Feature Analysis for PV Fault Detection Neural Network Using Linear PCA and Random Forest

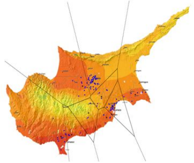
Skyler Verch[1], Gowtham Muniraju[1], Andreas Spanias[1], Yiannis Tofis[2]

[1] SenSIP center ASU [2] KIOS Research and Innovation Centre of Excellence

Solar facility ASU [1]

Smart Monitoring Device (SMD) [1]
Helps diagnose solar faults

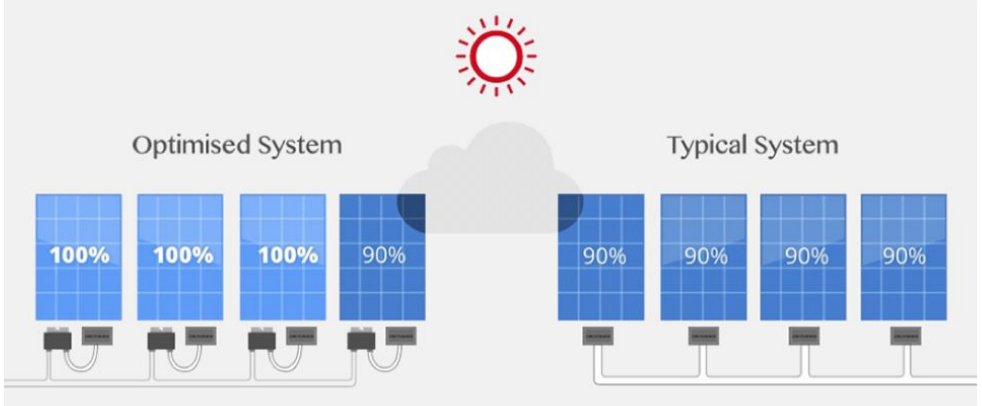
IRES SenSIP center
https://sensip.engineering.asu.edu/nsf-ires-project/



IRES project sponsored by NSF
Award 1854273

Motivation

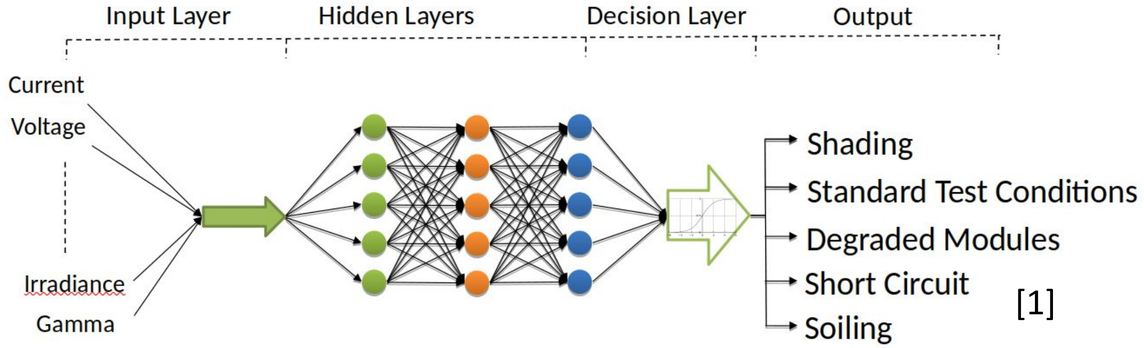
- Solar farms greatly benefit from sensor monitoring systems with the capacity to detect array faults and anomalies.
- Automatic and remote detection and correction of faults elevates the efficiency and robustness of a PV power plant.
- Reduces monitoring and maintenance cost.



Project Aim

- Identify which data features provide the most information to the neural network about classifying solar array faults.
- Reduce redundant or unimportant data

