Predicting Cascading Failures in Power Grids using Machine Learning Algorithms

Rezoan A. Shuvro, Mitun Talukder, Pankaz Das, Majeed M. Hayat

Department of Electrical and Computer Engineering,

Marquette University, Milwaukee, WI, USA

IEEE PES Power & Energy Society®

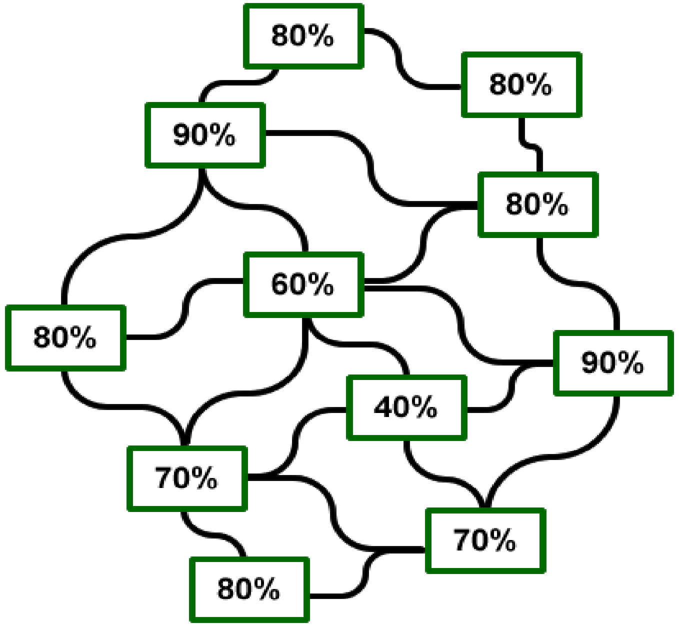
NSF

51st North American Power Symposium

Outline

- *Overview of Cascading Failures in power grid*
- *Problem, Challenges and Related Works*
- *Data Set and Features*
- *Results*
- *Summary*

Cascading Failures in Power Grid : Overview



Network running normally

Source: Wikipedia

A few lines/generators trip initially → Loads shifted to nearby elements in the system → Nearby elements are overloaded beyond their capacity → Elements are tripped; load are shifted onto other elements → Cascading chain of failures starts, which can lead to system blackout

Motivations for this work

In a recent study by cascading failure working group [1], following critical metrics for cascading failures were identified:

- Size and distribution of the blackout size
- Amount of Load shedding and Load shed distribution
- Critical transmission lines

Challenges:

- Limited amount of real-world data
- Generating synthetic cascade data using DC/AC optimal power flow formulations
- Validation
- Limited amount of prior works

Related Works: Very few works that use machine learning to predict cascading failures.

1. Gupta, S., Kambli, R., Wagh, S., & Kazi, F. (2015). Support-vector-machine-based proactive cascade prediction in smart grid using probabilistic framework. *IEEE Transactions on Industrial Electronics*, *62*(4), 2478-2486.
2. Pi, R., Cai, Y., Li, Y., & Cao, Y. (2018). Machine Learning Based on Bayes Networks to Predict the Cascading Failure Propagation. *IEEE Access*, *6*, 44815-44823.
3. Hink, R. C. B., Beaver, J. M., Buckner, M. A., Morris, T., Adhikari, U., & Pan, S. (2014, August). Machine learning for power system disturbance and cyber-attack discrimination. In *2014 7th international symposium on resilient control systems (ISRCS)* (pp. 1-8). IEEE.

Contribution of this work

- Development of a cascading failure simulation framework using MATPOWER
- Classification of cascading failures in power grids using various machine learning algorithms and their comparison.
- Predicting the severity of cascading failures (from amount of lien failures and load shedding) using linear regression technique.

