Surface Albedo Prediction using Artificial Neural Networks

Niki Kyriacou, Sameeksha Katoch, Andreas Spanias, Yiannis Tofis

***Abstract*—Surface albedo describes the fraction of sunlight reflected by a surface using a value from zero to one. Surface albedo is a key piece of information for algorithms used to maximize the performance of solar arrays through topology reconfiguration. Since surface albedo can vary widely due to environmental conditions, predicting surface albedo is incredibly useful. This paper explores the leveraging of artificial neural networks to predict surface albedo.**

Index terms: surface albedo, photovoltaic systems, artificial neural networks, machine learning, irradiance, shading prediction

I. Introduction

As the harmful effects of using non-renewable energy sources become more apparent, the shift towards renewable energy sources has rapidly increased. The most popular sources of renewable energy include biomass, geothermal, hydro, solar, and wind. When comparing the global potential for these renewable energy sources, 90% comes from solar, 9% comes from wind, and only 1% comes from other renewables [1]. Since solar comprises such a large percentage of the potential global renewable energy, the development of efficient PV systems is critical.

One difficulty with developing efficient PV systems is the fluctuation of solar energy produced as environmental factors change. One of the factors affecting solar power production is surface albedo. Surface albedo describes the fraction of sunlight reflected by a surface measured from zero to one, where zero describes a surface that absorbs all sunlight and one describing a surface that reflects all sunlight [2].

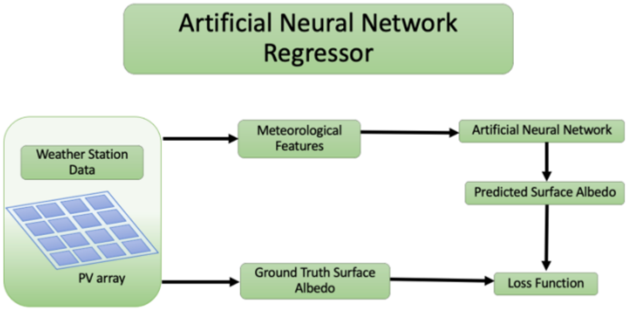


Figure 1: Block diagram for surface albedo prediction using artificial neural network regressor.

Surface albedo can change based on various environmental factors. Surface albedo tends to be higher during the early morning and late afternoon and dip during the midday. Different seasons also show different surface albedo measurements, with the highest values occurring during the winter months and the lowest during the summer months. Climate also affects surface albedo, with dry regions showing an average surface albedo of 0.5 compared to a surface albedo of 0.2 in more humid areas. Furthermore, the presence of snow could increase surface albedo to approximately 0.9 [3].

Surface albedo is related to irradiance, which describes the rate that solar energy hits a surface in units of Watts per square meter. Irradiance data has previously been used in machine learning algorithms to reconfigure the topology of solar arrays to maximize performance [4]. Since surface albedo is a function of irradiance, predicting surface albedo could further improve these topology reconfiguration models.

SenSIP lab has previously addressed several problems in solar array monitoring, control, and optimization [5-15]. Initial work was reported in 2012 where traditional statistical methods were proposed [6]. Machine learning methods were later considered, including PU learning [7][8]. In 2019, fault detection using neural nets and optimization methods were reported [9-11]. Most recently, SenSIP lab has reported PU learning for fault detection and published a study including neural net fault detection experiments and simulations on a quantum computer simulator [12][13].

This research proposes training artificial neural networks using one-year-long data from the NSRDB database, collected at 30-minute time resolution. The data includes nine major features which are utilized to predict surface albedo. The use of artificial neural networks to predict surface albedo should allow models that reconfigure the topology of solar arrays to maximize efficiency throughout the year and reduce the power output fluctuations in the PV systems.

II. Data

The data used in this research was obtained from the National Solar Radiation Database (NSRDB). This project used the data from the entire year of 2015, collected at intervals of 30 minutes. Each datapoint can be identified by latitude, longitude, time zone, elevation, month, day, hour, and minute. The specified data includes 12 major features which are used to predict surface albedo. These features include Diffuse Horizontal Irradiance (DHI), Direct Normal Irradiance (DNI), Global Horizontal Irradiance (GHI), Cloud Type, Dew Point, Solar Zenith Angle, Wind Speed, Precipitable Water, Wind Direction, Relative Humidity, Temperature, and Pressure.

