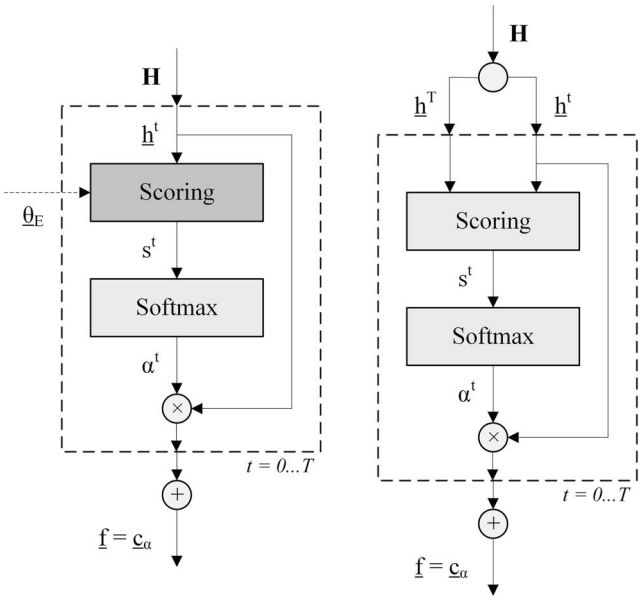

FIGURE 3 Global (left) and cosine (right) embedding models

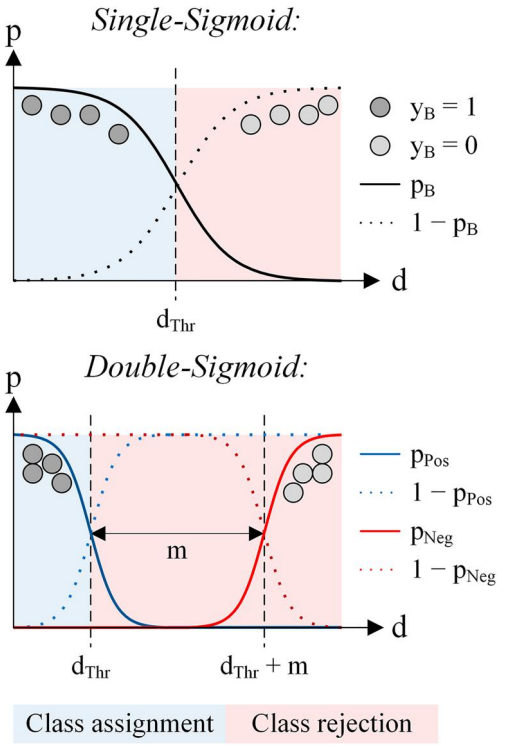

FIGURE 4 Basic principles of *Single-Sigmoid* (top) and *Double-Sigmoid* (bottom) classifiers

$$p_{\text{Neg}}(d) = \frac{1}{1 + \exp\left(-\frac{d-(d_{\text{Thr}}+m)}{\sigma}\right)}. \tag{7}$$

An additional bandwidth parameter σ is used to control and adapt the decision boundaries. The class label $\hat{y}$ is estimated by the outcomes of the positive prediction results for all classes in the support set $\bar{S}$ following a simple One-Vs-Rest strategy given by

$$\underline{p}_{\text{Pos}} = \left[p^{S}_{\text{Pos}}\right]_{S \in \bar{S}} \text{ and} \tag{8}$$

$$\hat{y} = \begin{cases} \max \underline{p}_{\text{Pos}} > 0.5 : \underset{S}{\text{argmax}}\, \underline{p}_{\text{Pos}} \\ \text{reject} \end{cases}. \tag{9}$$

In the training phase, support instances are chosen for each target class y and operational point. Query instances are sampled randomly for each instance of the support set such that the number of equal ($y_B = 1$) and non-equal ($y_B = 0$) pairs is balanced within each training batch. Binary cross-entropy is used as the loss function to train the classification model via backpropagation:

$$\mathcal{L} = -\begin{pmatrix} y_B \log\left(1 - p_{\text{Neg}}(d)\right) \\ + \left(1 - y_B\right) \log p_{\text{Neg}}(d) \\ + y_B \log p_{\text{Pos}}(d) \\ + \left(1 - y_B\right) \log\left(1 - p_{\text{Pos}}(d)\right) \end{pmatrix}. \tag{10}$$

By contrast, the frequently used rank-based losses (e.g. quadruplet loss [34], triplet loss [29] and contrastive loss) require expensive sampling strategies to generate informative training instances and focus on learning discriminative embeddings. Classification-based losses [31] mainly formulate a multi-class softmax function over the squared Euclidean distances without the need for expensive sampling strategies. These approaches strongly rely on the *closed-set* assumption (see Section 1.2) and lack of suitable rejection options.