Data set and Features/ Grid Parameters

Dataset:

A 76280 X 11 Matrix with 9 features

Output:

Amount of Load-shedding, L_s
Total Number of line fail, N_p

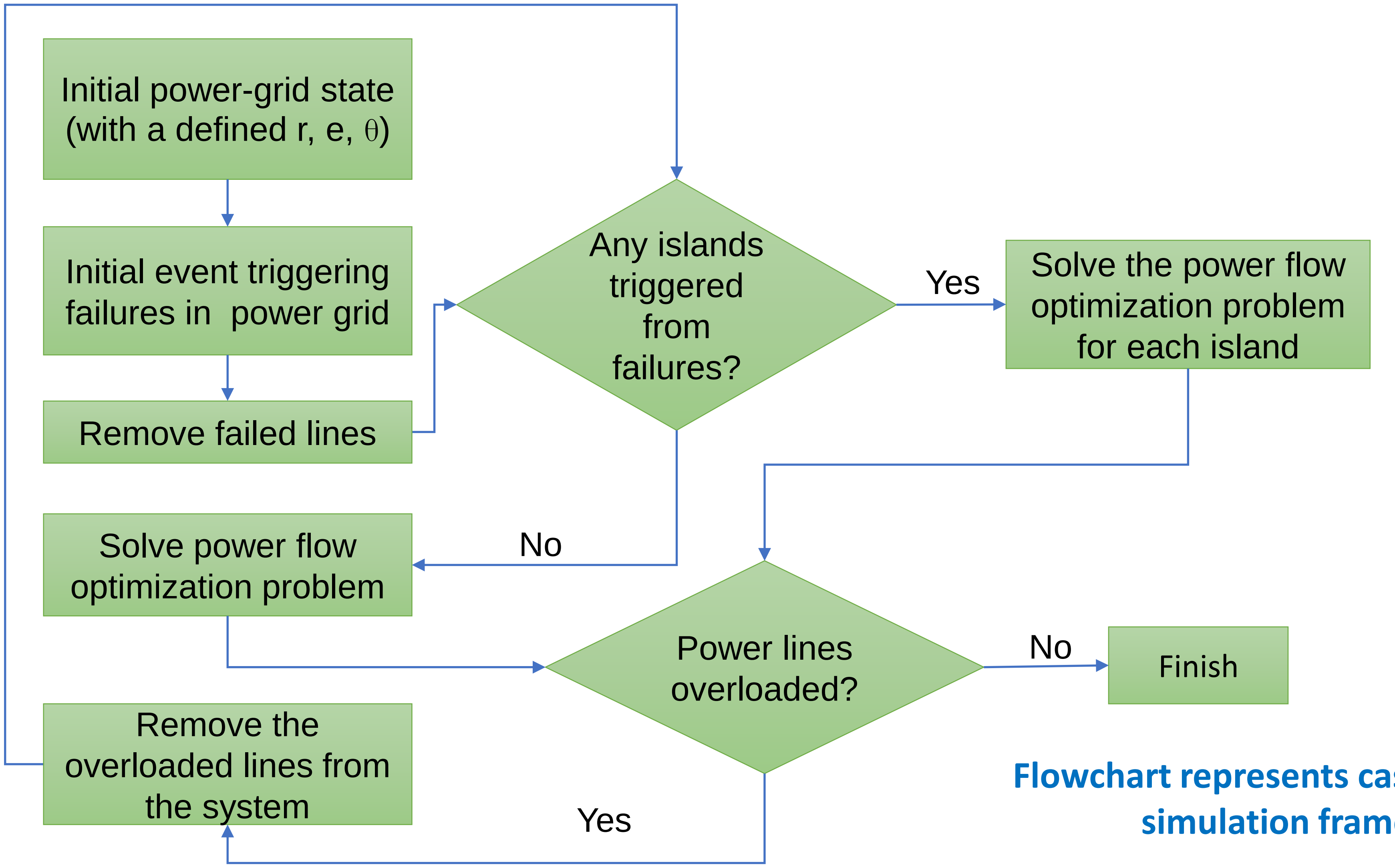
Source:

MATLAB using MATPOWER [1] m-files.