Before using the data to train the artificial neural network, the data underwent pre-processing. First, one-hot encoding was performed on the Cloud Type portion of the data since Cloud Type was a categorical variable followed by dropping the Fill Flag column. Next, sklearn's StandardScaler was used to standardize the dataset. Finally, the data was divided into an 80/20 train/test split.

III. Artificial Neural Network Regressor

This project trained SKLearn's Multi-layer Perceptron Regressor (MLPRegressor) to predict surface albedo. A multi-layer perceptron is a type of artificial neural network. This kind of network will typically have an input layer, an output layer, and one or more hidden layers. We use MLP Regression to predict a surface albedo value.

An overview of artificial neural network is shown in Figure 2. In this work, we used sklearn's MLPRegressor model which has several hyperparameters that can be tuned to optimize the output predictions. The model allows alteration of the number of hidden layers, the number of nodes in each layer, the number of iterations used, learning rate, batch size, solver, activation function, and several other features [6].

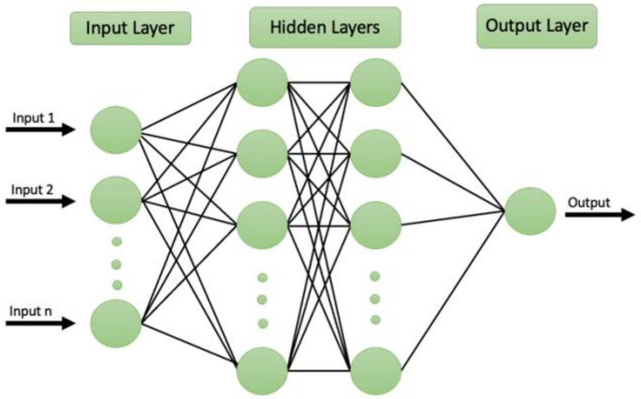


Figure 2: Depiction of an artificial neural network with two hidden layers.

IV. Results

The project began with analysis of the data from the NSRDB. To get an idea of the values being used, the DHI, GHI, and surface albedo for the year were plotted.

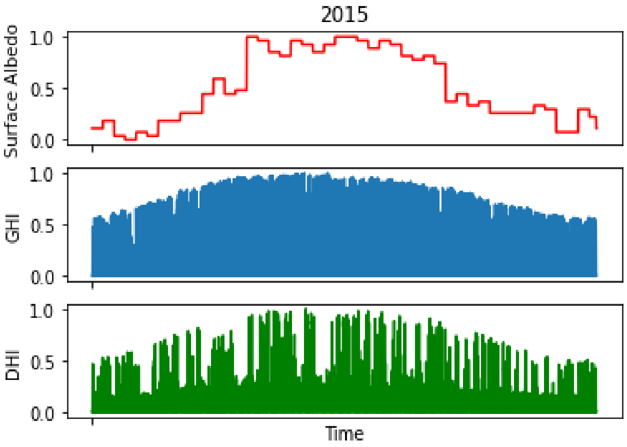


Figure 3: Graph depicting average DHI, GHI, and Surface Albedo for the year 2015.

A. Hyperparameter Tuning

Next, the first step in utilizing the artificial neural network was to decide how many hidden layers, nodes, and iterations would be used. It was decided that root mean square error (RMSE) would be used to as a metric to calculate the distance between ground truth and predicted surface albedo. In order to determine the best parameters for this artificial neural network, a number of simulations were run with various combinations of hidden layers, nodes, and iterations. Some of the first simulations run used one, two, and three hidden layers and tested the RMSE with nodes 1-100. These simulations were performed using 1000 iterations and the default parameters for all other settings of MLPRegressor.

Based on the relatively small size of the data set, it was decided that ten node layers would be used for this experiment.

Once the number of nodes was decided, simulations were run with one, two, and three hidden layers of ten nodes each over iterations 1-1000.

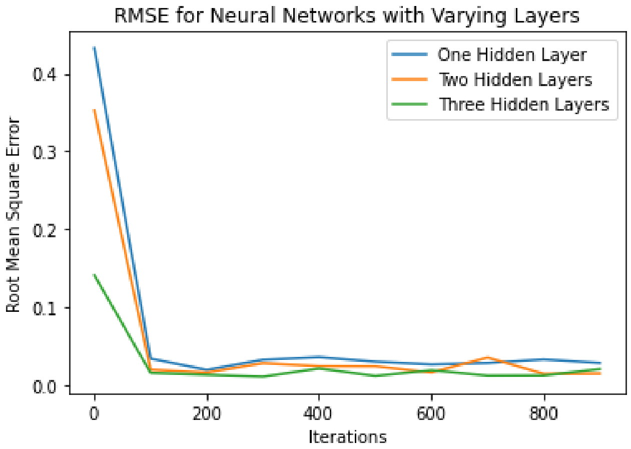


Figure 4: Graph depicting RMSE for artificial neural network using one, two, and three layers with 10 nodes each and iterations 1-1000.