3.3 | Conventional recurrent neural networks with One-Vs-Rest sigmoid classifiers

Based on the preliminary work in [25], the second classification approach utilises a One-Vs-Rest sigmoid classifier to enable the rejection of unknown classes (**DOCSig**). In accordance with the authors, this classification approach is called deep open classification (DOC) and is designed especially for neural network architectures. In contrast with the previous model, a non-Siamese architecture is required such that the PMU measurements $\boldsymbol{X}$ are directly fed into the classifier via the embedding functions described before. For each class C, a binary classifier is learnt based on a sigmoid function to compute probability p_C, whether the actual feature vector $\underline{f}$ belongs to the class ($p_C \to 1$) or not ($p_C \to 0$). This can be expressed as

$$p_C\left(\underline{f}\right) = 1 - \frac{1}{1 + \exp\left(-\left(\underline{w}_C \underline{f} + b_C\right)\right)}. \tag{11}$$

The class label $\hat{y}$ is estimated by the probability outcomes for all classes in the class configuration $\bar{C}$ following a One-Vs-Rest strategy similar to (8) and (9) from the previous section. A global probability threshold decides whether to reject the actual observation or not and is set to 0.5 by default. Compared to the Siamese GRU-based classifier, individual decision boundaries can be estimated for each class, which leads to a higher model capacity and an increased number of classifier parameters. Additionally, no storing of the support instances is required and a faster computation of the model predictions can be achieved due to the single branch model architecture.

3.4 | Evaluation procedure

This study strongly focusses on the rejection capabilities of the classification models. Therefore, the training and validation set only contains known classes, while the test sets only contain unknown classes. The classification accuracies as well as the $F1$-scores with macro averaging are computed for the training, validation and test set to evaluate the classifiers. For a correct evaluation, an additional rejection class C_{K+1} is added to the set of known classes $\overline{C}$ such that

$$\overline{C}^* = [C_1, \ldots, C_K, C_U] \tag{12}$$

4 | CLASSIFICATION RESULTS AND EVALUATION OF THE REJECTION CAPABILITIES

4.1 | Experimental setup

For a detailed analysis of the model's rejection capabilities, four scenarios are investigated including disturbance events, which differ to some extent from the known training events—see Table 2. *Similar disturbance types* correspond to test events with a high resemblance to the training events (e.g. short circuits at different line positions, load losses at different levels). *Near-by disturbance locations* correspond to test events with a close proximity to the training events (e.g. adjacent lines or stations). Table 3 shows the number of contingencies for the training and test datasets of the scenarios C.1 to C.4. Within this study, the following disturbance types are considered:

- steam power plant outages (SPP),
- base load power plant outages (BLPP),
- solar plant power losses at 25%, 50% and 75% compared to the steady state (PV^{XX}),
- line trips (Line),
- three phase-to-ground short circuits at 10%, 50% and 90% of the line length (SC^{XX}), and
- load losses at 25%, 50% and 75% compared to the steady state (LL^{XX}).

TABLE 2 Overview of the test scenarios

Scenario	Disturbance types	Disturbance location
C.1	Unknown and similar	Known
C.2	Unknown and dissimilar	Known
C.3	Known	Near-by locations
C.4	Known	Far-away locations

TABLE 3 Number of training events ('Tr') and test events ('Te') for the scenarios C.1 to C.4

	C.1		C.2		C.3		C.4	
Type	Tr	Te	Tr	Te	Tr	Te	Tr	Te
SPP	2	2	2	3	1	1	2	2
BLPP	2	2	2	3	3	3	2	2
PV^{25}	2	2	2	2	1	1	1	2
PV^{50}	4	1	4	-	2	2	2	3
PV^{75}	2	2	2	1	1	1	1	2
Line	4	-	4	4	4	4	4	4
SC^{10}	2	2	2	2	2	2	2	2
SC^{50}	1	2	1	1	1	1	1	1
SC^{90}	1	2	1	1	1	1	1	1
LL^{25}	1	1	1	1	1	1	1	1
LL^{50}	2	2	2	2	2	2	2	2
LL^{75}	1	2	1	-	1	1	1	1

