Predicting the onset of cascading events using machine learning

Dr. Panagiotis Papadopoulos, Senior Lecturer (Associate prof.) and UK Research and Innovation Future Leaders Fellow, University of Strathclyde

Contributors: Georgios Nakas, Dr. Tabia Ahmad, Aaron Leavy

IEEE PESGM 2022, Panel session: Enhancing power system operation through online analytics, 19/07/2022

Stability/Security assessment using machine learning

Recent approaches

- Stability prediction – predicting system is stable after a disturbance
  - Rotor angle stability (transient and small signal)
    - Decision trees, neural networks (including convolutional networks and LSTM), support vector machines, etc.
    - Binary and multiclass classification
  - Voltage stability
    - Short term and long term
    - Regression for stability margin
- Dynamic Security Assessment – predicting system is secure for credible contingencies
  - Commonly binary classification secure/insecure
  - Based on static calculations
  - Dynamics included
- Various challenges related to sampling, training dataset, performance increase, etc.

[1] L. Duchesne, E. Karangelos, and L. Wehenkel, "Recent Developments in Machine Learning for Energy Systems Reliability Management," Proceedings of the IEEE, vol. 108, no. 9, pp. 1656-1676, 2020.

Predicting cascading events

Complex interactions including the action of protection devices: *focus on dynamics*

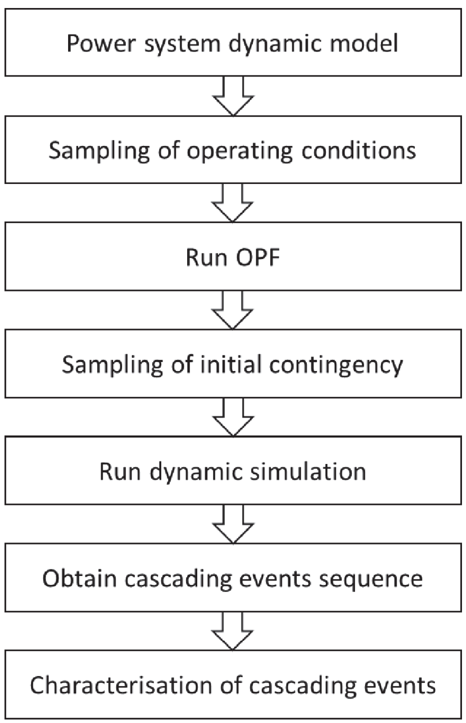
- Move from stability/security prediction to more complex dynamic phenomena
  - Hybrid dynamics (protection, tap changers, etc.)
  - Capture mechanisms related to rotor angle, voltage and frequency stability, including the impact of converters (modelled in an RMS framework)
  - Particular focus to dynamics, especially for last stages before collapse
- Use of machine learning models for time series data
  - Long Short Term Memory (LSTM) Recurrent Neural Networks to capture time dependencies
  - Spatio-Temporal Graph Convolutional Networks (ST-GCN)
  - Temporal Convolutional Networks (TCN)
- Predict a protection device will trip (a cascading event happens)
- Predict the reason for the next trip
- Practical aspects
  - Time window selection (fast prediction before the first cascade)
  - Important measurements to achieve good performance

Cascading failures – brief background

- Complex phenomenon involving various timescales
  - Thermal limits, short/long-term voltage, frequency, angular stability
  - Need to include dynamics
- Static approaches
  - DC or AC power flow
- Dynamic approaches
  - More accurate, high computational effort
- A time series prediction problem
  - Identification of cascading event (Is it happening or not?)
  - What is the reason/mechanism for upcoming trip?
  - Stop the evolution of cascades or mitigate impact

[2] Vaiman et al., "Risk Assessment of Cascading Outages: Methodologies and Challenges," in IEEE Trans. on Power Syst., vol. 27, no. 2, pp. 631-641, May 2012.