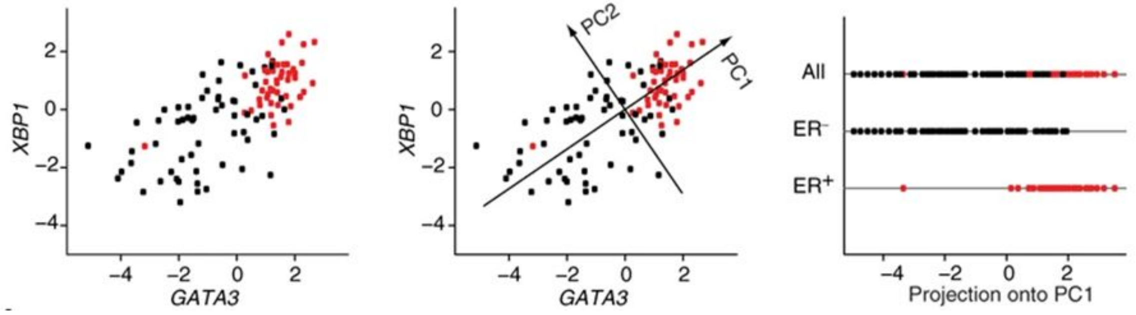
Neural Network architecture



IRES project sponsored by NSF
Award 1854273

Experimental Methods and ML tools

Linear Principal Component Analysis



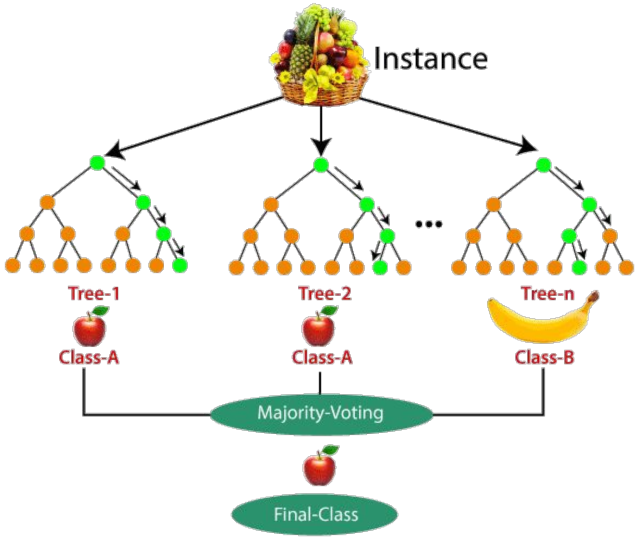
[8]

Implemented via Python

```
In [8]: #Creating neural network
        def NN_model(Dim,pca,X_train, X_test, y_train, y_test):
            x_train = pca.transform(X_train)
            x_test = pca.transform(X_test)
            #transform, just perfroms pca

            OL = 5
            # define the keras model
            model = Sequential()
            model.add(Dense(Dim, input_dim=Dim, activation='relu'))
            model.add(Dense(16, activation='relu'))
            #model.add(Dense(8, activation='relu'))
            model.add(Dense(OL, activation='sigmoid'))

            model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
            model.fit(x_train, y_train, epochs=150, batch_size=128, verbose=1, validation_data=(x_test, y_test)


            # evaluate the keras model
            _, accuracy = model.evaluate(x_test, y_test)
            print('Accuracy: %.2f' % (accuracy*100))

            return accuracy
```

Experimental Methods and ML tools cont...