The disturbance classification models use the frequency and voltage magnitudes from the busbars of 13 selected substations to build up the input matrix $\boldsymbol{X}$. A fixed timespan of 2 s at a reporting rate of 25 f.p.s. leads to $T = 50$ input timesteps. The training and test dataset comprises three operational points with different load and generation conditions. For all classification models, early stopping is used to control the number of iterations (max. 1000) and RMSprop is used as an update rule for the backpropagation algorithm with a fixed batch size of 50 instances. Additional hyperparameter settings are discussed in the subsequent paragraphs 4.2 and 4.3. The calculations are performed in Python using the Keras library [35] on a CPU machine with a 2.8 GHz CPU processor and 16 GB RAM.

4.2 | Parameter tests for the siamese GRU-based double-sigmoid classifier

For the evaluation of the DoubleSig classifier, different parameter tests with variation of the embedding function, the number of hidden units and the classifier parameters are performed. To limit the amount of necessary simulation runs, the learning rate ($lr = 0.005$) and the bandwidth parameter ($\sigma = 0.2$) are fixed for all the tests. The average training, validation and test accuracies for scenarios C.1 to C.4 are given in Table 4. For all simulation runs, the training and validation

accuracies achieve over 98% but the test accuracies decrease below 50%. The standard deviation for the test accuracies is quite high (12.77%) compared to its average (40.07%). This indicates a high sensitivity of the classification results with regard to the selection of the hyperparameters. The highest losses of the test accuracy occur for scenario C.1 (−84.34% compared to training and −82.03% compared to validation) and the lowest losses of the test accuracy can be observed for scenario C.2 (−47.96% compared to training and −45.25% compared to validation).

The use of the local embedding (mean: 43.67%) achieves higher test accuracies than the use of the global embedding (mean: 42.94%), cosine embedding (mean: 42.84%) or none embedding (mean: 30.83%), whereby none embedding refers to the case when the last hidden state $\underline{h}^{T}$ is used as the feature vector. When the number of hidden units is increased from 15 to 20, the test accuracies remain nearly unchanged.

Large impacts on the training and test accuracies can be observed when changing the threshold distance d_{Thr} and the margin parameter m. The corresponding results of the averaged training and test accuracies are given in Figure 5 and Figure 6. Irrespective of the margin, low threshold values lead to a decrease of the distances for equal and non-equal pairs as well as a minimisation of the intra-class and inter-class variances of the data points. On the one hand, a low inter-class variation deteriorates the correct differentiation between the known classes and decreases the training accuracy up to 5% especially for low threshold distances (see Figure 5). On the other hand, a low intra-class variation significantly improves the rejection capability and test accuracy of the classifier up to 20% (see Figure 6), because equal data points move closer together and the open space increases.

A proper adjustment of the margin parameter can additionally enhance the test accuracy especially in case of low threshold distances (see again Figure 6). In general, a high margin leads to larger differences between equal and non-equal pairs and inter-class variances, which increase the training and test accuracies. Nonetheless, too high margin values diminish the classification accuracies due to a poor model convergence. For a better illustration, Figures A1–A3 show the histograms of the distances between all training and test data points for selected *SingleSig* and *DoubleSig* classification models.

TABLE 4 Average accuracies for all parameter tests of the *DoubleSig* classifier

	Accuracy		
Scenario	**Training**	**Validation**	**Test**
C.1	0.9821	0.9590	0.1387
C.2	0.9868	0.9597	0.5072
C.3	0.9867	0.9576	0.4793
C.4	0.9859	0.9587	0.4777
Average	0.9854	0.9587	0.4007
STD	0.0461	0.0534	0.1277