Features	Feature range/values
Power grid loading level, r	[0.5, 0.5, 0.7, 0.8, 0.9]
Constraint on implementing load-Shedding, θ	[0.05, 0.1, 0.15, 0.20, 0.25]
Uncertainty over the capacity estimation, e (capacity estimation error)	[0.05, 0.1, 0.15, 0.20, 0.25]
Fixed fail probability of neighbors, f_p	[0.01, 0.02, 0.03, 0.04, 0.05, 0.06]
avg betwenness rounded ,B	[0, 1.7]
avg SP rounded, S_p	[0, 14]
Number of islands, $N_{islands}$	[1, 54]
flow capacity of the initially failed transmission lines, C_{flow}	[1,1100]
total Number of line fail, N_p	[2,114]
Amount of Load-shedding, L_s	[0,3352]
Total transmission capacity of the failed lines, C	[40, 13160]

1. Zimmerman, R. D., Murillo-Sánchez, C. E., & Thomas, R. J. (2011). MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems*, *26*(1), 12-19.

Data collection procedure



Flowchart represents cascading failure simulation framework

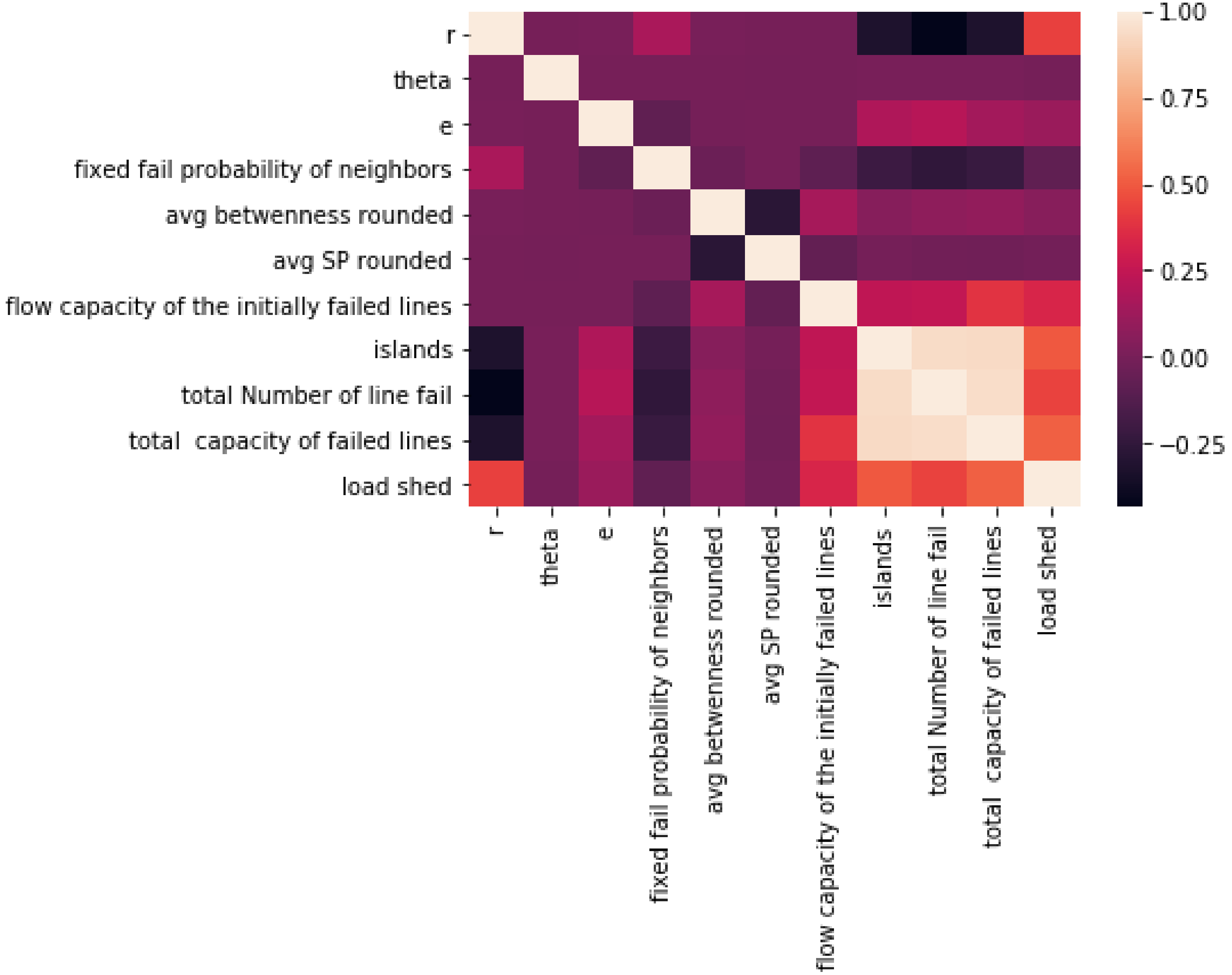
Correlation between Features

Correlation plot shows:

- High correlation between total number of failed lines and total capacity of the failed lines.

- Strong correlation between amount of load shed and loading level, total number of failed lines and total capacity of failed lines

- Topological features (average shortest path and average betweenness) have very low correlation with the outputs(total number of failed lines, amount of load shed)