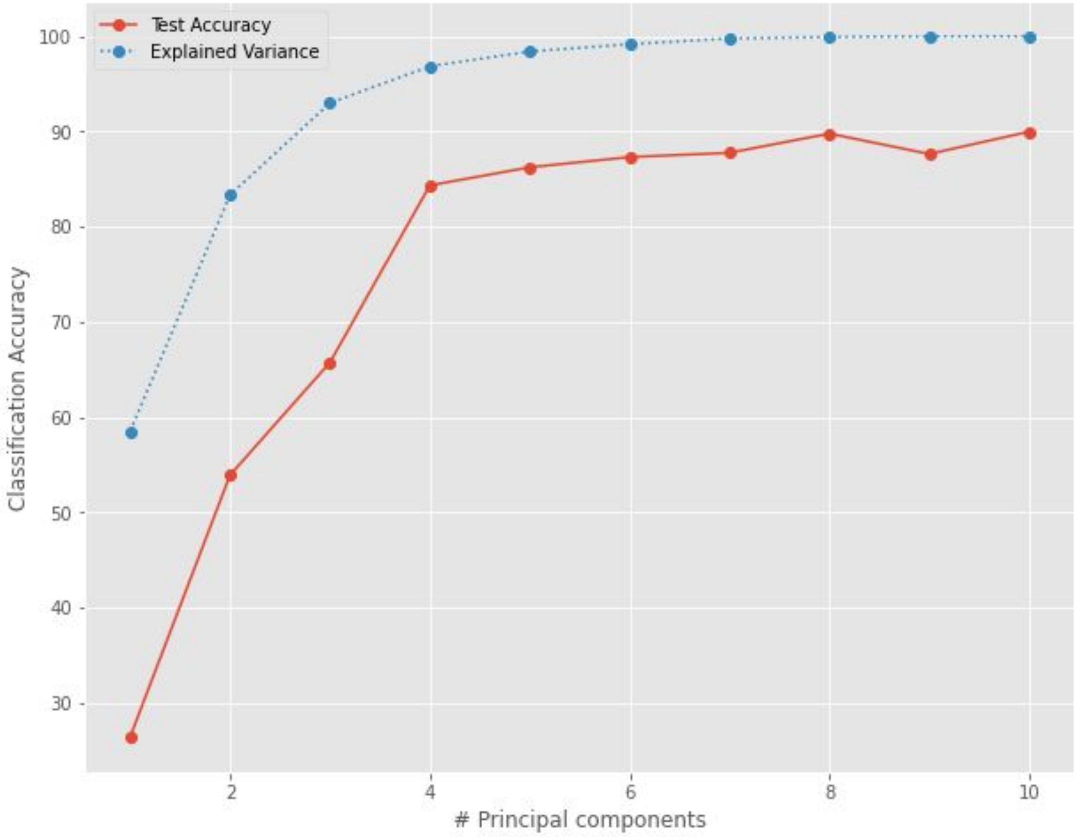
Random Forest Algorithm



Implemented via Python

```
In [6]: feature_names = Data.columns
        feature_names = list(np.array(feature_names[0:9]).astype(str))
        #feature_names = list(np.array(feature_names[0:10]).astype(str))
        print('Feature names : ',feature_names)

        Feature names :  ['DC_Power', 'Vmp', 'Imp', 'Temp', 'Irr', 'Pmp', 'Voc', 'Isc', 'Gamma']

In [7]: forest = RandomForestClassifier(random_state=0)
        forest.fit(X_train, y_train)

Out[7]: RandomForestClassifier(random_state=0)

In [8]: importances = forest.feature_importances_

In [9]: std = np.std([tree.feature_importances_ for tree in forest.estimators_], axis=0)

In [10]: forest_importances = pd.Series(importances, index=feature_names)
```

Linear Principal Component Analysis Results



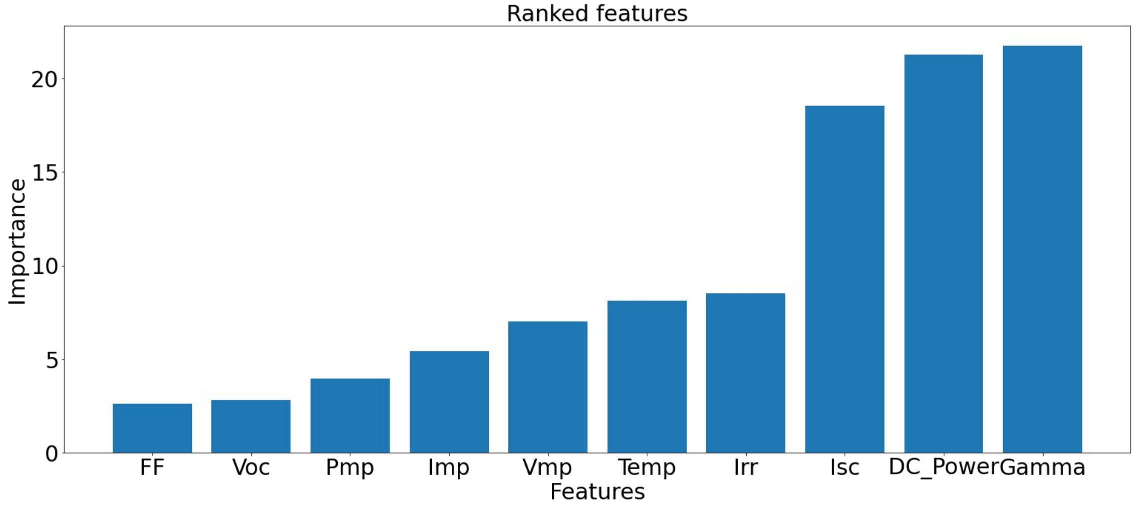
The feature space can be reduced to 4 principal components while retaining 85% accuracy.