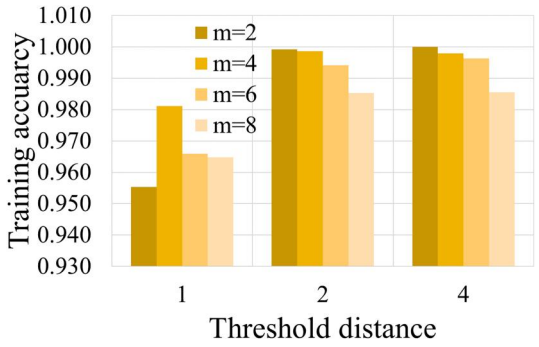

FIGURE 5 Average training accuracies of the *DoubleSig* classifier for different threshold distances and margins

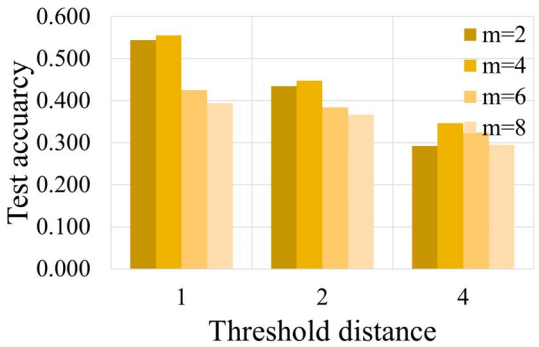

FIGURE 6 Average test accuracies of the *DoubleSig* classifier for different threshold distances and margins

4.3 | Parameter tests for the conventional GRU-based One-Vs-Rest sigmoid classifier

The average training, validation and test accuracy for the *DOCSig* classifier is shown in Table 5. Compared to the *DoubleSig* classifier, the validation accuracies (mean: 95.98%) and test accuracies (mean: 12.24%) are quite low. In this case, too, scenario C.1 shows the highest losses of the test accuracy (−93.89% compared to training and −90.18% compare to validation) and scenario C.2 shows the lowest losses of the test accuracy (−82.50% compared to training and −78.65% compared to validation). Compared to the case of the *DoubleSig* classifier, parameter changes have only a marginal impact on the accuracy results. Variations of the embedding function lead to test accuracy changes between −0.35% and +1.16%. The validation accuracies can be increased using the *global embedding* (+2.56%) and the *cosine embedding* (+2.92%) compared to *none embedding*.

4.4 | Performance analysis for the best hyperparameter settings

Based on the parameter tests in Sections 4.2 and 4.3, stratified 10-fold cross-validation runs are performed for the best hyperparameter settings of the *DoubleSig* and *DOCSig* model.

TABLE 5 Average accuracies of the *DOCSig* classifier

	Accuracy		
Scenario	**Training**	**Validation**	**Test**
C.1	0.9974	0.9603	0.0585
C.2	0.9961	0.9576	0.1711
C.3	0.9991	0.9636	0.1475
C.4	0.9954	0.9576	0.1126
Average	0.9970	0.9598	0.1224
STD	0.0068	0.0268	0.0358

Additional results of the *SingleSig* classification model are investigated for a better comparison of the findings. In contrast with the parameter tests, the accuracies and F1-scores with macro averaging are evaluated and discussed. Table 6 provides the chosen hyperparameters that are used within scenarios C.1 to C.4.

For a more detailed comparison of the proposed model architectures, the number of trainable parameters as well as the average training and inference times have been evaluated. Here, the inference time corresponds to the elapsed time to predict the class label $\hat{y}$ from a single observation $\boldsymbol{X}$ (in case of the *DOCSig* model) or $\boldsymbol{X}_{\mathrm{Q}}$ (in case of the *SingleSig* and *DoubleSig* model).

Compared to the *DOCSig* model, the *SingleSig* and *DoubleSig* models lead to an increased training effort regarding the number of trainable parameters as well as the training times. It should be noted that the *SingleSig* and *DoubleSig* models are trained in a binary classification regime with randomly sampled positive and negative queries, which lead to a different training and convergence behaviour than in the case of the *DOCSig* model. The inference times required for the online classification of grid disturbances are of particular interest in the context of this study. Considering a PMU reporting rate of 25 f.p.s. or 40 ms, all classification models show sufficient time periods to predict the disturbance event from given PMU input signals.

