SAND2020-5720C

An evaluation of the value of **principal component analysis** for photovoltaic IV trace classification of physically-induced failures

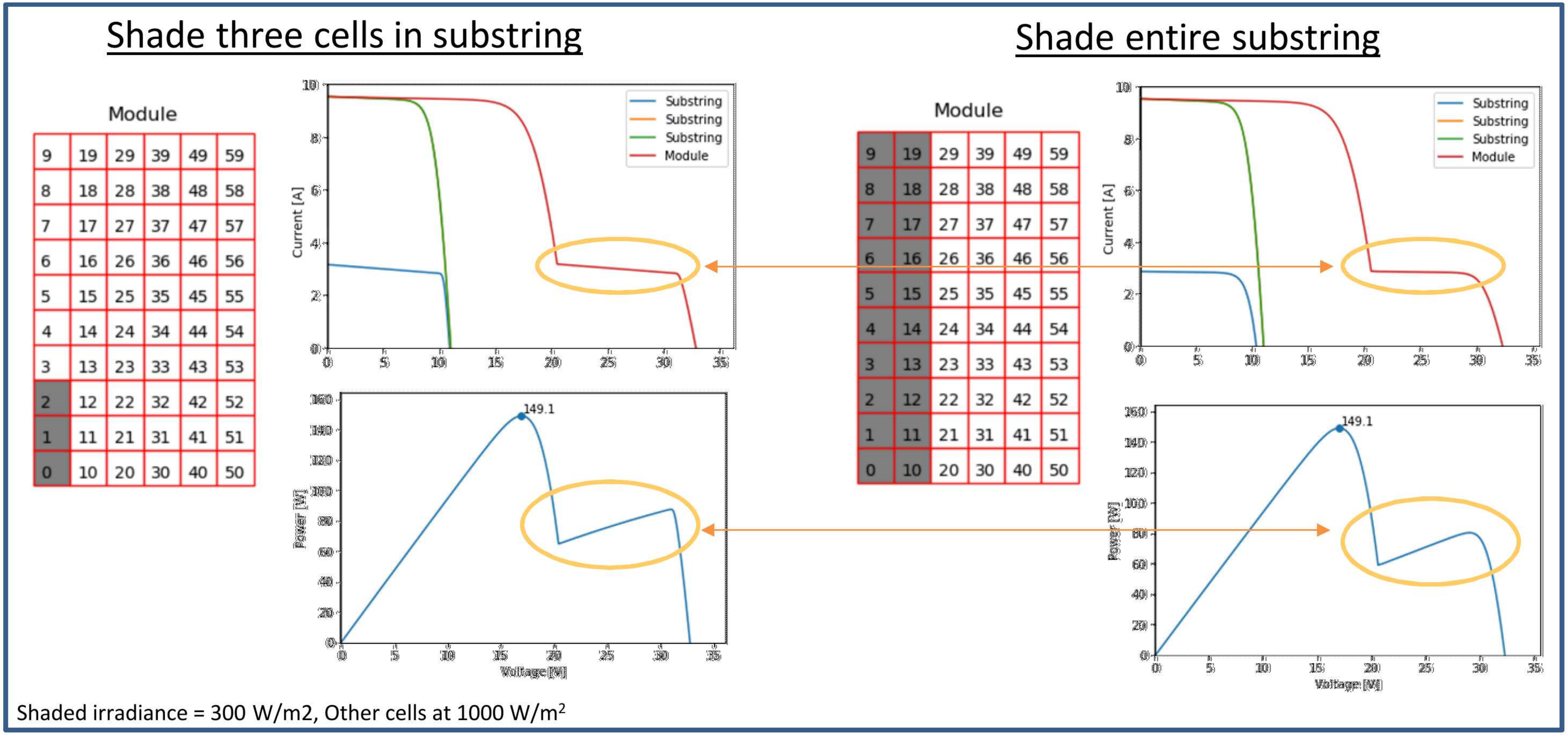
M.Hopwood, T.Gunda, H.Seigneur, J.Walters
University of Central Florida
Florida Solar Energy Center (FSEC)
Sandia National Laboratories

IEEE v.PVSC, June-August 2020

UCF

Background and Motivation

- **Improved diagnostics** for PV failures are ***critical*** **for ensuring reliability**
- IV traces are a common technique used to evaluate string or module performance
- IV traces have been classified by feature extraction (I_{sc},V_{oc},FF, R_{SH}, R_s, etc.), **but some failure characteristics *may* be missed.** For example,



Plots courtesy of Josh Stein, Sandia National Labs

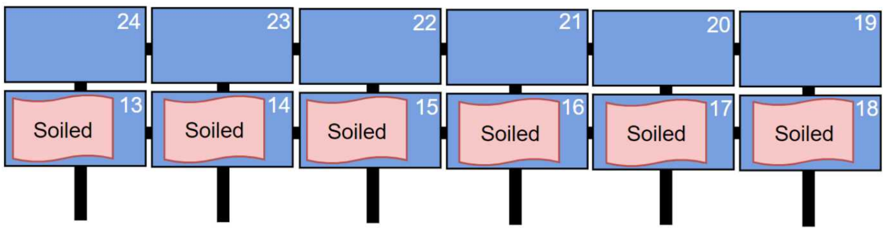
- **Principal component analysis** improves feature variance, and has shown success in IV classification [1]