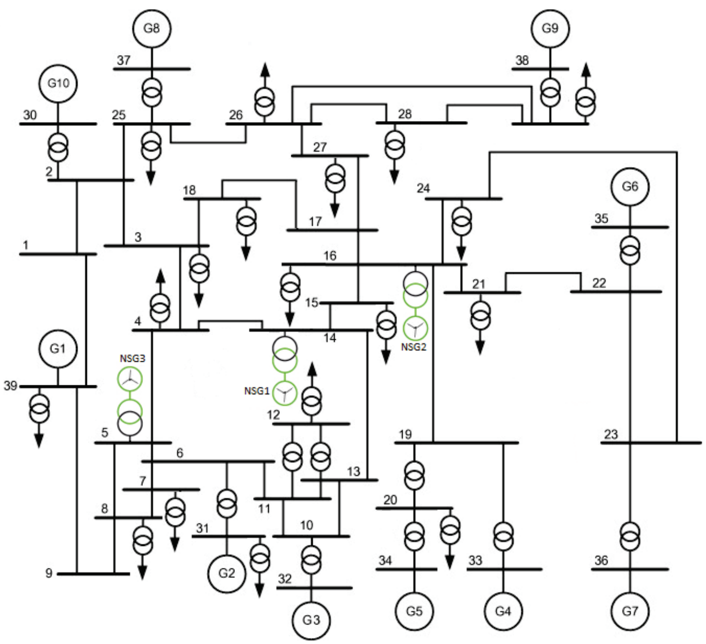
Modelling of cascading failures



- Dynamic model (RMS) to capture dynamics up to 120-180s
  - Capture rotor angle, voltage, frequency dynamics and associated devices/control
  - Including protection devices
  - Operating condition variation: System loading and wind generation
  - Initialised by OPF
- Characterise cascades
  - Capture detail of evolution: Sequence, components (locational aspects) and reason (instability mechanisms)
  - Load loss, length of cascades

Modelling of cascading failures

Modified IEEE 39-bus



- Type 4 wind generation
- Protection devices
  - Synchronous generators
    - Over/under-speed
    - Pole-slip
    - Under-voltage
    - Over-excitation limiter
  - Non synchronous generation
    - Over/Under-voltage
    - Over/Under-frequency
  - Loads
    - UFLS (4 stages)
- Other relevant devices
  - AVRs, PSS, Governors, wind with FRT, (secondary f control)
  - Load tap changers

If

Example rules

- V<V1pu for t>t1ms

or

- V<V2pu for t>t2ms

or

- f<f1pu for t>t3ms

trip

[3] G. A. Nakas, P. N. Papadopoulos, “Investigation of the Impact of Load Tap Changers and Automatic Generation Control on Cascading Events”, PowerTech 2021, Madrid, Spain (online), June 28 – July 2, 2021.
[4] G. A. Nakas, P. N. Papadopoulos, “Investigation of Cascading Events in Power Systems with Renewable Generation”, ISGT Europe 2020, Delft, Netherlands (online), 26-28 October 2020.