Figure 4 depicts the results of the simulations over 1-1000 iterations. As seen in the graph, using two hidden layers at 1000 iterations produced the lowest RMSE. From all the simulations discussed thus far, it was decided to use two hidden layers with ten nodes each at 1000 iterations.

The next hyperparameter explored was the learning rate. Simulations were run with three different learning rates of 0.001, 0.0001, and 0.0001.

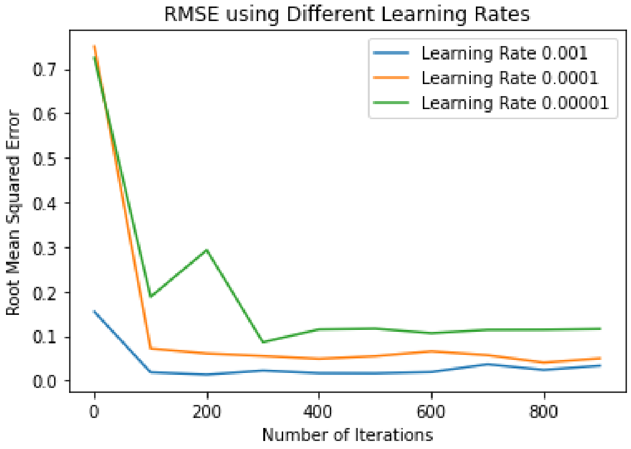


Figure 5: Graph depicting RMSE for artificial neural network using two hidden layers with 10 nodes each, 1000 iterations, and learning rates 0.001, 0.0001, and 0.00001.

Figure 5 depicts the results of the simulations with the three different learning rates over 1000 iterations. As seen in the graph, MLPRegressor's default learning rate of 0.001 was found to produce the lowest RMSE.

The next hyperparameter explored was batch size. Eight different simulations were run using two hidden layers with ten nodes each, the default learning rate, 1000 iterations, and batch sizes from eight to 64.

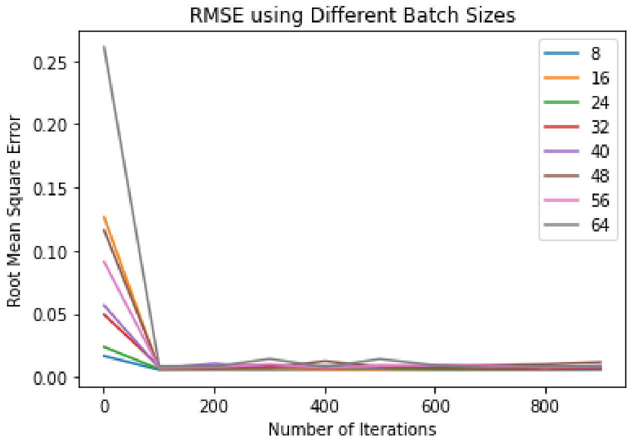


Figure 6: Graph depicting RMSE for artificial neural network using two hidden layers with 10 nodes each and batch sizes 8-64 over 1000 iterations.

Figure 6 depicts the results of the simulations run using eight different batch sizes. Based on these results, it was decided that a batch size of 64 would be used.

The next hyperparameter explored was the solver. The two solvers tested were the default solver Adam and Stochastic Gradient Descent (SGD).

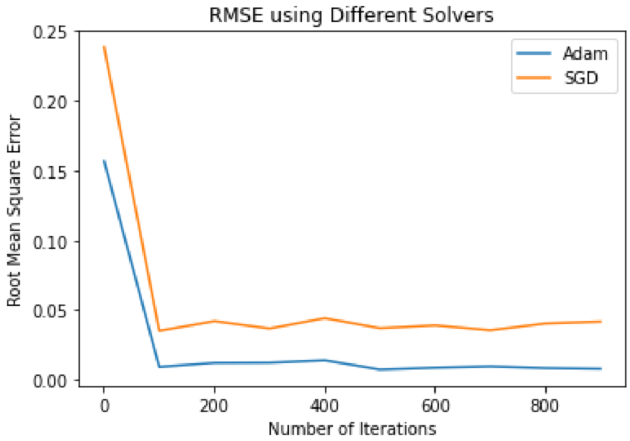


Figure 7: Graph depicting RMSE for artificial neural network using two hidden layers with 10 nodes each, batch size 64, and solvers Adam or SGD over 1000 iterations.