UCF

Physical implementation of failure modes

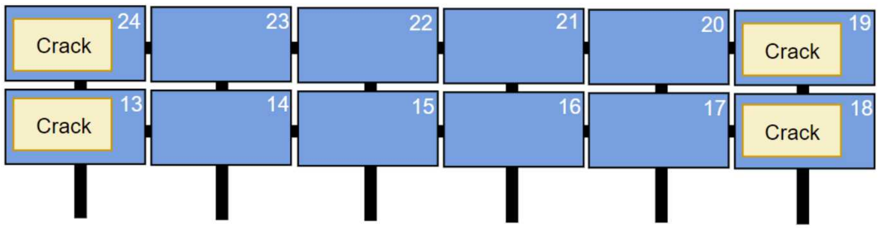
- Located at Florida Solar Energy Center (FSEC) in Cocoa, Florida
- A control string and a test string are implemented with 12 modules each
- The test string has three modes: *unstressed*, *partial soiling*, and *cell cracking*

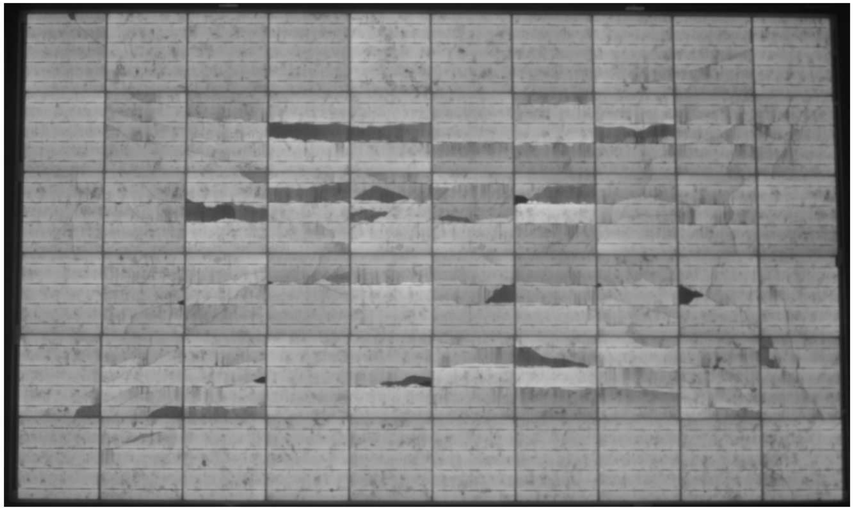
Partial Soiling



- Semi-transparent polymer film was laid on top of the bottom six modules [2]

Cell Cracking

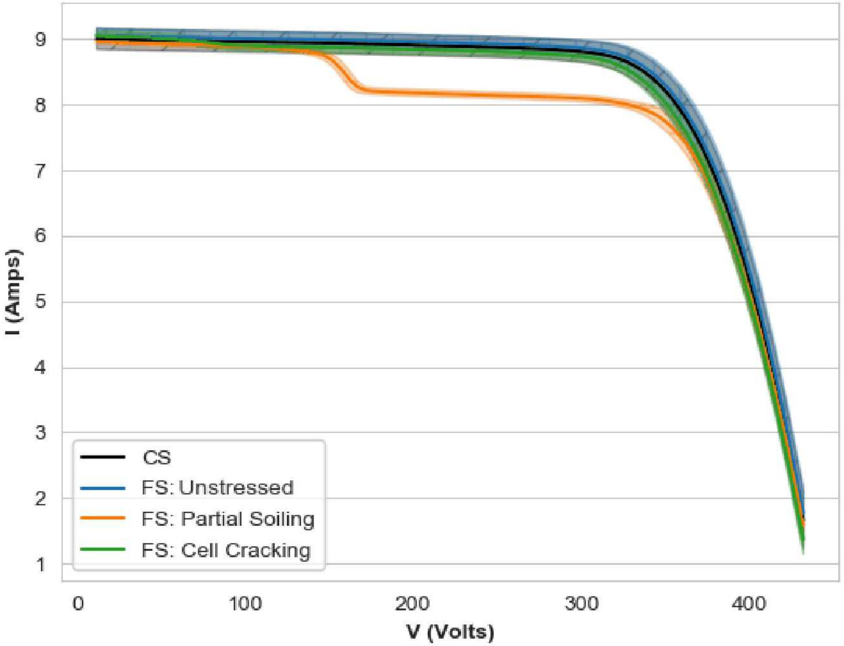




- Modules underwent a sequence of increasingly damaging thermomechanical loads [3,4]