Figure 7 shows the averaged validation and test accuracy results for all three classification models. The corresponding, averaged F1-scores are given in Figure 8. The *SingleSig* classifier shows the highest validation accuracy of all four test scenarios with 99.23% compared to the *DoubleSig* classifier with 95.08% and the *DOCSig* classifier with 95.07%. The same applies for the macro F1-scores of the validation data (*SingleSig*: 94.55%, *DOCSig*: 91.47%, *DoubleSig*: 91.43%). In the case of scenario C.1, the rejection of unknown classes is most difficult and very low accuracies (<25%) can be observed. As a consequence, test instances of similar disturbance types at known disturbance locations seem to share a lot of common features with the training instances. When the disturbance type or the disturbance location differs from the training examples (scenarios C.2 to C.4), the test results improve significantly for all three classification models. In all four scenarios, the test accuracies of the *DoubleSig* model (mean: 56.23%) are remarkably higher than those of the *SingleSig* model (mean: 32.47%) and the *DOCSig* model (mean: 13.84%). This especially accounts for the test accuracies of scenario C.2 (*DoubleSig*: 71.22%, *SingleSig*: 43.28%, *DOCSig*: 17.76%). Because there are no observations from known classes in the test data, the macro F1-scores are at a very low level between 1% and 3% for all three classification models. As a consequence, the test results suggest a superior performance of the *DoubleSig* classifier compared to the *SingleSig* and *DOCSig* classifier for a robust rejection of unknown disturbance events combined with correct recognition and differentiation of known disturbance events.

The results discussed so far assume a fixed probability threshold of 0.5 to accept or reject observations. Changing this value may strongly affect the classifier's ability to recognise known and unknown disturbance events. To evaluate this, the validation and test accuracies of the *DoubleSig* and *DOCSig* models are computed for different probability thresholds for scenario C.2—see Figure 9. In both cases, the test accuracies increase linearly within a wide probability range between 0.05 and 0.9 without significant changes of the validation accuracies. Beyond probability thresholds of 0.9, the number of misclassifications in the validation set increases drastically. For both models, raising the probability threshold from the default value of 0.5 to a higher level between 0.8 and 0.9 can clearly improve the test accuracies or rejection capabilities.

5 | CONCLUSION AND OUTLOOK

This study proposes the use of Siamese recurrent neural networks for robust disturbance classification based on PMU data. In contrast with existing approaches, the recognition of previously unknown disturbances is explicitly investigated, which leads to an *open-set* recognition problem. A *Double-Sigmoid* classifier is proposed and combined with a Siamese GRU network to correctly recognise and distinguish between known and unknown disturbance events. Additional trainable and non-trainable attention models are introduced to create representative feature embeddings and to improve the classification results. Within different test scenarios for a large 400 kV power transmission system, the proposed Siamese GRU classifier is compared with a state-of-the-art *open-set* classifier and shows superior performance within different training and test scenario conditions. For a high recognition rate of unknown disturbances, the hyperparameters of the *Double-Sigmoid classifier* need to be selected carefully. Additionally, the test results highly depend on the nature of the unknown events. Unknown and similar event types at known event locations are very hard to recognise with test accuracies below 25%. In contrast with that, high test accuracies over 70% can be achieved for dissimilar event types or events at far-away locations.

Further improvements of the robustness and test performance of the classification models are necessary. Possible

TABLE 6 Classification models with best hyperparameters

Parameter	Value *(DoubleSig)*	Value *(SingleSig)*	Value *(DOCSig)*
Embedding	Local	Local	Cosine
Hidden units	20	20	10
Learning rate	0.005	0.005	0.005
Distance d_{Thr}	2	2	-
Margin m	4	-	-
Bandwidth σ	0.2	0.1	-
# of trainable parameters	3.870	3.870	1.330
Training time (scenario C.2)	0.17 h	0.16 h	0.04 h
Inference time (scenario C.2)	1.45 ms	1.30 ms	0.32 ms

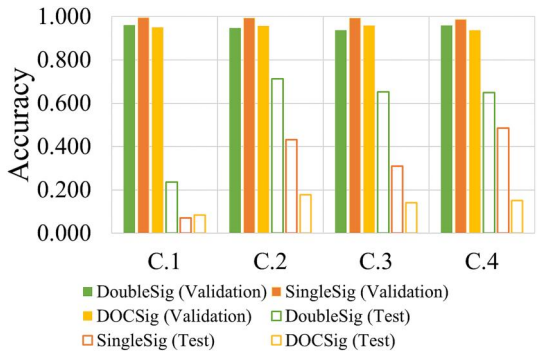

FIGURE 7 Average validation and test accuracies for the best hyperparameter models after 10-fold cross-validation