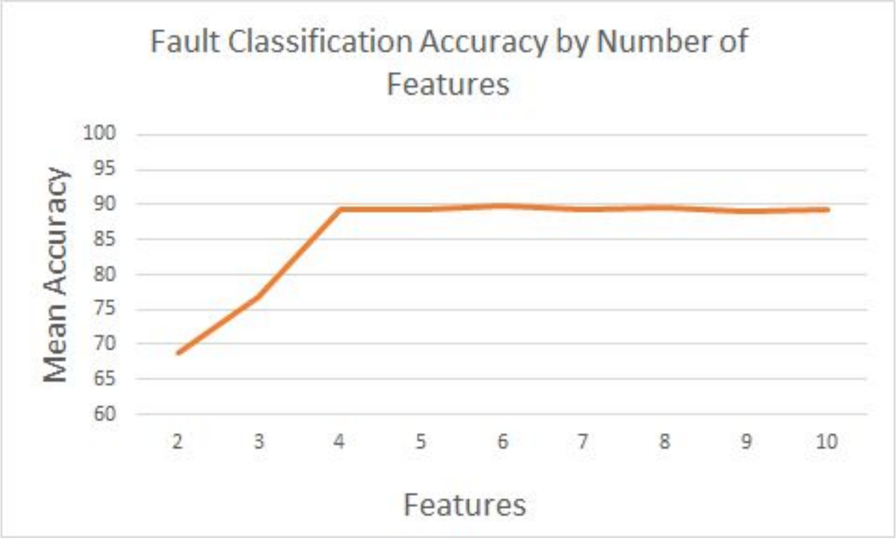
However, the Neural Network does not attain the target accuracy until 8+ principal components.

Data from the National Renewable Energy Laboratory (NREL)

Random Forest Algorithm Results

Using the random forest generated importances, the top 4 features provide the neural network with enough information to classify faults at target accuracy.





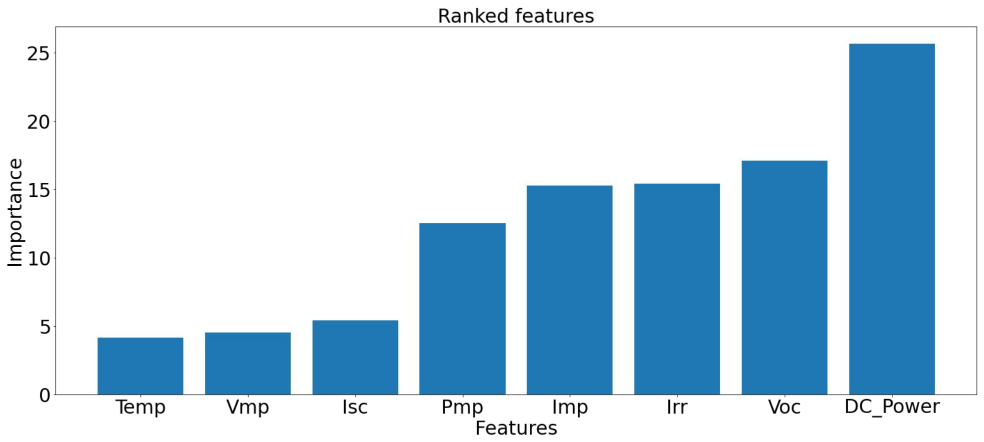
Conclusions

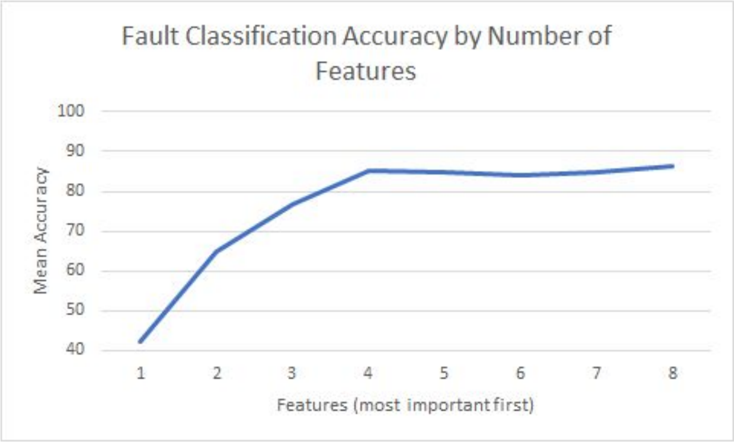
- The feature space can be effectively optimized using both Linear Principal Component Analysis and Random Forest importance rankings.
- The top 4 features in either method accounted for the majority of neural network classification ability.
- Dependent features can play an important role in reducing the feature space