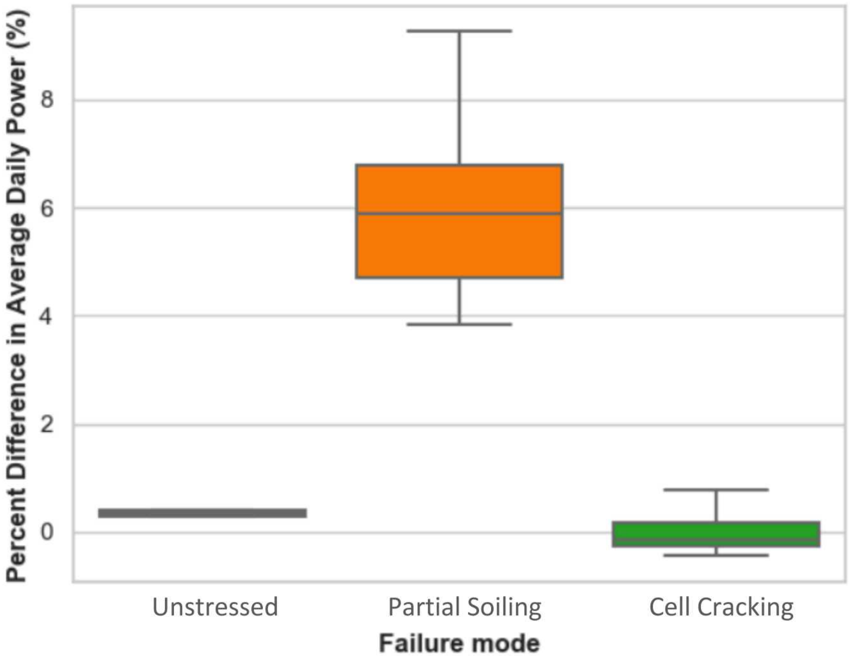
Methodology: First looks at the data

Average IV curve per mode



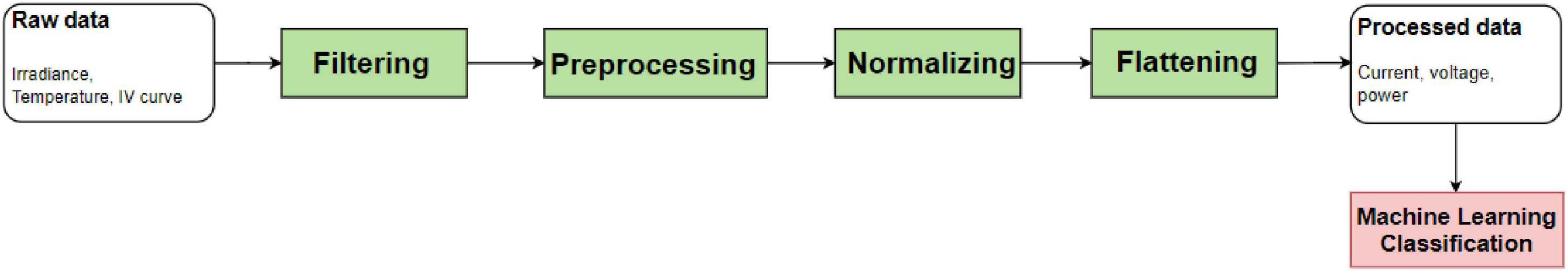
Average IV curve profiles for control string (CS), and three modes in the faulted string (FS) shows identifiable trends in each failure mode. A standard deviation region is included on all samples.

Percent difference in average daily power per mode

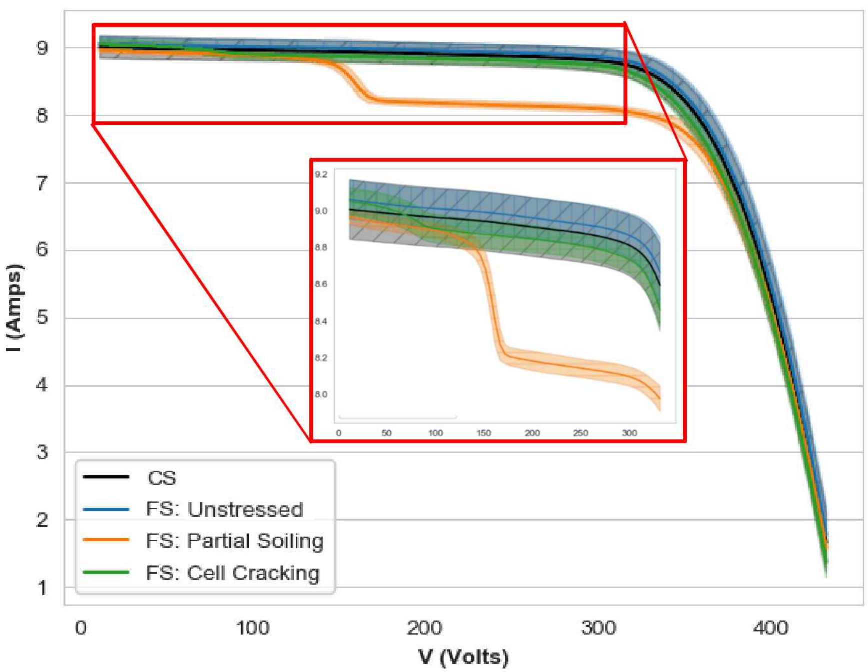


Max power point tracking (MPPT) data shows large power loss in partial soiling failure but relatively small, sometimes undetectable, loss in cell cracking failure

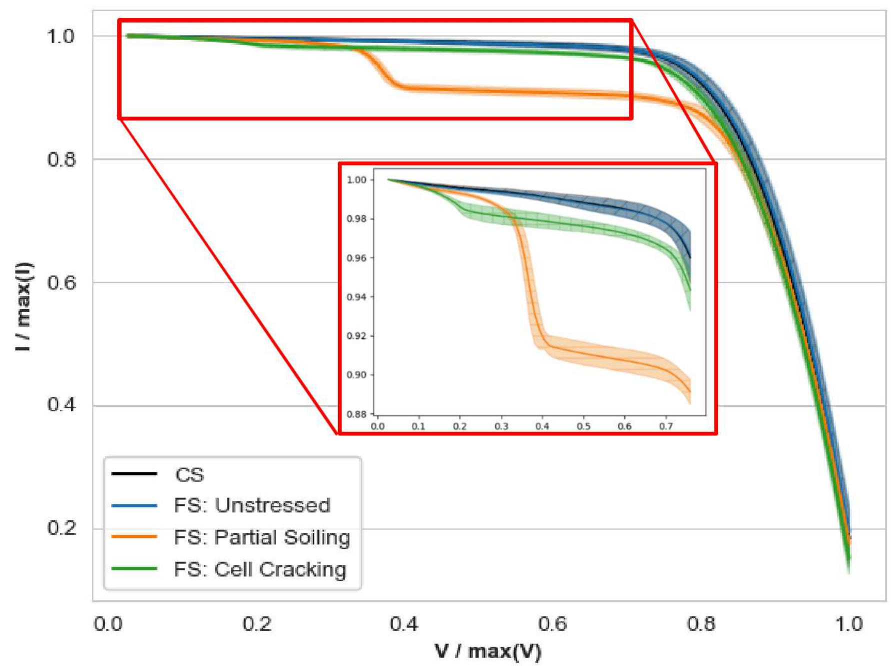
Methodology: Data filtering and processing