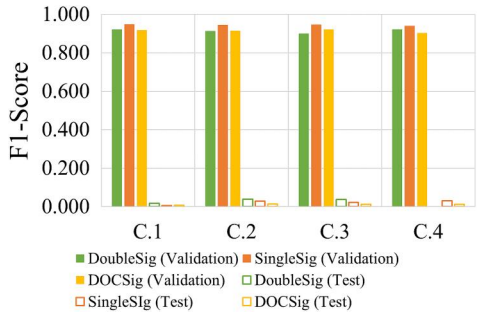

FIGURE 8 Average validation and test F1-scores for the best hyperparameter models after 10-fold cross-validation

alternative approaches could include the integration of 'virtual' unknown classes within episodic training strategies (similar to [32]), the use of range-based loss functions for training of Siamese networks (e.g. triplet loss or quadruplet loss) or the integration of a compact abating probability model to better estimate the *open space* risk.

ACKNOWLEDGMENTS

Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTERESTS

The authors declare no conflicts of interest.

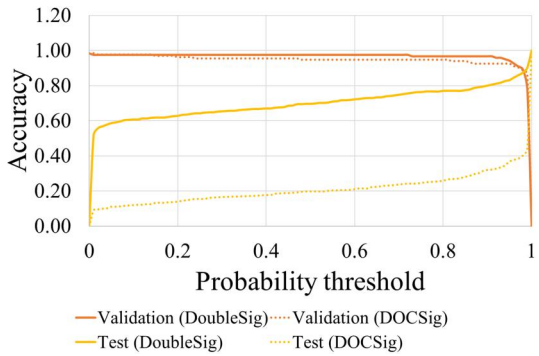

FIGURE 9 Validation and test accuracies for scenario C.2 with different probability thresholds

DATA AVAILABILITY STATEMENT

The author elects to not share data.

ORCID

André Kummerow https://orcid.org/0000-0002-9620-6008

REFERENCES

1. Brosinsky, C., et al.: A new development platform for the next generation of power system control center functionalities for hybrid AC-HVDC transmission systems. In: 2017 IEEE Power & Energy Society General Meeting: 16-20 July 2017, Chicago, IL, pp. 1–5 (2017)
2. Kummerow, A., et al.: A novel framework for synchrophasor based online recognition and efficient post-mortem analysis of disturbances in power systems. Appl. Sci. 10, 5209–15 (2020). https://doi.org/10.3390/app10155209
3. Kummerow, A., et al.: PMU-based online and offline applications for modern power system control centers in hybrid AC-HVDC transmission systems. In: Proceedings of International ETG Congress 2019: 08.–09.05.2019 in Esslingen am Neckar, pp. 405–410. VDE VERLAG GMBH, Offenbach (2019)
4. Usman, M.U., Faruque, M.O.: Applications of synchrophasor technologies in power systems. J. Mod. Power Syst. Clean Energy. 7(2), 211–226 (2019). https://doi.org/10.1007/s40565-018-0455-8
5. Dahal, N., et al.: Event stream processing for improved situational awareness in the smart grid. Expert Syst. Appl. 42(20), 6853–6863 (2015). https://doi.org/10.1016/j.eswa.2015.05.003
6. Bian, J.J., Slone, A.D., Tatro, P.J.: Protection system misoperation analysis. In: 2014 IEEE PES General Meeting: Conference & Exposition; 27–31 July 2014, National Harbor, pp. 1–5. (2014)