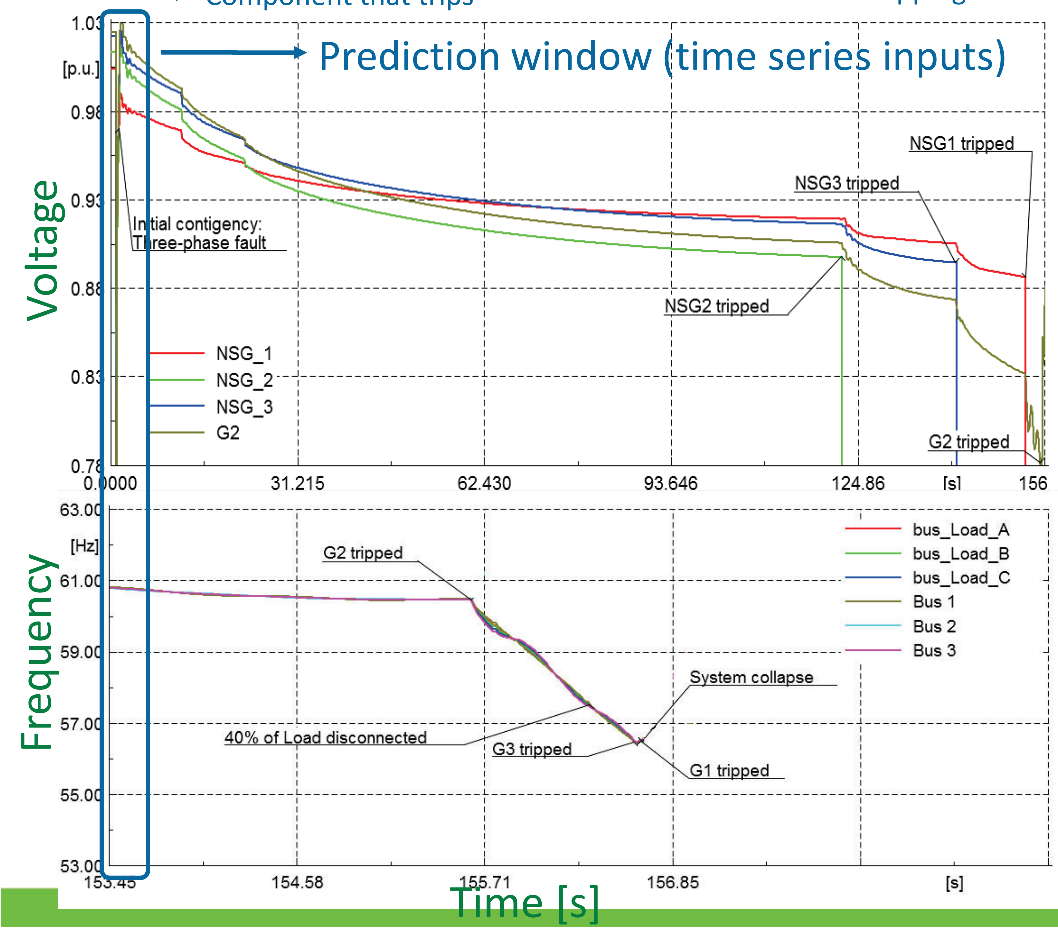
An example of cascading events

1st event 2nd event 3rd event

[(NSG2-UV), (NSG3-UV), (NSG1-UV), (G2-UV), (Load_C-UF), (Load_A-UF), (Load_B-UF), (G3-US), (G1-US)]

Component that trips

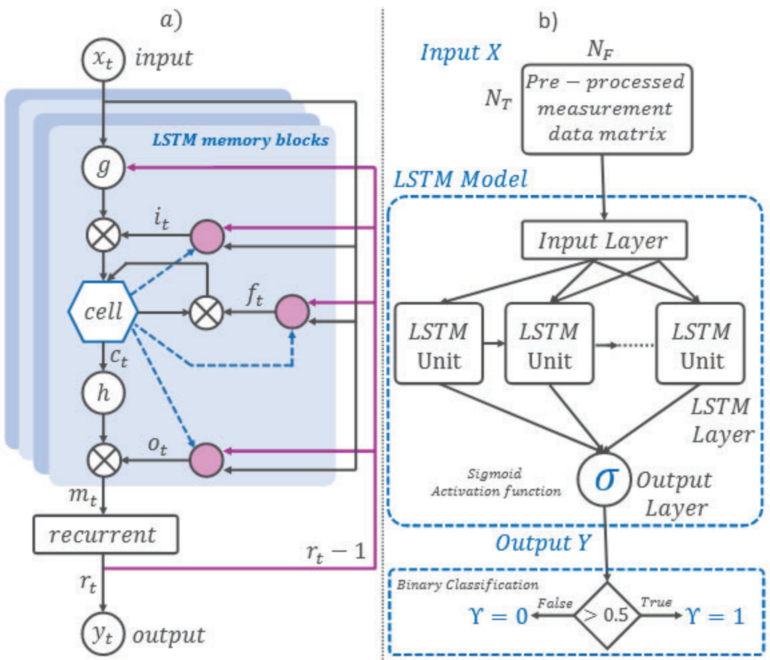
Reason for tripping



Time of Event (s)	Event Description
1	Three – Phase Fault on 90% of Line 5-7
1.07	Line 5-7 trips by Line Distance Protection Relays
122.26	NSG2 disconnected by UV Relay
141.30	NSG3 disconnected by UV Relay
152.80	NSG1 disconnected by UV Relay
155.61	G2 disconnected by UV Relay
156.34	40% of Load C disconnected by UF Relay
156.34	40% of Load A disconnected by UF Relay
156.34	40% of Load B disconnected by UF Relay
156.62	G3 disconnected by Under – Speed Relay
156.65	G1 disconnected by Under – Speed Relay

Data Study Group team. (2022, January 5). Data Study Group Final Report

Long Short Term Memory – LSTM



LSTM network configuration: a) A schematic diagram of a LSTM memory cell. b) Structure of the LSTM model.