Preprocessed



Preprocessed & Normalized



Methodology: Data flattening

Raw data (Irradiance, Temperature, IV curve) → **Filtering** → **Preprocessing** → **Normalizing** → **Flattening** → **Processed data** (Current, voltage, power)

Flattening

"Trace-level"

Datetime	IV Voltage	IV Current	IV Power	y
3/1/2020 08:00:00	$[V_{0,0}, V_{0,1}, \dots, V_{0,m}]$	$[I_{0,0}, I_{0,1}, \dots, I_{0,m}]$	$[P_{0,0}, P_{0,1}, \dots, P_{0,m}]$	0
3/1/2020 08:30:00	$[V_{1,0}, V_{1,1}, \dots, V_{1,m}]$	$[I_{1,0}, I_{1,1}, \dots, I_{1,m}]$	$[P_{1,0}, P_{1,1}, \dots, P_{1,m}]$	0
...	$[V_{n,0}, V_{n,1}, \dots, V_{n,m}]$	$[I_{n,0}, I_{n,1}, \dots, I_{n,m}]$	$[P_{n,0}, P_{n,1}, \dots, P_{n,m}]$	k

→

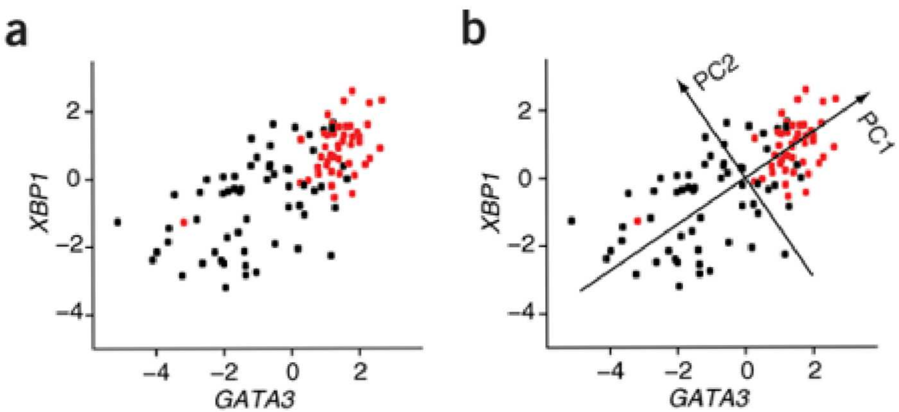
"Point-level"

IV Voltage	IV Current	IV Power	y
$V_{0,0}$	$I_{0,0}$	$P_{0,0}$	0
$V_{0,1}$	$I_{0,1}$	$P_{0,1}$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$V_{0,m}$	$I_{0,m}$	$P_{0,m}$	0
$V_{1,0}$	$I_{1,0}$	$P_{1,0}$	0
$V_{1,1}$	$I_{1,1}$	$P_{1,1}$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$V_{1,m}$	$I_{1,m}$	$P_{1,m}$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$V_{n,m}$	$I_{n,m}$	$P_{n,m}$	k

Per Fadhel et al [1]

Brief overview of ML techniques

Principal Component Analysis (PCA)



Principal components are evaluated as linear combinations of the input variables, constituting new axis in the input feature space. Figure from [5]

Flattened data

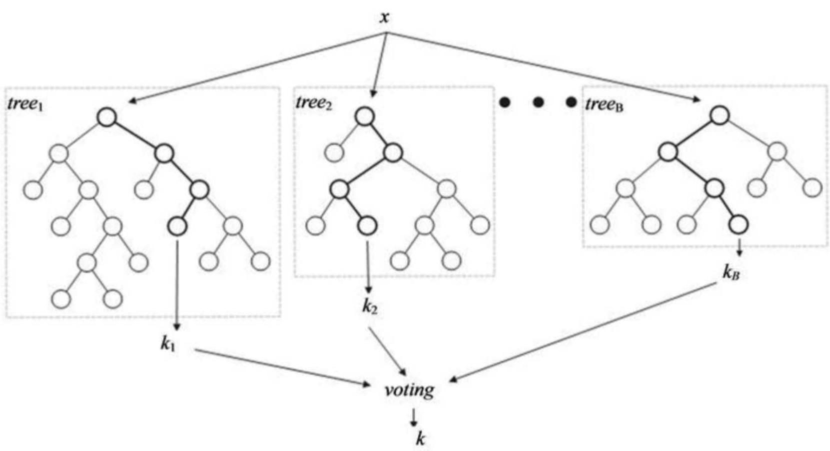
IV Voltage	IV Current	IV Power
$V_{0,0}$	$I_{0,0}$	$P_{0,0}$
$V_{0,1}$	$I_{0,1}$	$P_{0,1}$
$\vdots$	$\vdots$	$\vdots$
$V_{0,m}$	$I_{0,m}$	$P_{0,m}$
$V_{1,0}$	$I_{1,0}$	$P_{1,0}$
$V_{1,1}$	$I_{1,1}$	$P_{1,1}$
$\vdots$	$\vdots$	$\vdots$
$V_{1,m}$	$I_{1,m}$	$P_{1,m}$
$\vdots$	$\vdots$	$\vdots$
$V_{n,m}$	$I_{n,m}$	$P_{n,m}$