Algorithms for cascading failure simulation

Algorithm 1 Finding maximum overloaded lines probabilistically during cascading failures

Require: e, PF, Capacity
Ensure: FailedIndex

1: **for** $i \leftarrow 1$ to M **do**
2: $\quad P_{\text{lf}}(i) \leftarrow \text{abs}(\text{PF}(i))/((1-e) * \text{Capacity}(i))$
3: ProbTest $\leftarrow 0$
4: **for** $i \leftarrow 1$ to M **do**
5: $\quad$ **if** rand $<$ LinkProb(i) **then**
6: $\quad\quad$ **if** (ProbTest $< P_{\text{lf}}(i)$) **then**
7: $\quad\quad\quad$ ProbTest $= P_{\text{lf}}(i)$);
8: $\quad\quad\quad$ FailedIndex $= i$
return FailedIndex

Algorithm 2 Solving power flow in each islanded grid

Require: mpc, s, c
Ensure: GD, PF

1: SG = sparse(AdjMatrix)
2: [s, c] = graphconncomp(G)
3: **for** $i \leftarrow 1$ to s **do**
4: $\quad$ struct mpc$[s] \leftarrow$ mpc
5: $\quad$ [PF, GD] = rundcopf(mpc$[s]$)
return GD, PF

Algorithms used to find the maximum overloaded line in one iteration and solving power flow in each island.

ML Algorithms used and Metric for evaluation

- We use the following algorithms for classification:
  - Logistic regression
  - KNN (k nearest neighbor)
  - Random forest
  - Decision tree
  - Support vector machine
  - Adaboost
  - Linear regression (for prediction)

Classification Metric:

- Precession - recall - f1-score

Regression metric:

- mean absolute error - mean square error - root mean square error

Predicting Cascading failures

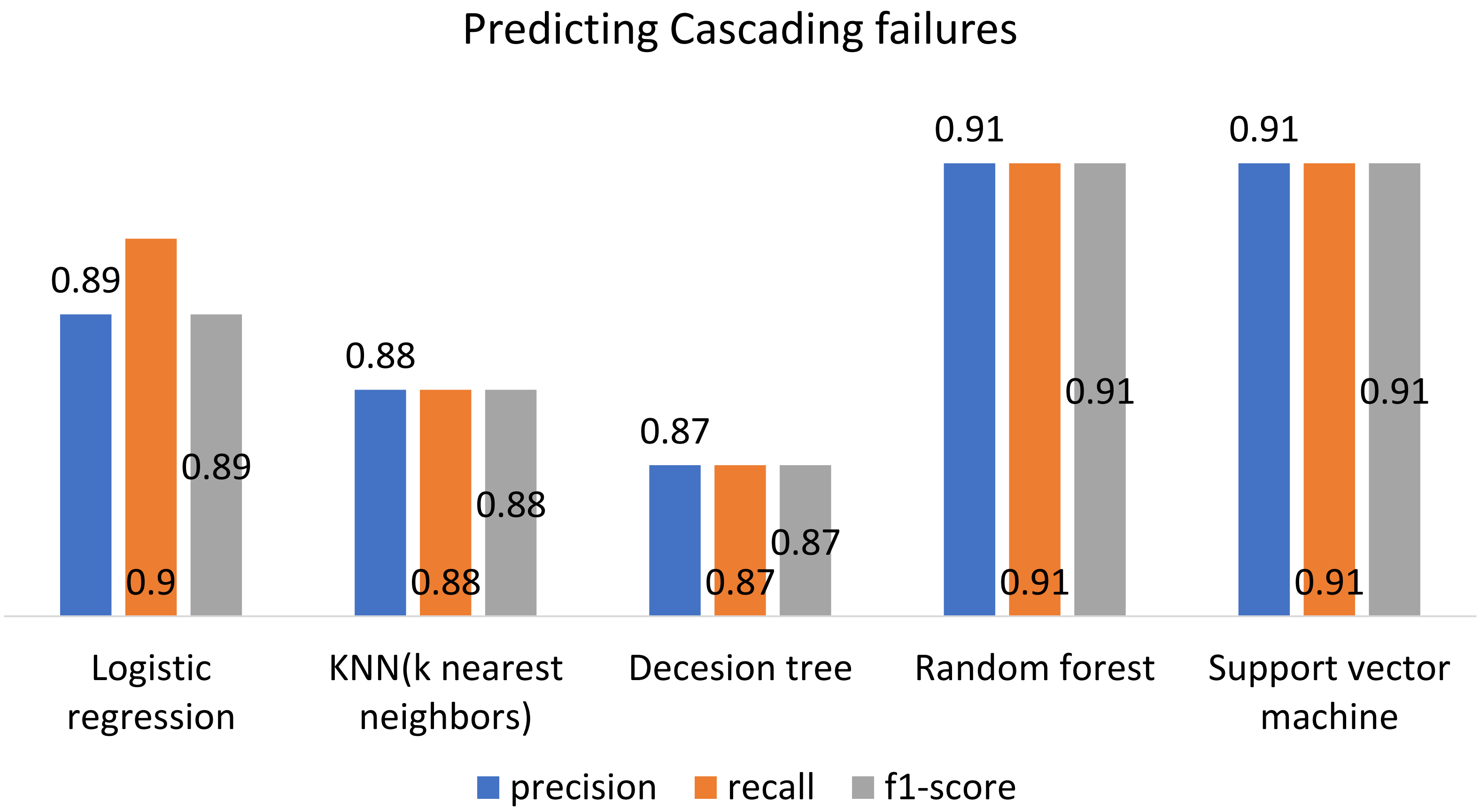
Total Number of transmission lines =186

Initial condition: 2 line failures

No cascade = Less than 10 failures

small cascade = greater than 10 and Less than 25 failures

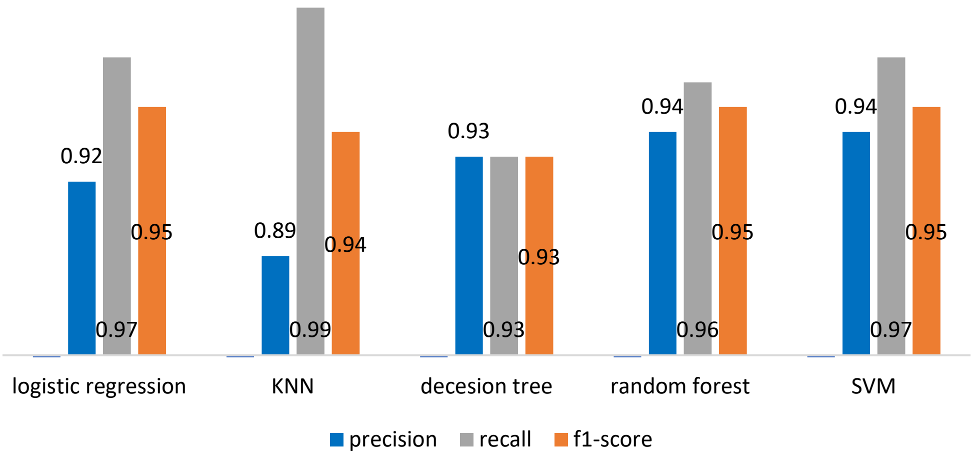
Large cascade = greater than 25 failures