- Aim: Predict a cascading failure is about to happen
  - Train ML with simulated time series data *offline*
  - Use in real time utilising available *online* measurement data
  - Binary classification (cascade happens or not)
- Long Short Term Memory – a type of Recurrent Neural Network
  - Very effective for handling time series data
  - (RNNs) have additional recurrent connections in the hidden layers that can maintain information from past states (form a memory)
  - LSTMs can retain memory for longer time and remembers over variable time intervals
  - LSTMs have memory cell with Input, Output and Forget Gates
  - Aims to solve problem of vanishing (or exploding) gradient make it difficult to retain memory for long time
- Useful for cascading failures
  - Phenomena spanning various timescales are involved
- Prediction window
  - PMU measurements (100ms duration)
- Input X: Number of features × Number of time steps
- Output Y: The probability of the appearance of cascading events

Case study – performance of online prediction

Prediction performance

Model	Accuracy(%)	Precision(%)	Recall(%)	F1 score(%)
MLP	92.6	97.3	87.6	92.2
RNN	93.8	94.3	93.2	93.8
LSTM	95.6	96.5	94.6	95.5

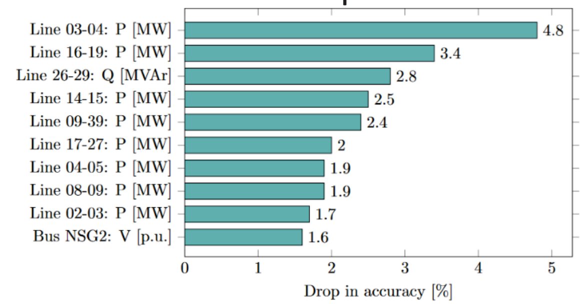
Confusion matrix

	Actually positive (1)	Actually negative (0)
Predicted Positive	473	17
Predicted negative	27	483

Performance with limited availability and measurement noise

LSTM	Accuracy(%)	Precision(%)	Recall(%)	F1 score(%)
10 features	84	92.1	74.4	82.3
15 features	90.9	90.2	91.8	90.9
20 features	94.4	96.1	92.6	94.3
With noise	95.6	96.5	94.6	95.5