$\xrightarrow{\text{PCA}}$

Transformed data

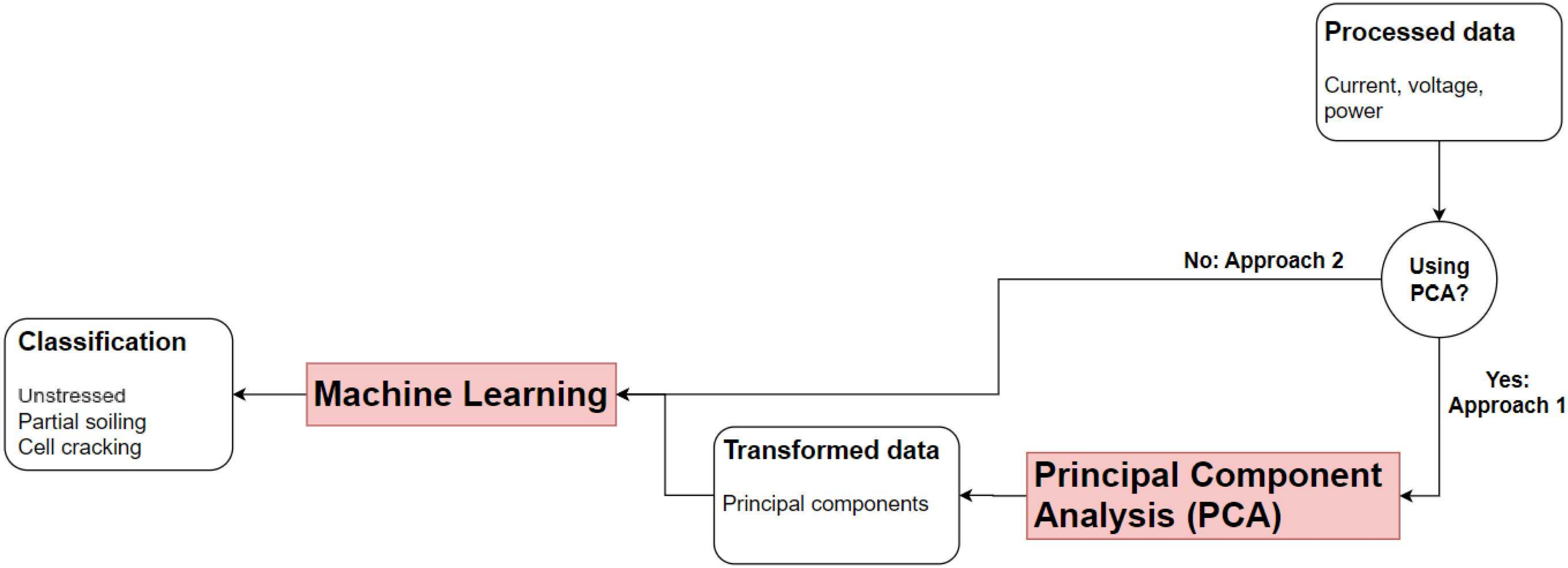
PC1	PC2
$PC1_1$	$PC2_1$
$PC1_2$	$PC2_2$
$\vdots$	$\vdots$
$PC1_{n*m}$	$PC2_{n*m}$

Random Forest (RF)



Random Forest (RF) is an *ensemble of decision trees*. Figure from [6]

Methodology: Feature reduction and classification

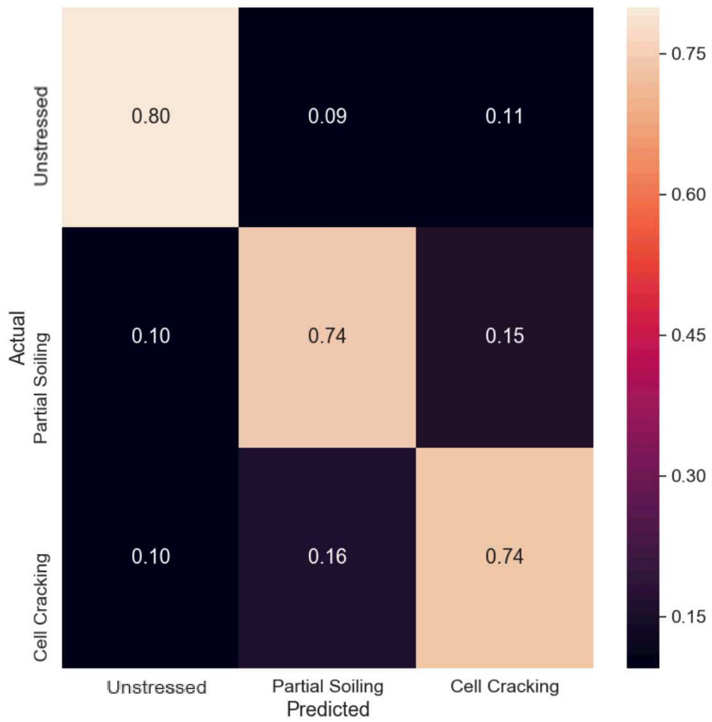


Two approaches are studied:

1. With PCA: Conduct PCA on input features, push principal components into machine learning model
2. Without PCA: Push input features into machine learning model