As seen in figure 7, the solver Adam resulted in a lower RMSE than SGD. Therefore, it was decided that Adam would be used going forward [7] [8].

The final hyperparameter tested was the activation function. Three different activation functions were tested. These activation functions were the rectified linear unit function (relu), logistic sigmoid function (logistic), and the hyperbolic tangent function (tanh).

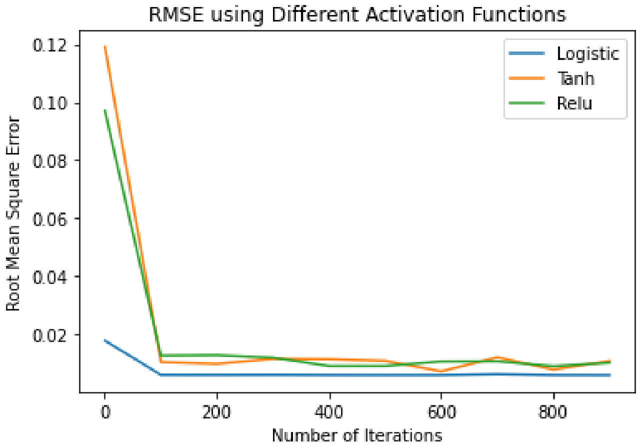


Figure 8: Graph depicting RMSE for artificial neural network using two hidden layers with 10 nodes each, batch size 64, and activation functions relu, tanh, or logistic over 1000 iterations.

The results of the simulations using the different activations functions are depicted in figure 8. Based on the results of these simulations, it was decided that the default activation function relu would be used going forward.

This concluded the hyperparameter tuning portion of the project.

B. *Feature Ranking*

Next, one feature at a time was removed from the dataset and the change in prediction performance was measured to rank which features correlated most strongly to surface albedo. This simulation was done ten times and the averages were computed.

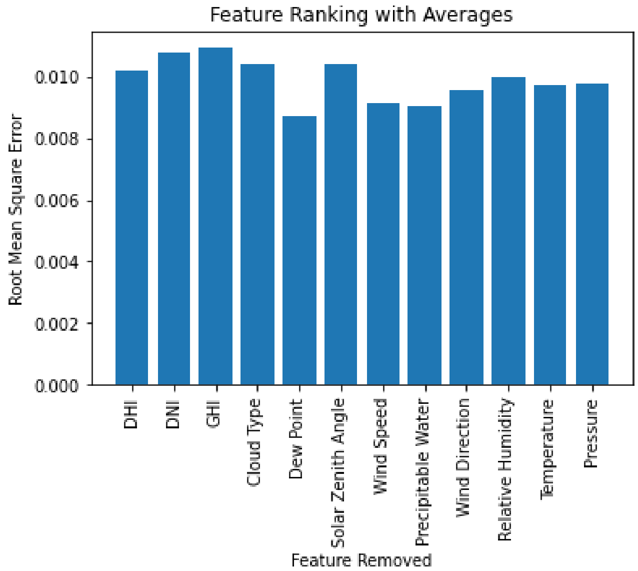
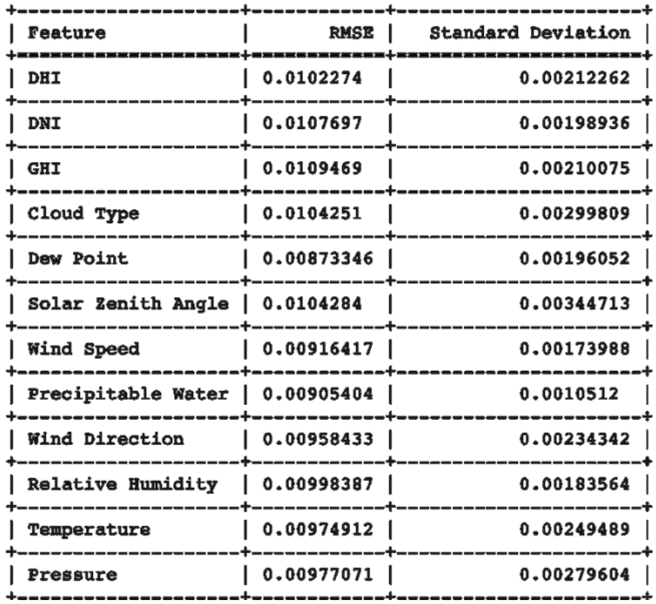


Figure 9: Graph depicting average RMSE when the identified feature was removed.

