**Predicting Composition of Photo Voltaic Cells Using Neural Networks**

**Introduction:** A better understanding of IV (current voltage) curve data collected from photo voltaic cells may lead to the construction of solar cells with improved electrical properties. With this in mind, IV curve data from different types of solar cells were acquired from the Photovoltaics and Electrochemical Systems Branch, NASA Glenn Research Center. Neural networks were created to predict the chemical composition of three classes of solar cells. The success of these predictions varied with class.

**Language and API:** The neural networks were built in Python using Keras with TensorFlow as a backend, and the data to train the network were acquired from researchers at NASA Glenn Research Center. Data preprocessing steps were conducted using Sci-Kit Learn, and electrical properties of the curve were calculated using R.

**Solar Cell Data:** The IV curves from over 7000 solar cells, representing 22 different types of solar cell materials, were used in this study. The electrical parameters of short circuit current (SCC), reverse saturation current (RCS), shunt resistance (RSH), open circuit voltage (VOC), series resistance (RS), and the diode ideality (DIF) were calculated from the IV curves using the proprietary R code developed at NASA Glenn. The neural network models used to predict the composition of the solar cells were trained on data sets based on these parameters. The solar cell data was first separated into three classes by solar cell material type: “non-alloy compound/elemental,” “single junction and “triple junction.” Non-alloy compound/elemental solar cells are composed of non-alloy compounds and/or elements. Single junction materials are composed of only one material type , whereas triple junction materials possess are composed of layers of three materials each with a different bandgap. There were 5 non-alloy compound/elemental, 15 single junction, and 3 triple junction solar cell materials.

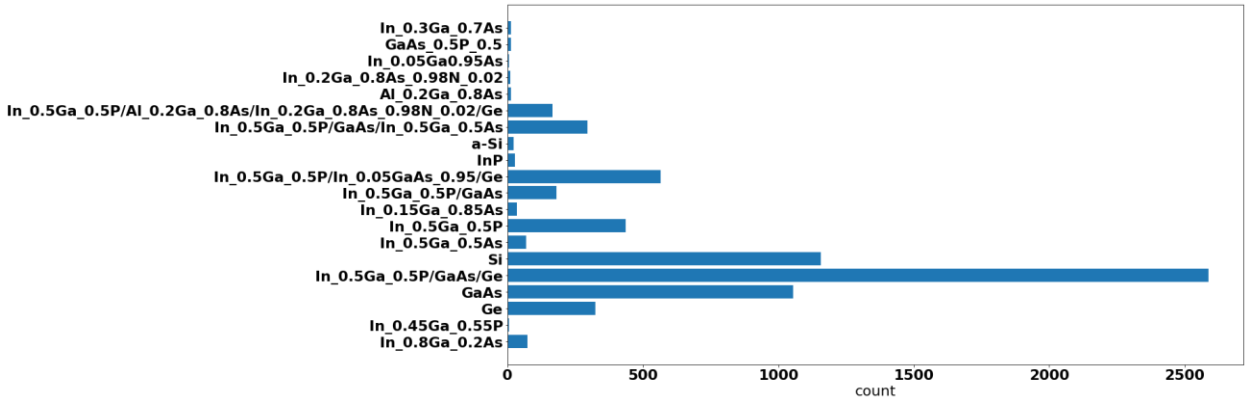



Figure 1: The y-axis is a listing of solar cell materials represented in the received data set. The x-axis is the count of a particular material.

**Neural Network Architecture:** A fully connected neural network was constructed to form the basis of three models for predicting solar cell chemical composition: one model for each of the

non-alloy compound/elemental, single junction, and triple junction type solar cells material type classes. The inputs consisted of five of the electrical parameters derived via the R code from the IV curves, and dropout regularization was used to guard against overfitting. The outputs of each neural network model are probabilities that the correct material is predicted. The material choices in each model are limited to the number of materials in each solar cell material class: 5 for non-alloy compound/elemental, 15 for single junction, and 3 for triple junction.