Feature Importance

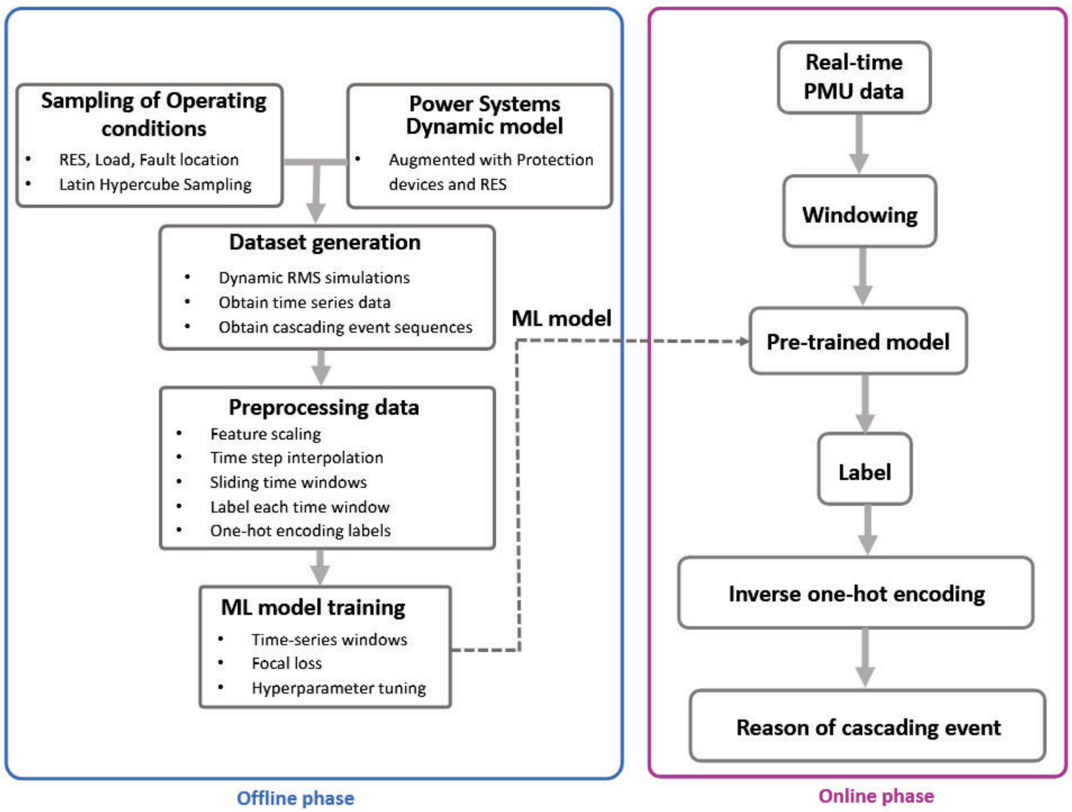


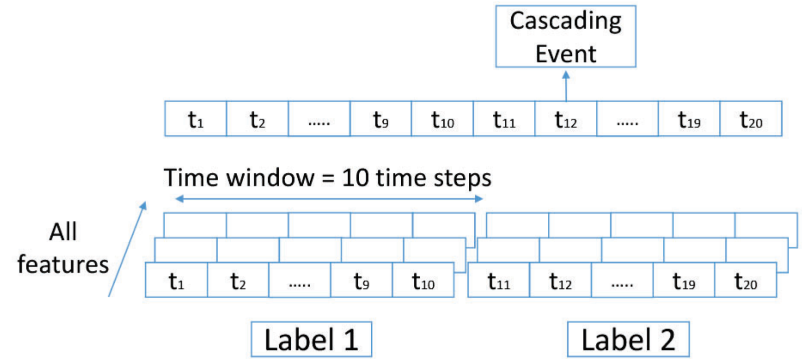
- *Predictions possible with 100ms time window and takes 0.042s*
- The LSTM model exhibits the highest Accuracy Recall and F1 score
- 17 False Positives predictions and 27 False Negatives out of 1000 unseen data samples

Practical aspects:

- Machine learning model performance can vary for different operating conditions
- Feature importance analysis using the permutation technique
  - Identify which features mostly affect the model performance
  - Using only the 20 most important out of the 178 total features leads to a reduction of 1.26% in accuracy
  - Added noise has no effect on the model performance

Predicting the reason for next trip using TCNs



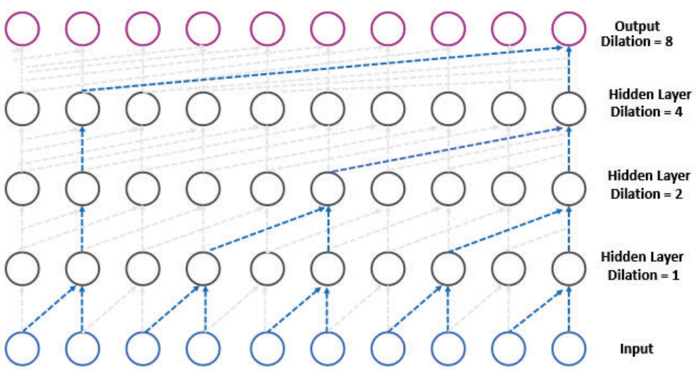


Time windowing process.

- Each time window is assigned a label, according to the reason of the next cascading event
- 7 possible cases of cascading event reasons\ classes: No cascading event, Out-of-step, Over-frequency, Over-voltage, Under-frequency, Under-voltage and Distance protection
  - *Information for corrective control*
- Can be used in an online setting as a continuous process
- LSTM, GRU, TCN
- Focal loss function

[5] G. A. Nakas, P. N. Papadopoulos, “Online Identification of Cascading Event Sequences in Power System using Deep Learning”, to be submitted.

Temporal Convolutional Networks



A stack of dilated causal convolutional layers.

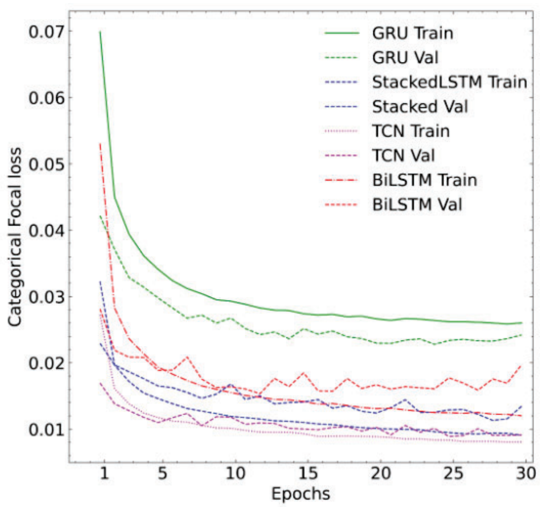
- Temporal Convolutional Network (TCN)
  - Used for sequence modelling
  - Computations can be more parallelized compared to RNNs
  - Do not suffer from vanishing/exploding gradients
  - Less memory requirements compared to gated RNNs
- The operation of TCN can be expressed as:
  $TCN = 1D\ FCN + causal\ convolutions$
- Causal convolutions, which are convolutions that for the output at time t only data points from time t and earlier in the previous layer are considered
- 1D fully-convolutional network (FCN), where the size of each hidden layer is the same as the input size
- TCN uses dilated convolutions to extend the receptive field while using fewer layers

[6] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” arXiv:1803.01271v2 [cs.LG], 2018

Predicting the reason for next trip – case study

MODEL EVALUATION METRICS.

