ASIA-PACIFIC
SOLAR RESEARCH CONFERENCE

Process Optimization of Commercial Solar Cell Manufacturing Using Machine Learning

Casper A. Eijkens[1], Yoann Buratti[1], Ziv Hameiri[1]

[1]*School of Photovoltaic and Renewable Energy Engineering, UNSW, Sydney, NSW 2052*
E-mail: c.eijkens@student.unsw.edu.au

Abstract: The manufacturing of photovoltaic devices is a combination of consecutive processes that turns silicon, chemicals, gasses and metals into a solar cell. Each process follows a recipe that needs to maximize the performance of the fabricated solar cells. The configuration of a process is a complex optimization problem, since many interactions between process variables exist. Typically, process optimization involves statistical modeling of experimental data using low order polynomials. However, polynomial functions struggle to fit many variables and non-linear patterns and can therefore only model a few processes. In this study, this problem is addressed by using machine learning (ML) algorithms, which can accurately model complex patterns in high-dimensional data. ML simultaneously learns multiple processes and the interactions among their variables that normally remain unexplored. The capacity to model manufacturing processes is tested for several ML algorithms using simulated data. An artificial neural network is found to model a complete production line with 49 input variables, predicting cell efficiencies with a mean absolute error of 0.29%. Furthermore, using an optimization algorithm, the neural network identified a new process configuration with a significantly better cell efficiency, demonstrating the advantage of ML models for PV manufacturing process optimization.

1. INTRODUCTION

Typically, process optimization involves polynomial fitting of design of experiment (DoE) data to establish a model between process configurations and solar cell attributes [1]. However, polynomials cannot model non-polynomial patterns and are neither time nor cost effective with an increased number of variables [1]. Given that PV manufacturing has several processes and many interacting variables, a polynomial-based method cannot extract all the information regarding the process and is restricted to modeling only a few processes [1].

The aim of this study is to explore machine learning (ML) applications in process optimization of PV manufacturing. ML seems more appropriate for modelling PV manufacturing processes, because it has the capacity to learn a wide variety of non-linear patterns in high-dimensional data [2]. ML models include any interconnections between processes and allow precise optimization of multiple processes. As a proof-of-concept, simulated manufacturing data is modeled with several ML algorithms (artificial neural network (ANN) [2], support vector regression (SVR) [2], random forests (RF) [2] and AdaBoost (AB) [2]) and their performance is then compared to a polynomial fitting (POLY).

II. BACKGROUND

Although applications of ML in PV manufacturing have been previously investigated [1,3,4,5], they have been used generally for optimization of a limited number of processes. Furthermore, in these studies, a comparison between different ML algorithms is missing and only small data sets are used for verification [1,3,4,5]. Typically, these studies implement ANN, however, the motivation for choosing this algorithm is lacking. In this research, we aim to provide new insight into the performance of various ML algorithms in process optimization and to explore their capacity to model a complete solar cell manufacturing line.

III. METHODS

Manufacturing data is simulated with the virtual production line (VPL) software that was developed in collaboration between UNSW and PV Lighthouse [6]. VPL simulates the manufacturing of solar cells using ten fabrication processes and 49 input variables. The solar cells electrical properties are modeled using PC1D [7]. VPL includes many interactions to imitate the complexity of solar cell fabrication. In this study, we create three different data sets with different sets of equations (modes) to model the interconnections of the VPL processes. For each mode, 9,500 random process configurations are simulated for training and 500 for testing. For a fair comparison, each algorithm is evaluated to identify its optimal settings, such as the hidden layer structure for ANN, a property that determines the complexity of the model [2].

The performance of the algorithms is assessed by predicting the photovoltaic cell efficiency (PCE) based on the process variables and is measured with the root mean squared error (RMSE) [2] and the coefficient of determination (R^2) [2] in a five-fold cross-validation [2]. The algorithm scoring the lowest RMSE (ANN) is then trained on a balanced data set of 280,000 samples to test its capacity to model a complete production line. Finally, a genetic algorithm [1] is used to optimize the ANN model in order to identify the configuration of each process that achieves the highest PCE.

IV. RESULTS and DISCUSSION

After optimizing the settings of the ML algorithms and POLY, three differently simulated data sets are modeled with POLY, SVR, RF, AB, an ANN with two hidden layers (ANN2) and three hidden layers (ANN3). For each data set, the models are trained on 9,500 data points and tested on the remaining 500. The RMSE scores of these tests are shown in Figure 1(a). The respective performance of the methods is equivalent for the three data sets. This result suggests that ANN is most appropriate for modeling a complete PV manufacturing line, as it scored the lowest RMSE in this comparison.
Next, a balanced data set of 280,000 data points is modeled with ANN3 and its performance is assessed by predicting 15,000 data points that were not seen before. A scatterplot of actual PCE versus predicted PCE is shown in Figure 1(b), scoring an RMSE of 0.29 and an R^2 of 0.99, which is an impressive result. This model is then optimized with a genetic algorithm, which identifies a configuration that achieves a 1.6% increase in PCE compared to the maximum PCE that was included in the data.
Since PV manufacturing lacks broad variation in the data, future research will focus on combining the variance of DoE data with the size of production data.