7. Konakalla, S.A.R., de Callafon, R.A.: Feature based grid event classification from synchrophasor data. Procedia Comput. Sci. 108, 1582–1591 (2017). https://doi.org/10.1016/j.procs.2017.05.046
8. Cardoso, P.E.A.: Deep Learning Applied to PMU Data in Power Systems, Dissertation. Faculdade de Engenharia, Universidade do Porto, Porto (2017)
9. Yadav, R., Pradhan, A.K., Kamwa, I.: Real-time multiple event detection and classification in power system using signal energy transformations. IEEE Trans. Ind. Inf. 1 (2018). https://doi.org/10.1109/TII.2018.2855428
10. Singh, A.K., Fozdar, M.: Supervisory framework for event detection and classification using wavelet transform. In: Institute of Electrical and Electronics Engineers, Power & Energy Society et al. 2017 - IEEE PES General Meeting, pp. 1–5. (2017)
11. Biswal, M., Brahma, S.M., Cao, H.: Supervisory protection and automated event diagnosis using PMU data. IEEE Trans. Power Deliv. 31(4), 1855–1863 (2016). https://doi.org/10.1109/TPWRD.2016.2520958
12. Brahma, S., et al.: Real-time identification of dynamic events in power systems using PMU data, and potential applications—models, promises, and challenges. IEEE Trans. Power Deliv. 32(1), 294–301 (2017). https://doi.org/10.1109/TPWRD.2016.2590961
13. Fries, S., Rückemann, C.-P. (eds.). Finding needles in a haystack: line event detection on smart grid PMU data streams: ENERGY 2016 The Sixth International Conference on Smart Grids, Green Communications and IT Energy-Aware Technologies : June 26–30, 2016. Curran Associates Inc, Lisbon (2016)
14. Ma, R., Basumallik, S., Eftekharnejad, S.: A PMU-based multivariate model for classifying power system events. ArXiv, abs/1812.00246 (2018)
15. Shaw, P., Jena, M.K.: A novel event detection and classification scheme using wide area frequency measurements. IEEE Trans. Smart Grid, 1 (2020). https://doi.org/10.1109/TSG.2020.3039274
16. Darvishi, H., et al.: Sensor-fault detection, isolation and accommodation for digital twins via modular data-driven architecture. IEEE Sensors J. 21(4), 4827–4838 (2021). https://doi.org/10.1109/JSEN.2020.3029459
17. Mandal, S., et al.: Nuclear power plant thermocouple sensor fault detection and classification using deep learning and generalized likelihood ratio test. IEEE Trans. Nucl. Sci. 1, 1 (2017). https://doi.org/10.1109/TNS.2017.2697919
18. Zhu, Y., Liu, C., Sun, K.: Image embedding of PMU data for deep learning towards transient disturbance classification. In: 2018 IEEE International Conference on Energy Internet (ICEI), Beijing, pp. 169–174. (2018)
19. Li, Z., et al.: A power system disturbance classification method robust to PMU data quality issues. IEEE Trans. Ind. Inf. 1, 1 (2021). https://doi.org/10.1109/TII.2021.3072397
20. Mahapatra, K., et al.: Power system disturbance classification with online event-driven neuromorphic computing. IEEE Trans. Smart Grid, 1 (2020). https://doi.org/10.1109/TSG.2020.3043782
21. Scheirer, W.J., et al.: Toward open set recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35(7), 1757–1772 (2013). https://doi.org/10.1109/TPAMI.2012.256
22. Rocha, A., Scheirer, W.J.: The open set recognition problem and its implications and opportunities in visual computing. Forensics Secur (2016)
23. Bendale, A., Boult, T.: Towards open world recognition. https://arxiv.org/pdf/1412.5687
24. Bendale, A., Boult, T.: Towards open set deep networks. http://arxiv.org/pdf/1511.06233v1
25. Shu, L., Xu, H., Liu, B.: DOC: deep open classification of text documents. https://arxiv.org/pdf/1709.08716
26. Zhang, X.-Y., Liu, C.-L., Suen, C.Y.: Towards robust pattern recognition: a review. Proc. IEEE. 108(6), 894–922 (2020). https://doi.org/10.1109/JPROC.2020.2989782
27. Kummerow, A., et al.: Simultaneous online identification and localization of disturbances in power transmission systems. In: 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), Bucharest, Romania, pp. 1–5. (2019)
28. Kaya, M., Bilge, H.Ş.: Deep metric learning: a survey. Symmetry. 11(9), 1066 (2019). https://doi.org/10.3390/sym11091066
29. Hoffer, E., Ailon, N.: Deep metric learning using triplet network. In: Feragen, A., Pelillo, M., Loog, M. (eds.) Lecture Notes in Computer Science. Similarity-Based Pattern Recognition: Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, October 12–14, 2015: Proceedings, vol. 9370, pp. 84–92. Springer, Cham, Heidelberg (2015)
30. Meyer, B.J., Harwood, B., Drummond, T.: Deep Metric Learning and Image Classification with Nearest Neighbour Gaussian Kernels. http://arxiv.org/pdf/1705.09780v3
31. Snell, J., Swersky, K., Zemel, R.S.: Prototypical Networks for Few-Shot Learning (2017)
32. Liu, B., et al.: Few-Shot Open-Set Recognition Using Meta-Learning (2020). https://arxiv.org/pdf/2005.13713
33. Zhou, P., et al.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany, pp. 207–212.
34. Chen, W., et al.: Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-identification. https://arxiv.org/pdf/1704.01719
35. Chollet, F., et al.: Keras. GitHub