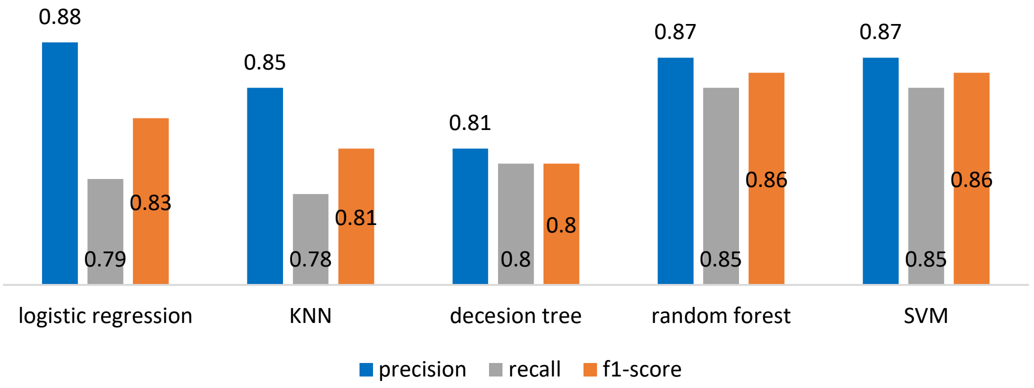
Random forest and SVM gives best accuracy

Predicting Cascading failures

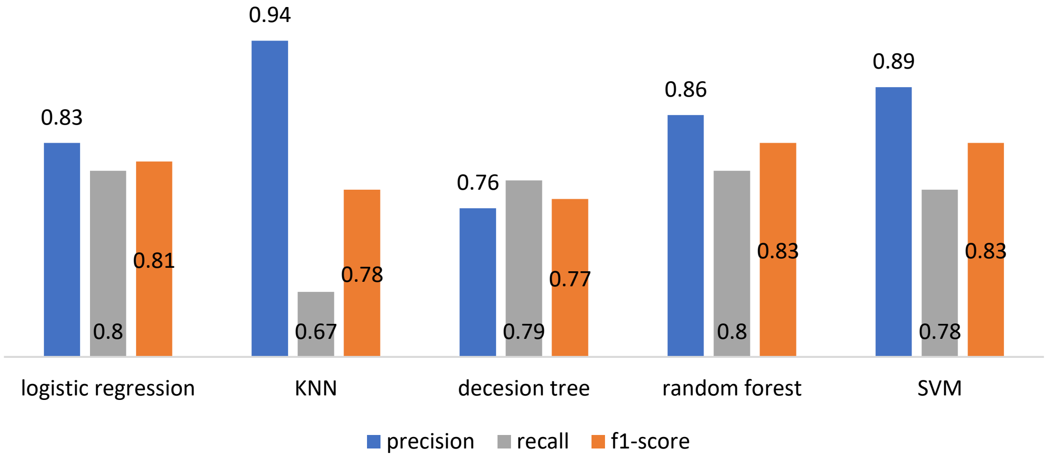
Predicting Cascading failures (No cascade)



Predicting Cascading failures (small cascade)