Future Work

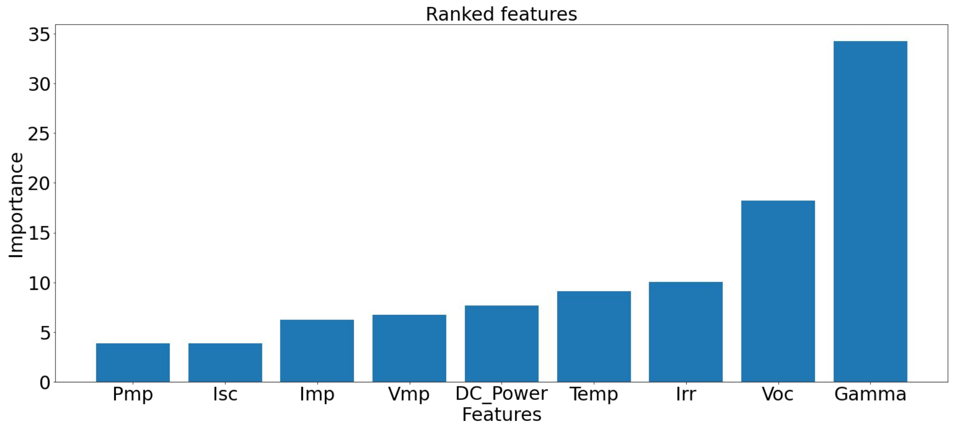
Identifying the role of dependent features (features that are combinations of others)

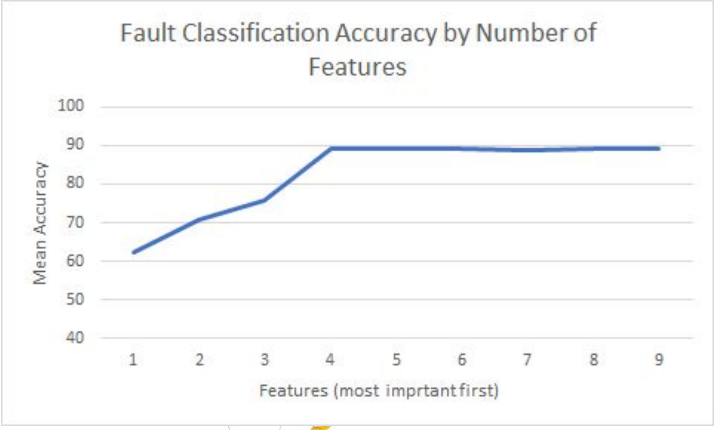
Shown right:
The inclusion of gamma as power divided by irradiance improves overall classification accuracy





Without Gamma





With Gamma

Ira A. Fulton Schools of Engineering
ARIZONA STATE UNIVERSITY

University of Cyprus

IRES SenSIP center
https://sensip.engineering.asu.edu/nsf-ires-project/

IRES project sponsored by NSF Award 1854273

Challenges and Reflection

- I familiarized myself with several facets of solar panel optimization and solar data.
- I learned data science and machine learning techniques, and implemented them in python.
  - Setting up python environment
  - Reading and writing data from excel
  - PCA, random forest and neural network functionality

Ira A. Fulton Schools of Engineering
ARIZONA STATE UNIVERSITY
University of Cyprus
IRES SenSIP center
https://sensip.engineering.asu.edu/nsf-ires-project/
IRES project sponsored by NSF
Award 1854273

References

[1] S. Rao, S. Katoch, V. Narayanaswamy, G. Muniraju, C. Tepedelenlioglu, A. Spanias, P. Turaga, R. Ayyanar, and D. Srinivasan, "Machine learning for solar array monitoring, optimization, and control,"Synthesis Lectures on Power Electronics, vol. 7, no. 1, pp. 1–91, 2020.

[2] Z. Chen, F. Han, L. Wu, J. Yu, S. Cheng, P. Lin, and H. Chen, "Random forest based intelligent fault diagnosis for pv arrays using array voltage and string currents,"Energy conversion and management, vol. 178, pp.250–264, 2018.

[3] R. Fazai, K. Abodayeh, M. Mansouri, M. Trabelsi, H. Nounou, M. Nounou, and G. E. Georghiou, "Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems,"Solar Energy, vol. 190, pp. 405–413, 2019.

[4] M. Hajji, M.-F. Harkat, A. Kouadri, K. Abodayeh, M. Mansouri, H. Nounou, and M. Nounou, "Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems,"European Journal of Control, vol. 59, pp. 313–321, 2021.

[5] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. M. Pavan, "A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks,"Renewable Energy, vol. 90, pp. 501–512, 2016.

[6] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in 2014 Science and Information Conference, 2014, pp. 372–378.

[7] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments,"Philosophical Transactions of the Royal Society A:Mathematical, Physical and Engineering Sciences, vol. 374, no. 2065, p.20150202, 2016.

[8] M. Ringner, "What is principal component analysis?" Nature biotechnology, vol. 26, no. 3, pp. 303–304, 2008.