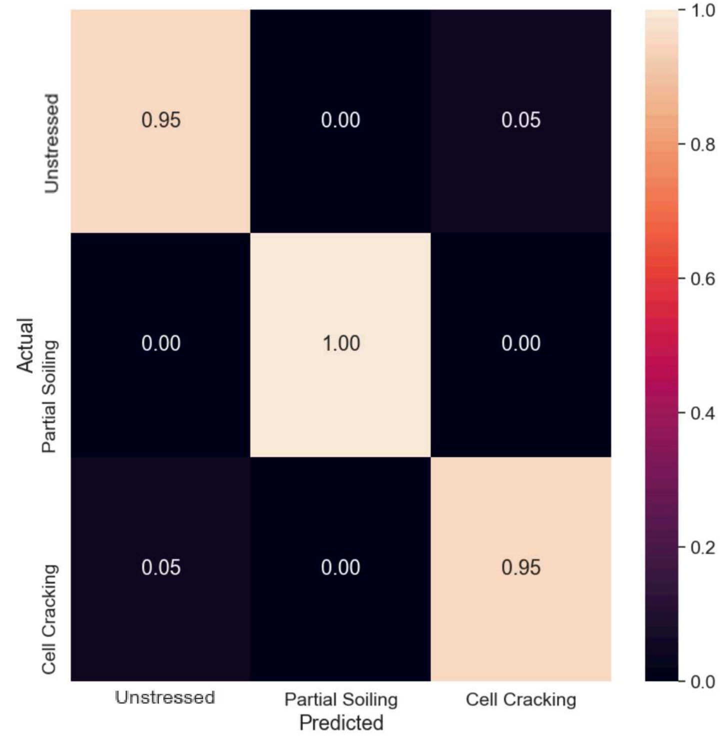
Results: Model accuracy can be evaluated at two levels

Points-level confusion matrix



Extrapolated to

Trace-level confusion matrix



Each point in the IV curve is assigned a classification.

The points are collected back into their respected IV trace. The classification which occurs the most is designated as the trace-level classification.

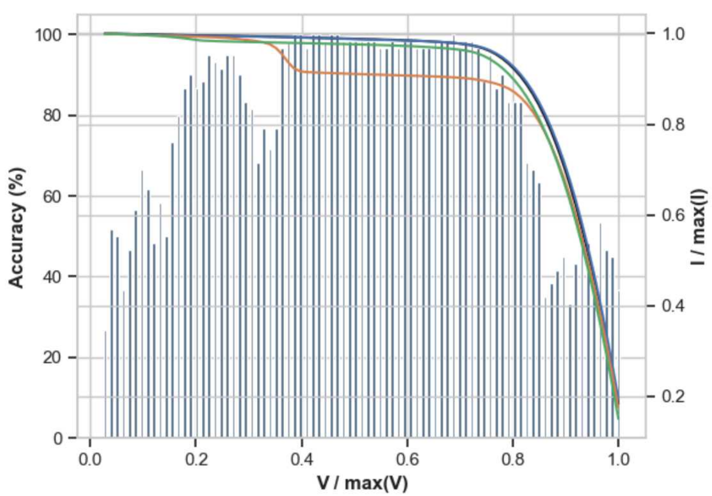
Results: Accuracy evaluations

Studying the model accuracy over 20 repetition to quantify variance across iterations

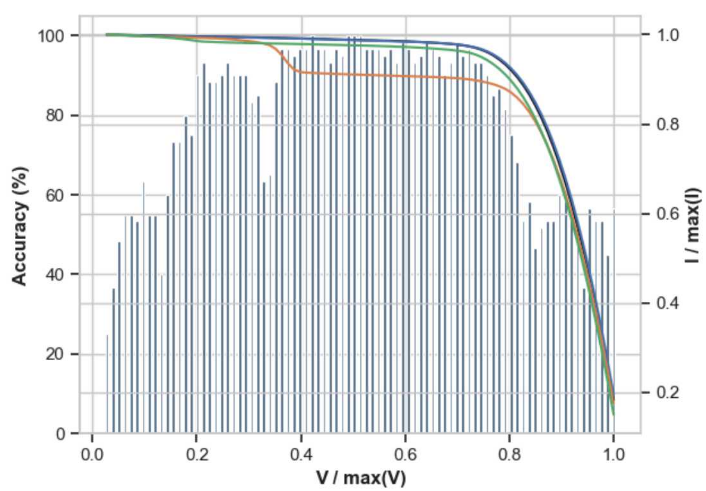
	Average accuracy, % (SD)	
	Point-level	Trace-level
With PCA	78.8 (1.22)	98.8 (1.28)
Without PCA	79.3 (1.80)	98.0 (1.63)