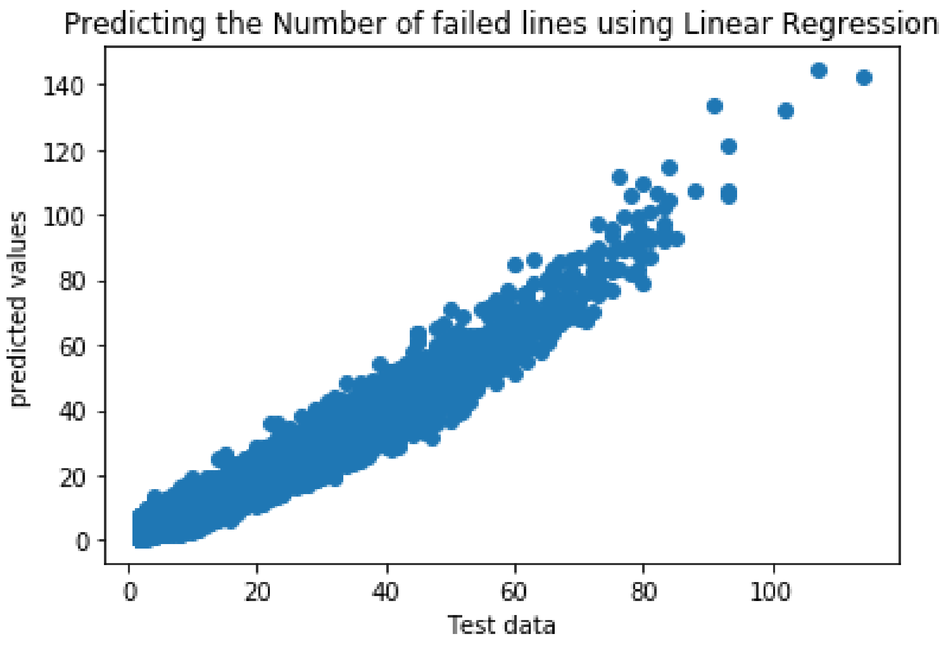


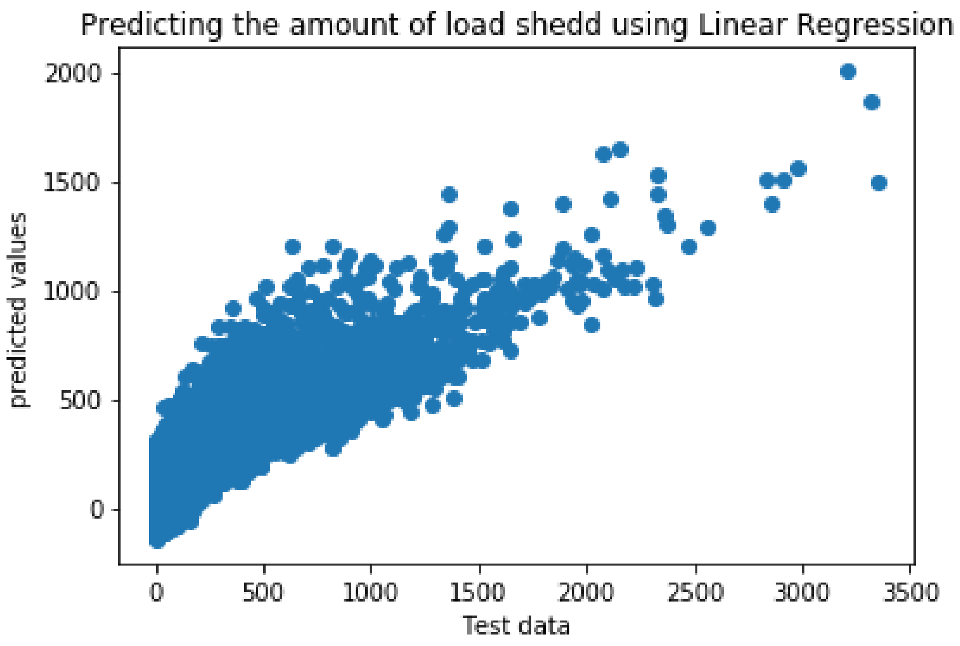
Predicting Cascading failures (Large cascade)



Random forest and SVM gives best accuracy

Linear regression to predict the number of failed lines and the amount of Load shedding





Metrics	Error
Mean Absolute Error	1.90
Mean Square Error	7.07
Root Mean Square Error	2.66

Metrics	Error
Mean Absolute Error	82.92
Mean Square Error	15334.84
Root Mean Square Error	123.83

Linear regression works well for predicting the number of lines

To minimize error for load shed prediction, nonlinear regressor such us SVR, RFR can be used

Summary

- Machine learning algorithms were used to predict cascading failures in power grids.