How to cite this article: Kummerow, A., Monsalve, C., Bretschneider, P.: Siamese recurrent neural networks for the robust classification of grid disturbances in transmission power systems considering unknown events. IET Smart Grid. 5(1), 51–61. (2022). https://doi.org/10.1049/stg2.12051

APPENDIX

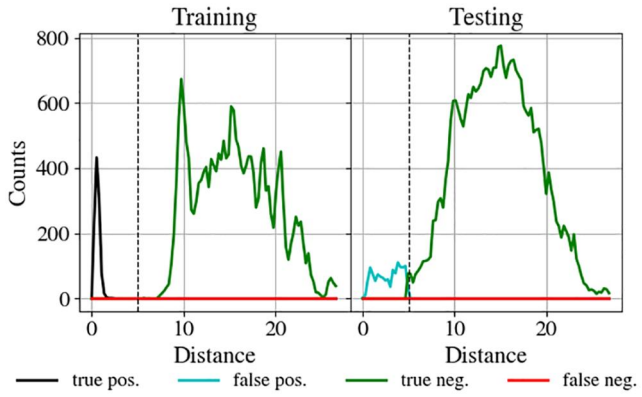

FIGURE A1 Distances for a *SingleSig* classifier at scenario C.2 with 20.37% test accuracy ($d_{\mathrm{Thr}} = 5$ highlighted with '--')

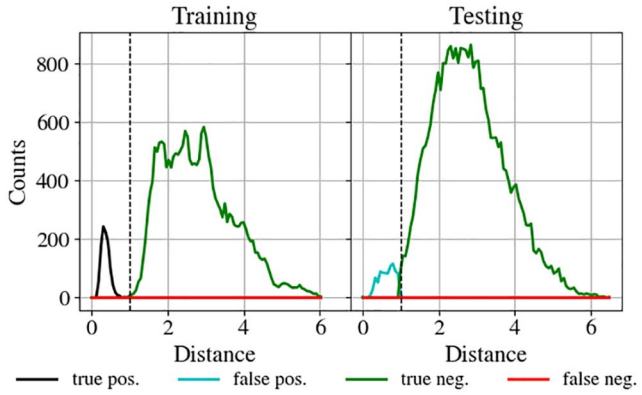

FIGURE A2 Distances for a *SingleSig* classifier at scenario C.2 with 41.29% test accuracy ($d_{\mathrm{Thr}} = 1$ highlighted with '--')

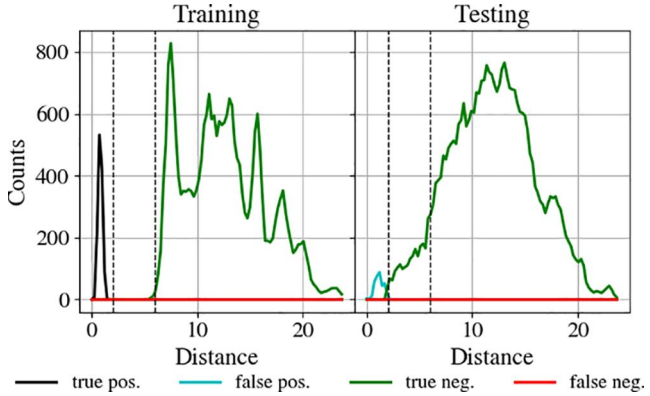

FIGURE A3 Distances for a *DoubleSig* classifier at scenario C.2 with 76.29% test accuracy ($d_{\mathrm{Thr}} = 2$ and $m = 4$ highlighted with '--')