Based on the results of averaging the features, the four features that demonstrated the highest RMSE when removed were GHI, DNI, Solar Zenith Angle, and Cloud Type. This would imply that these features are most strongly correlated to surface albedo prediction.



```
+--------------------+-----------+----------------------+
| Feature            |      RMSE |   Standard Deviation |
+====================+===========+======================+
| DHI                | 0.0102274 |           0.00212262 |
+--------------------+-----------+----------------------+
| DNI                | 0.0107697 |           0.00198936 |
+--------------------+-----------+----------------------+
| GHI                | 0.0109469 |           0.00210075 |
+--------------------+-----------+----------------------+
| Cloud Type         | 0.0104251 |           0.00299809 |
+--------------------+-----------+----------------------+
| Dew Point          | 0.00873346 |          0.00196052 |
+--------------------+-----------+----------------------+
| Solar Zenith Angle | 0.0104284 |           0.00344713 |
+--------------------+-----------+----------------------+
| Wind Speed         | 0.00916417 |          0.00173988 |
+--------------------+-----------+----------------------+
| Precipitable Water | 0.00905404 |          0.0010512  |
+--------------------+-----------+----------------------+
| Wind Direction     | 0.00958433 |          0.00234342 |
+--------------------+-----------+----------------------+
| Relative Humidity  | 0.00998387 |          0.00183564 |
+--------------------+-----------+----------------------+
| Temperature        | 0.00974912 |          0.00249489 |
+--------------------+-----------+----------------------+
| Pressure           | 0.00977071 |          0.00279604 |
+--------------------+-----------+----------------------+
```

Figure 10: Table depicting average RMSE and standard deviation when the identified feature was removed.

To gain more insights about the feature space, it was determined that principal component analysis should be done on the data. The first step in this process was to determine the variance of the dataset.

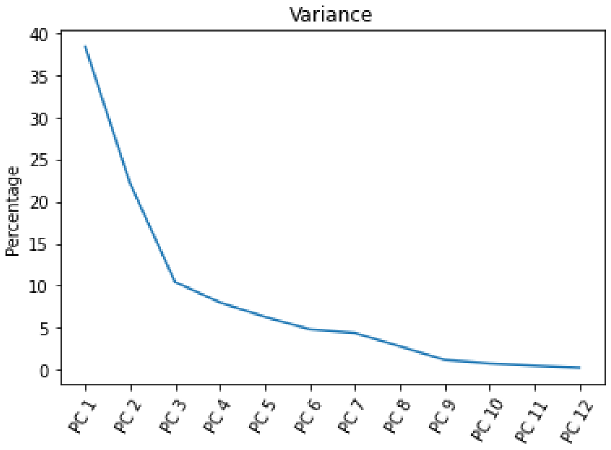


Figure 11: Graph depicting the variance of the dataset.

As seen in figure 11, the first three components contribute to 71% of the total variance. After the first three components, the change in variance diminishes significantly.

Next, simulations were run to measure the prediction performance using different numbers of principal components, from one to seven.

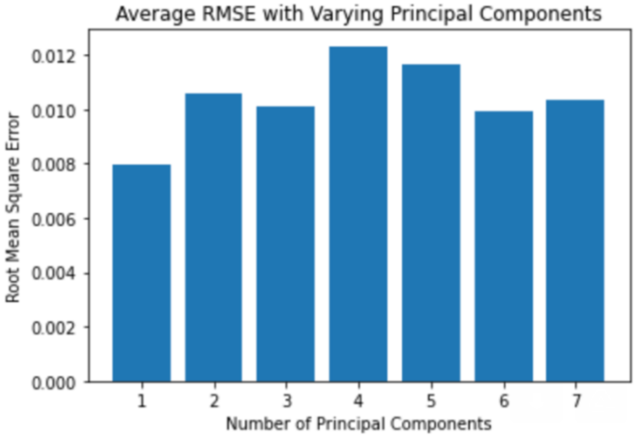


Figure 12: Table depicting average RMSE with different numbers of principal components.

As can be seen in figure 12, the lowest RMSE was obtained using only one principal component. One reason for this could be that the first principal component accounts for almost 40% of the variance.