Studying the distribution of accuracy along a trace in one test

With PCA

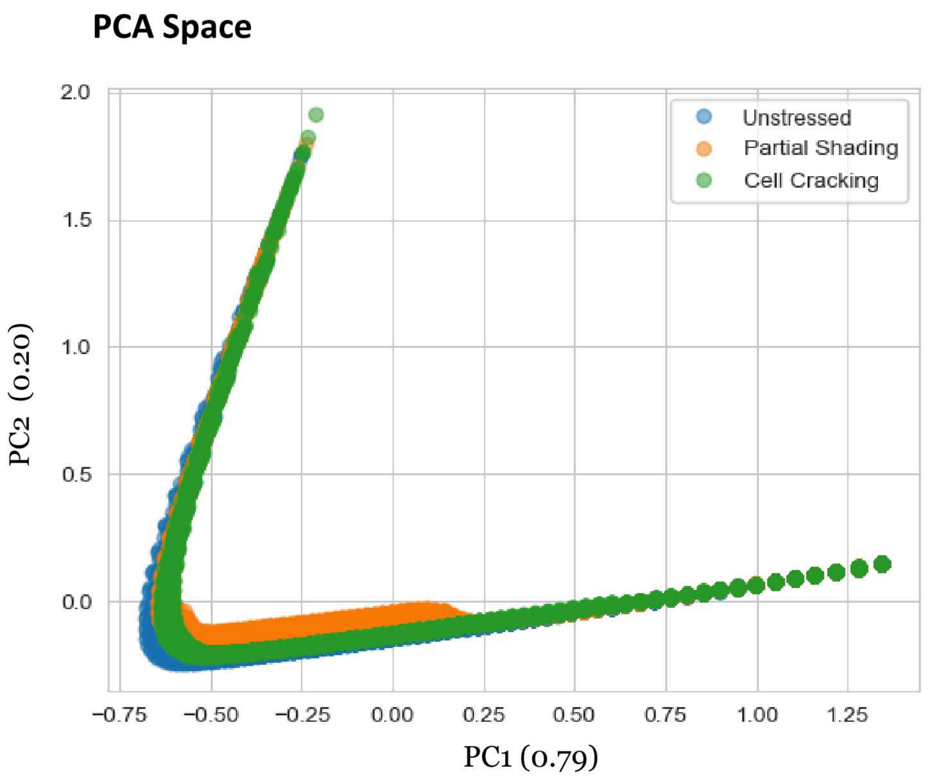


Without PCA



1. **Higher accuracies** located where failure modes *visually differentiate*
2. Similar accuracy profiles are seen on both With/Without PCA

Results: PCA has minimal effect on the feature space

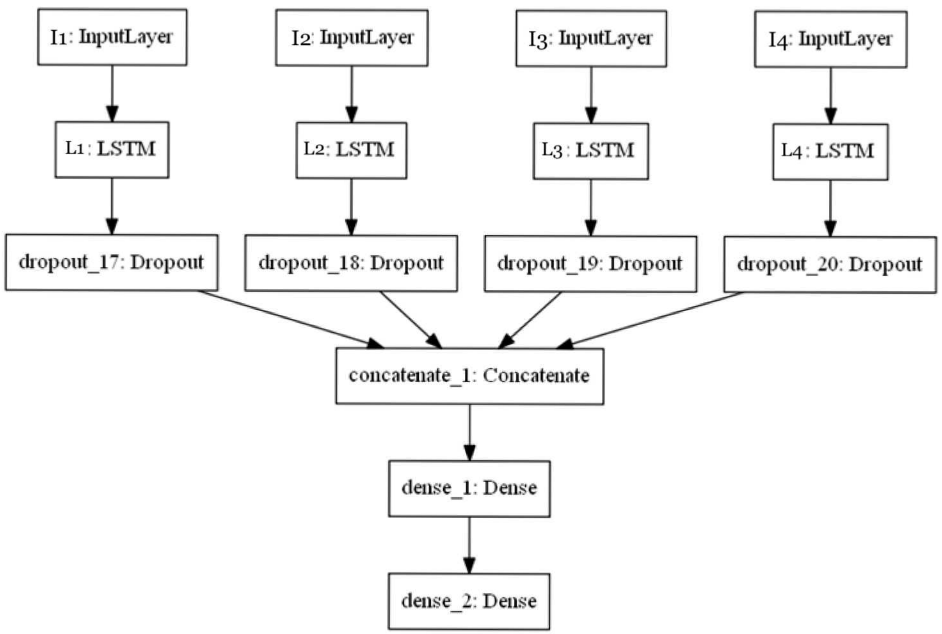


Conclusion

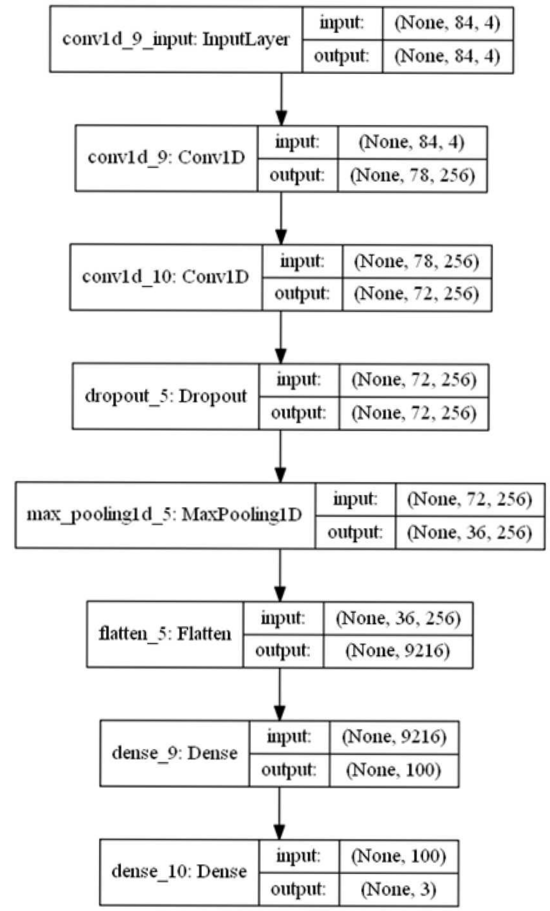
- In our case, PCA gives only marginal improvement in accuracy (+0.8%, on average)
- High accuracies (>98%) are found *even though* we incorporate failure modes which minorly affect the IV curve
- Preprocessing steps are essential towards differentiating our failure modes
- Model deployment is running successfully with similar accuracies

Future work: IV pattern recognition with neural networks

Multi-headed LSTM Architecture



1D CNN Architecture



Why neural networks?

