**Characterization of impurity contamination in solar cells with the assistance of machine learning**

**Oleg Olikh**, Oleksii Zavhorodnii

*Taras Shevchenko National University of Kyiv, Kyiv, Ukraine*

*email: olegolikh@knu.ua*

**Abstract**

Research related to sustainable clean energy technology and artificial intelligence is currently one of the most intriguing areas of interest. Publications at the intersection of these two crucial directions are also emerging. For example, machine learning techniques are utilized for solar cells (SCs) design and for the prediction of material properties essential to their production. Our work focuses on developing the physical foundations of a method aimed at assessing impurity concentrations in solar cells, based on applying machine learning to data obtained from current-voltage characteristics measurements. Such a method is express, low-cost, and does not require additional equipment, making it significant for material engineering applications.

The capability of this method is demonstrated by its application to monocrystalline silicon SC, which make up about 90% of worldwide photovoltaic production capacity, and iron atoms - ubiquitous yet efficiency-reducing impurities. Input features used to determine iron concentration NFe included SC parameters (base depth and doping level) and changes in photoelectric parameters (short-circuit current, open-circuit voltage, efficiency, and fill factor) after the decay of iron-boron pairs. The machine learning methods included artificial deep neural networks (DNN) and random forest (RF). The training and test datasets were generated by SC simulation performed using SCAPS-1D software over a temperature range of 290-340 K, under AM1.5 and monochromatic light (940 nm) illumination conditions. The hyperparameters of DNN and RF were optimized through a thorough tuning process. The predictive capabilities of deep neural networks and random forests for iron concentration prediction (range of 1010-1014 cm-3) were explored, depending on the number of input features used. It has been observed that the mean squared error for the test set could be down to 2 10-3, and random forest predictions were less accurate—see Fig.

|  |  |
| --- | --- |
|  |  |
| Fig. The prediction results of the DNN (left panel) and RF (right panel) for the test dataset, obtained for AM1.5 illumination condition. The black lines are the identify lines servings as the references | |