V. CONCLUSION AND FUTURE WORK

In conclusion, the variance distribution indicates that the there are three principal components responsible for most of the variance in the dataset. The three features that correlated to surface albedo prediction most strongly were GHI, DNI, and Solar Zenith Angle. For future work, it could be beneficial to use an auto-encoder for dimensionality reduction and compare those results to the results obtained in this principal component analysis [9]. That could potentially give us more information about the feature space and how the features are related to each other.

Acknowledgment

This project was sponsored by National Science Foundation Award 1854273.

References

[1] W. Hoffmann, *The economic competitiveness of renewable energy : pathways to 100% global coverage*. Salem, Massachusetts ; Hoboken, New Jersey: Scrivener Publishing : John Wiley & Sons, 2014, p. 47.

[2] Kotak, Y., Gul, M., Muneer, T. and Ivanova, S., 2015. Investigating the Impact of Ground Albedo on the Performance of PV Systems.

[3] B. Marion, “Albedo Data Sets for Bifacial PV Systems,” *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*, 2020.

[4] M. A. Gacusan and V. Muthukumar, "Cloud Motion Vector Estimation Using Scalable Wireless Sensor Networks," 2018 31st IEEE International System-on-Chip Conference (SOCC), 2018, pp. 13-18, doi: 10.1109/SOCC.2018.8618507.

[5] 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.

[6] H. Braun, S. Buddha, V. Krishnan, C. Tepedelenlioglu, A. Spanias, S.i Takada, T. Takehara, M. Banavar, and T. Yeider., Signal Processing for Solar Array Monitoring, Fault Detection, and Optimization, Synthesis Lectures on Power Electronics, Morgan & Claypool, Book, 1-111 pages, ISBN 978-1608459483, Sep. 2012.

[7] U. Shanthamallu, A. Spanias, C. Tepedelenlioglu, M. Stanley, "A Brief Survey of Machine Learning Methods and their Sensor and IoT Applications," *Proceedings 8th International Conference on Information, Intelligence, Systems and Applications (IEEE IISA 2017)*, Larnaca, August 2017.

[8] K. Jaskie and A. Spanias, "Positive and Unlabeled Learning Algorithms and Applications: A Survey," *Proc. IEEE IISA 2019*, Patras, July 2019.

[9] S. Rao, A. Spanias, C. Tepedelenliglu, "Solar Array Fault Detection using Neural Networks", IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), Taipei, May 2019.

[10] S. Rao, G. Muniraju, C. Tepedelenlioglu, D. Srinivasan, G. Tamizhmani and A. Spanias, "Dropout and Pruned Neural Networks for Fault Classification in Photovoltaic Arrays, *IEEE Access*, 2021.

[11] V. Narayanaswamy, R. Ayyanar, A. Spanias, C. Tepedelenlioglu, "Connection Topology Optimization in PV Arrays using Neural Networks'," IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), Taipei, May 2019.

[12] K. Jaskie, J. Martin, and A. Spanias, “PV Fault Detection using Positive Unlabeled Learning,” Applied Sciences, vol. 11, Jun. 2021.

[13] G. Uehara, S. Rao, M. Dobson, C. Tepedelenlioglu and Andreas Spanias, "Quantum Neural Network Parameter Estimation for Photovoltaic Fault,” *Proc. IEEE IISA 2021*, July 2021.

[14] H. Braun, S. T. Buddha, V. Krishnan, C. Tepedelenlioglu, A. Spanias, M. Banavar, and D. Srinivansan, “Topology reconfiguration for optimization of photovoltaic array output,” *Elsevier Sustainable Energy, Grids and Networks* (*SEGAN*), pp. 58-69, Vol. 6, June 2016.

[15] M20-254P Dropout and Pruned Neural Networks for Fault Classification in Photovoltaic Arrays, G. Muniraju, S. Rao, A. Spanias, C. Tepedelenlioglu, Provisional US 63/039,012 , 06/15/2020.

[16] Scikit-learn.org. n.d. *sklearn.neural_network.MLPRegressor — scikit-learn 0.24.2 documentation*. [online] Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegrMLPRe.html>.

[17] Kingma, D. and Ba, J., 2015. *Adam: A Method for Stochastic Optimization*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1412.6980>.

[18] Bottou, L., 1991. *Stochastic Gradient Learning in Neural Networks*. [online] Leon.bottou.org. Available at: https://leon.bottou.org/publications/pdf/nimes-1991.pdf.

[19] Ladjal, S., Newson, A. and Pham, C., 2019. *A PCA-like Autoencoder*. [online] arXiv.org. Available at: https://arxiv.org/abs/1904.01277.