- Could scale well with more failure modes which have less variability

UCF

References

[1] Fadhel, S., et al. "PV shading fault detection and classification based on IV curve using principal component analysis: Application to isolated PV system." *Solar Energy* 179 (2019): 1-10.

[2] Walters, Joseph, H. Seigneur, E. Schneller, M. Matam and M. Hopwood, "Experimental Methods to Replicate Power Loss of PV Modules in the Field for the Purpose of Fault Detection Algorithm Development," *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, Chicago, IL, USA, 2019, pp. 1410-1413, doi: 10.1109/PVSC40753.2019.8980896.

[3] Rowell, Michael W., et al. "The effect of laminate construction and temperature cycling on the fracture strength and performance of encapsulated solar cells." *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)(A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)*. IEEE, 2018.

[4] Schneller, Eric J., et al. "The Impact of Cold Temperature Exposure in Mechanical Durability Testing of PV Modules." *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*. IEEE, 2019.

[5] Ringnér, Markus. "What is principal component analysis?." *Nature biotechnology* 26.3 (2008): 303-304.

[6] Nguyen, Cuong, Yong Wang, and Ha Nam Nguyen. "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic." (2013).

UCF

Collaboration – synergy through teamwork

UNIVERSITY OF CENTRAL FLORIDA

Joe Walters, Hubert Seigneur,
Manjunath Matam, Michael Hopwood

PORDIS

Ryan Smith

Sandia National Laboratories

Josh Stein, C Birk Jones
Thushara Gunda

NEXTera ENERGY RESOURCES

Jon Pantano

OSIsoft

Mike Mihuc

SOLAR ENERGY TECHNOLOGIES OFFICE
U.S. Department Of Energy

Inna Kozinsky, Tassos Golnas

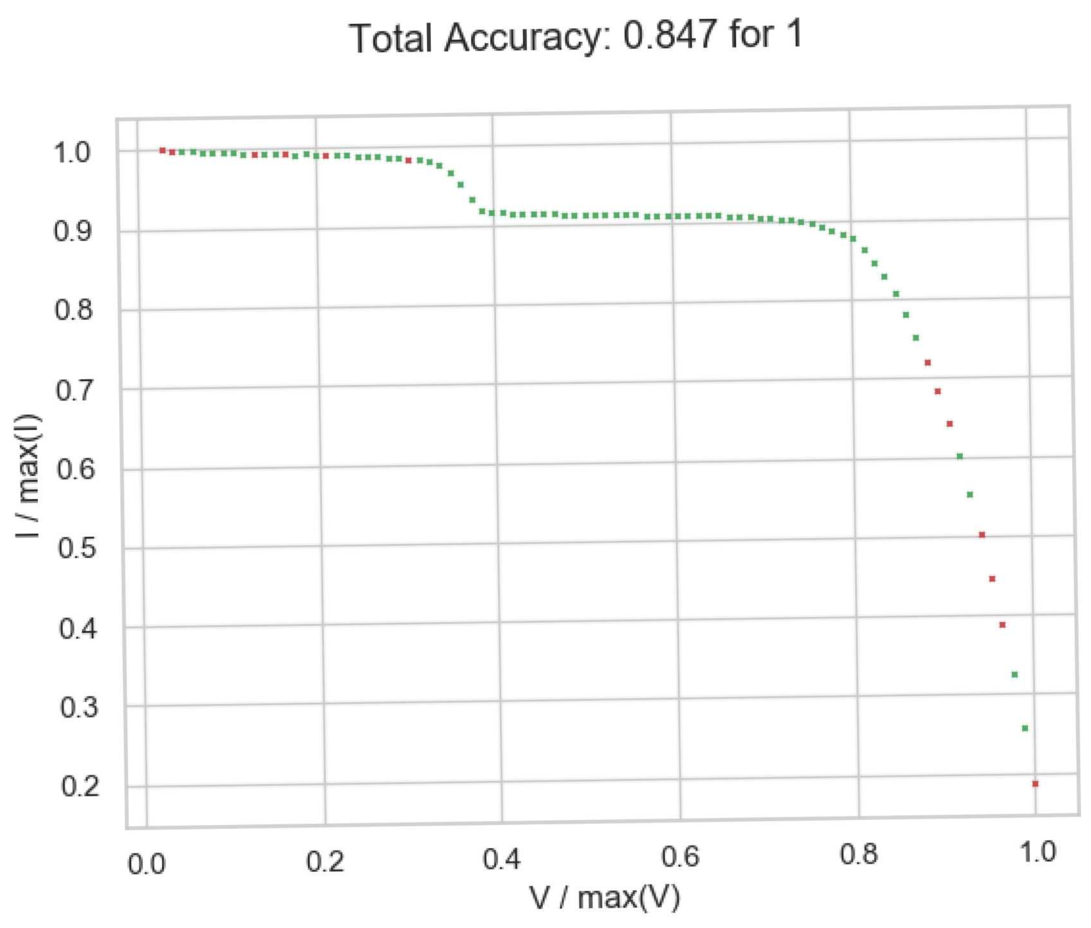
UCF

Thank you!

✉ michael.hopwood@knights.ucf.edu

UCF

APPENDIX


Total Accuracy: 0.847 for 1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
I / max(I)
0.0
0.2
0.4
0.6
0.8
1.0
V / max(V)

With PCA

Feature	Importance
PC1	0.499
PC2	0.501

Without PCA

Feature	Importance
Current	0.538
Power	0.320
Voltage	0.142