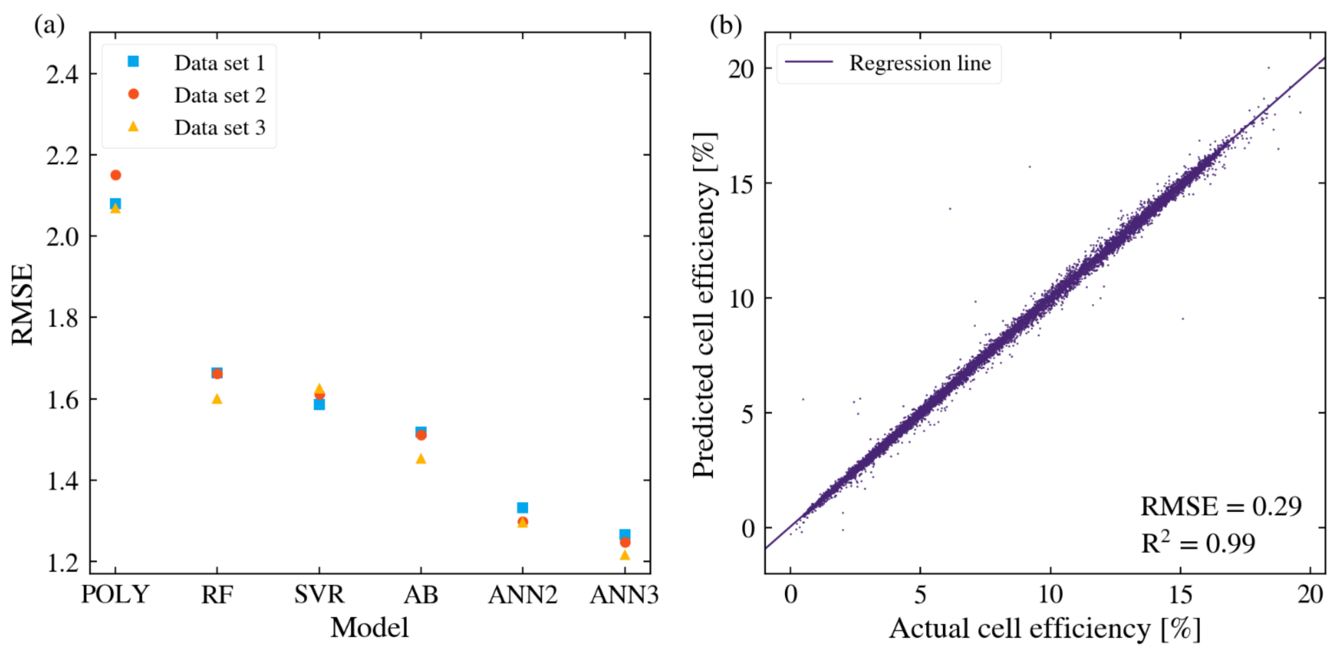

Figure 1: (a) comparison of five ML regression models and polynomial fitting on three different data sets and (b) scatterplot of actual versus predicted of 15,000 data points using ANN3.

V. CONCLUSIONS

In this research, the application of machine learning in the optimization for commercial PV manufacturing was explored. Several ML algorithms were tested on simulated production data and it was demonstrated that all ML algorithms performed better than standard polynomial fitting, with ANN-based algorithms performing the best. An ANN was trained on 280,000 data points to model a complete fabrication process and it was tested on 15,000 samples with an RMSE of 0.29, which is very impressive for such a complex process. This ANN model was then used for optimization, and a new process configuration was identified with a PCE that is 1.6% higher than the maximum PCE in the training set. In the full paper, our ML approach will be developed further to combine DoE data with production data to minimize the amount of experimental data needed.

References

[1] Šibalija, T. V., & Majstorovic, V. D., 2016, *Advanced Multiresponse Process Optimisation*, Berlin, Germany: Springer

[2] Goodfellow, I., Bengio, Y., & Courville, A., 2016, *Deep learning*, MIT press.

[3] Bae, H., 2010, *Optimization of silicon solar cell fabrication based on neural network and genetic programming*, Soft Computing, 14(2), 161-169

[4] Kim, H. S., 2009, *A comparison and analysis of genetic algorithm and particle swarm optimization using neural network models for high efficiency solar cell fabrication processes*, IEEE International Conference on Fuzzy Systems, pp. 1879-1884

[5] Li, Z. Y., & Li, H. S., 2013*, ANN-Model for Plasma Chemical Vapor Deposition,* Process In Advanced Materials Research, Vol. 699, pp. 92-95

[6] Abbott, M., Macintosh, K., 2015, https://factory.pvlighthouse.com.au/, PV Lighthouse

[7] Haug, H., & Greulich, J., 2016, *PC1Dmod 6.2 – Improved Simulation of c-Si Devices with Updates on Device Physics and User Interface*, Energy Procedia, 92, 60-68.