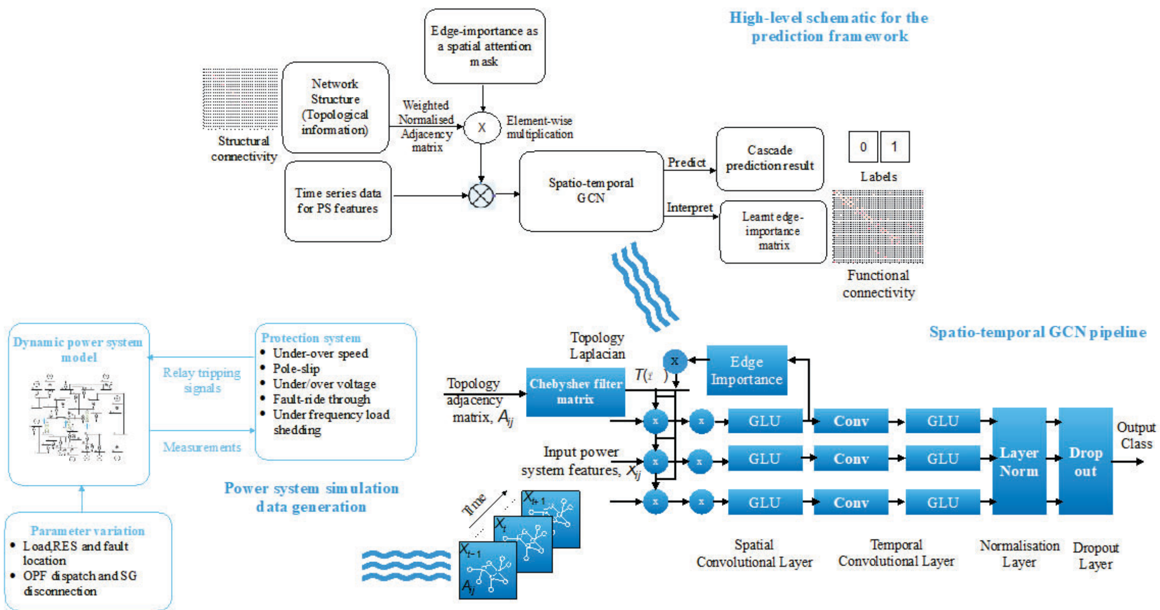
		StackedLSTM			BiLSTM			GRU			TCN		
		Acc (%)			Acc (%)			Acc (%)			Acc (%)		
		97.0			96.9			96.6			97.4		
Reason	Samples	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
No event	272174	98	99	99	98	99	98	97	99	98	98	99	99
Out-of-Step	71	52	17	26	37	27	31	50	37	42	67	28	40
OF	1781	96	95	96	96	95	95	93	95	94	97	96	96
OV	9726	90	88	89	92	89	91	91	85	88	93	90	92
UF	45789	96	93	94	95	90	93	96	89	93	97	92	95
UV	3648	86	67	75	85	79	82	84	76	80	82	83	83
Distance	20475	92	87	89	94	88	91	91	88	89	93	89	91



Learning curves of the models.

- *Predictions using 100ms time window and prediction within 0.077s on average*
- The TCN model performs best, with a total accuracy of 97.4% and achieves the lower losses during training and validation.
- The two LSTM based models, the Stacked LSTM and the Bidirectional LSTM, perform with a similar accuracy, 97% and 96.9% respectively.
- The GRU, which is the simplest model, has the lowest total accuracy, at 96.6%, and the highest training and validation losses.