- Results suggest that all the algorithms gives higher prediction accuracy(Random forest and Support vector Machines are the best with 0.91 precision)

- We predicted the number of failed transmission lines with low error using linear regression

- We predicted the amount of Load shed using linear regression, but the error was high.

- We didn't observe strong correlation between cascading failures and topological features which is consistent with [1]

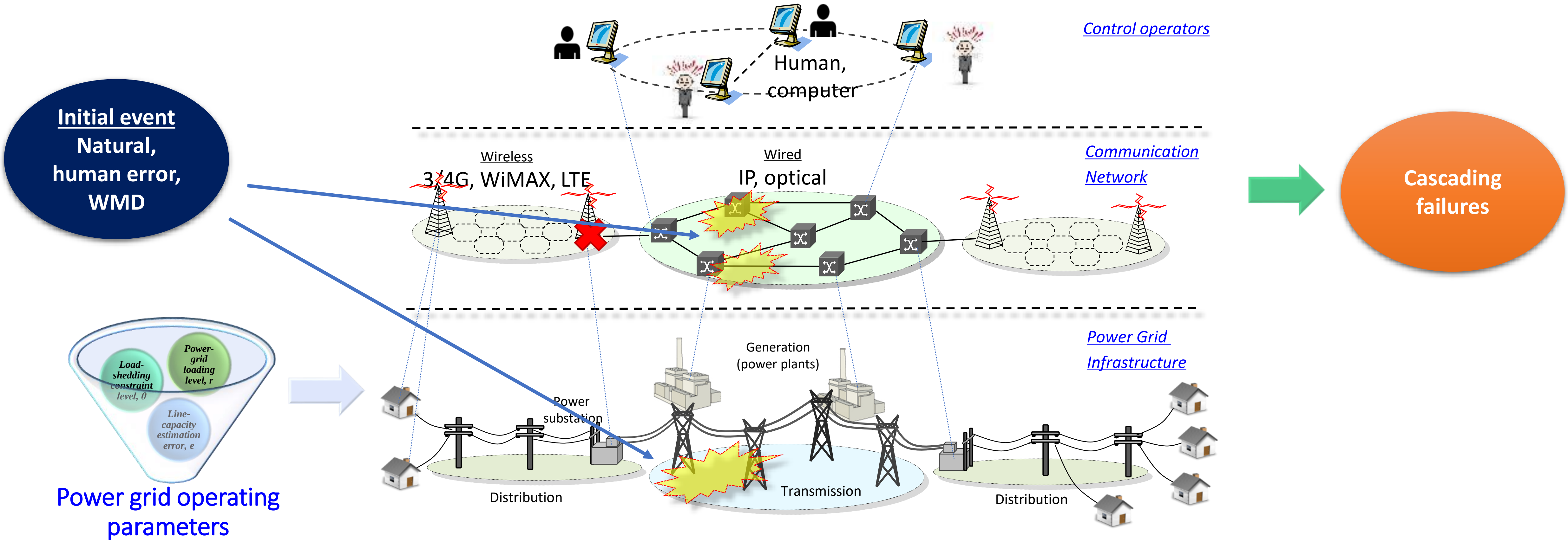
- Future works includes finding the distribution of transmission line failures and amount of load shedding as well as improvement on the accuracies

1. Korkali, M., Veneman, J. G., Tivnan, B. F., Bagrow, J. P., & Hines, P. D. (2017). Reducing cascading failure risk by increasing infrastructure network interdependence. *Scientific reports*, *7*, 44499.

Thank you for your time

Questions?

Overview of complex Power grid Infrastructure



Annex1:

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$\text{F1} = 2 \text{ x } \frac{Precision * Recall}{Precision + Recall}$$