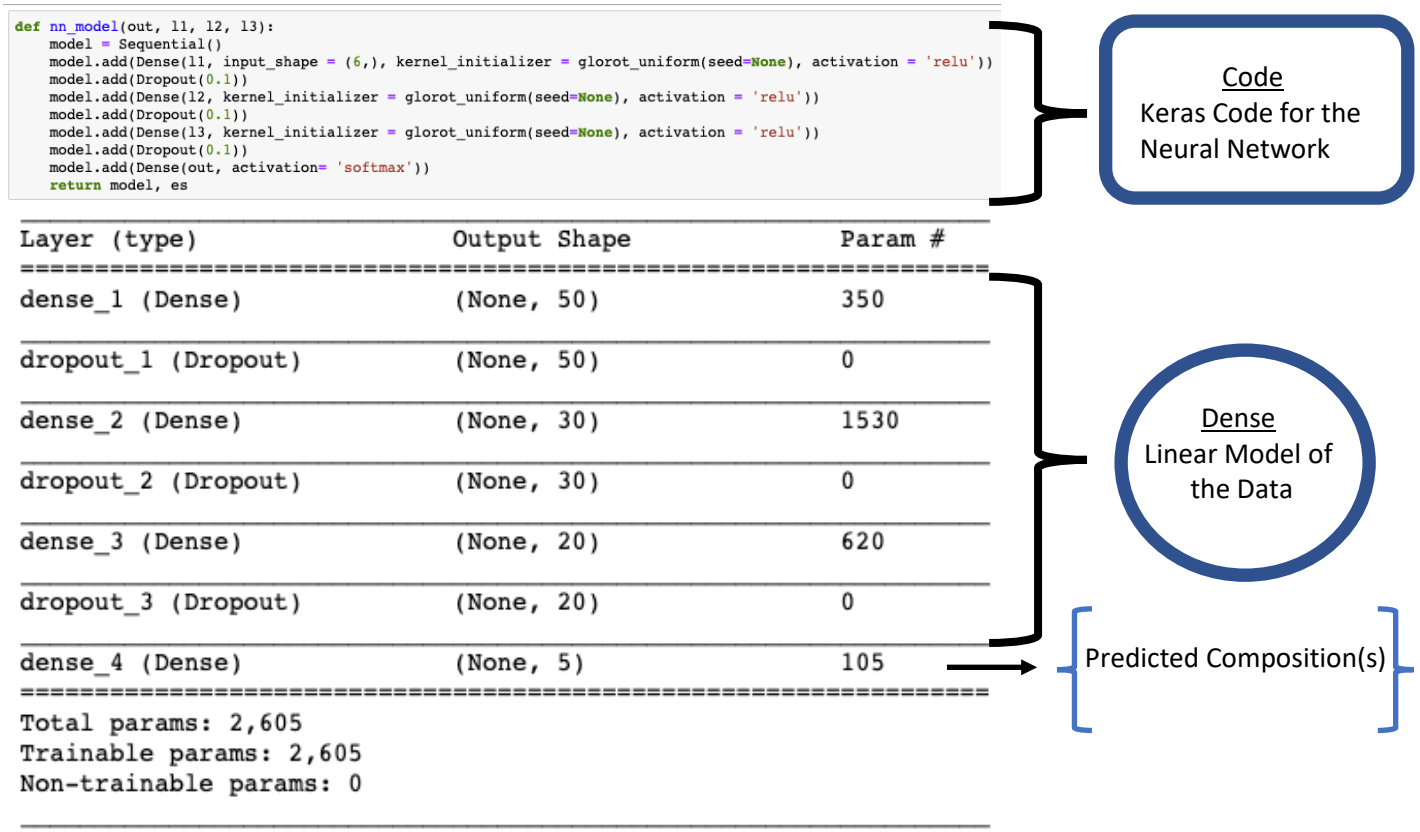
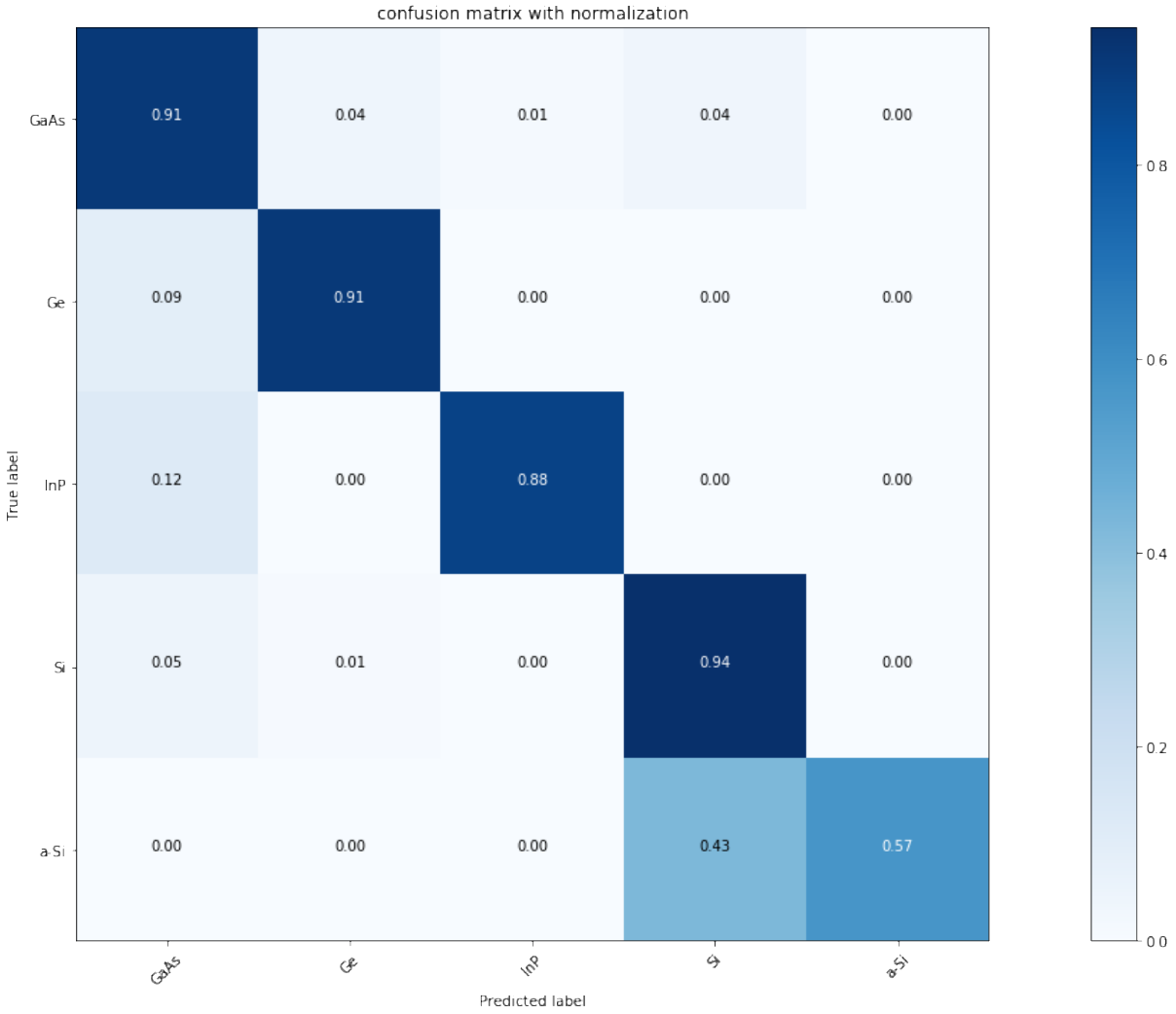