Graph Convolutional Networks to reveal topological aspects



- Cascading failures in power systems exhibit non-local propagation patterns
- Exploiting graph convolution by adding a one-dimensional convolutional layer behind the graph convolutional layer.
- Binary classification (cascade prediction)
- Weighted adjacency matrix $\tilde{A} = \mathrm{A} + \mathrm{I}$, where A is based on Y-bus
- $\tilde{A}$ replaced by $(\mathrm{A} + \mathrm{I})$*M (element-wise multiplication), where M is learned from data including system dynamic behaviour
- Voltages (100ms time window) at n buses as inputs
- Class weighted binary cross entropy (BCE)

Prediction performance improvement

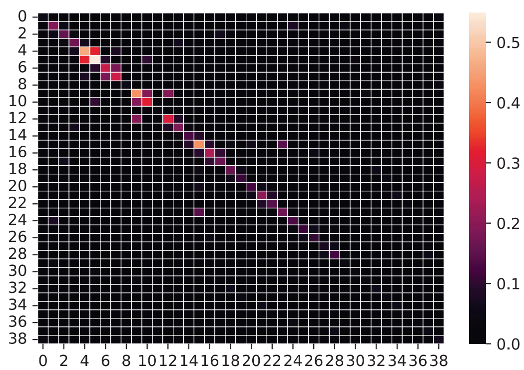
Model	Accuracy(%)	Precision(%)	Recall(%)	F1 score(%)
Vanilla st-GCN	92.56±1.55	84	93.22	88.36
St-GCN+Edge Imp	96.83±1.03	96.45	96.36	96.41

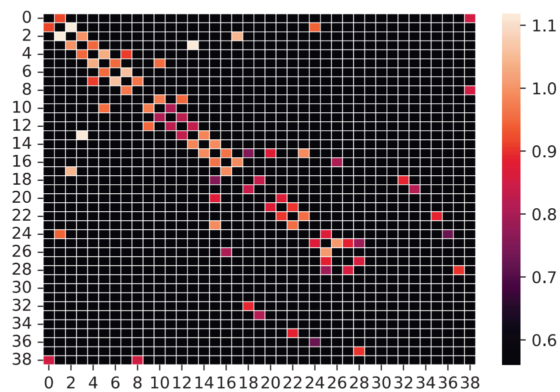
[7] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.
[8] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, 2016.
[9] P. D. Hines, I. Dobson, and P. Rezaei, “Cascading power outages propagate locally in an influence graph that is not the actual grid topology,” IEEE Trans. on Power Syst., vol. 32, no. 2, pp. 958–967, 2016.
[10] T. Ahmad, Y. Xhu, P. N. Papadopoulos, “Predicting Cascading Failures in Power Systems using Graph Convolutional Networks”, *NeurIPS 2021 Workshop on Climate Change AI*

Revealing functional connections from Graph

- ***Edge importance*** **matrix**, ***M as an additional trainable parameter*** inside the *spatio-temporal GCN* can reveal the influence of a set of nodes and edges in model prediction.
- ***Diagonal elements*** of M depict the relative importance of various ***power system buses (graph nodes)*** , while the ***off-diagonal elements of M*** depict the ***importance of power system lines (graph edges)*** in the prediction of cascades.

imag(Y_{bus}) mirroring the **physical topology of the grid**





*M is a dense matrix, with **significant off diagonal elements**, revealing the importance of **not only local but also multi-hop interactions related to dynamic behaviour.***

Deeper colour gradient means the buses/lines are "closer" in their susceptance-based connectivity and functional-connectivity respectively.

[11] T. Ahmad, P. N. Papadopoulos, "Prediction of Cascading Failures and Simultaneous Learning of Functional Connectivity in Power System", submitted.

Conclusions

- Need to capture dynamics for cascading failures
  - Especially for the last steps of collapse
- Machine learning methods for the prediction of the onset of cascading failures
  - More complex dynamic phenomena
  - LSTMs, TCNs, ST-GCNs
- Online application: Binary classification and reason for cascade identification
  - Can enable corrective control actions to stop or mitigate the effect of cascades
- Exploration of topological aspects
  - Complex mechanisms of propagation
- Interaction with ML community

Contact details

UK Research
and Innovation

Panagiotis Papadopoulos
University of Strathclyde

Dr. Panagiotis (Panos) N. Papadopoulos
Senior Lecturer and UKRI Future Leaders Fellow, University of Strathclyde
Web: https://www.strath.ac.uk/staff/papadopoulospanagiotisdr/
LinkedIn: https://www.linkedin.com/in/panagiotis-papadopoulos-92b32a27/
Email: panagiotis.papadopoulos@strath.ac.uk