Figure 3: Keras code for the fully connected neural network.

**Results:** Of the three neural network models, the non-alloy compound/elemental solar cell model performed the best. Given the calculated electrical properties of an IV curve, this model correctly predicted which of the 5 solar cell materials generated that IV curve with a sample-weighted average accuracy of 91% (Figure 3a). When presented with calculated electrical data from a given Si IV curve (1156 samples), it correctly predicted Si 94% of the time. This model had the most difficulty distinguishing between a-Si and Si; when presented with calculated electrical data from a given a-Si IV curve (22 samples), the model correctly predicted a-Si only 53% of the time, incorrectly predicting Si the remaining 43% of the time. The single junction model fared well, correctly predicting the composition a sample-weighted average 81% of the time among 3245 samples in 15 materials (Figure 3b). With this model, there were three materials (totaling 85 samples) where the prediction was accurate 100% of the time. There were also a few materials, one of which was a-Si (22 samples), where it had a prediction accuracy of 0%. The triple junction model fared the worst (Figure 3c) at 78%. This reduced accuracy is likely due to the disproportionate representation of $In_{0.5}Ga_{0.5}P$/GaAs/Ge (2588 of 3447 samples) in the data set. Consequently, the model defaulted to predicting $In_{0.5}Ga_{0.5}P$/GaAs/Ge, failing to correctly predict the other compounds the majority of the time.




Figurer 4a: The confusion matrix for the non-alloy compound/elemental solar cell neural network model. Number of sampled IV curves from each of 5 given non-alloy compound/elemental materials: GaAs (1055), Ge (323), InP (27), Si (1156), a-Si (22)

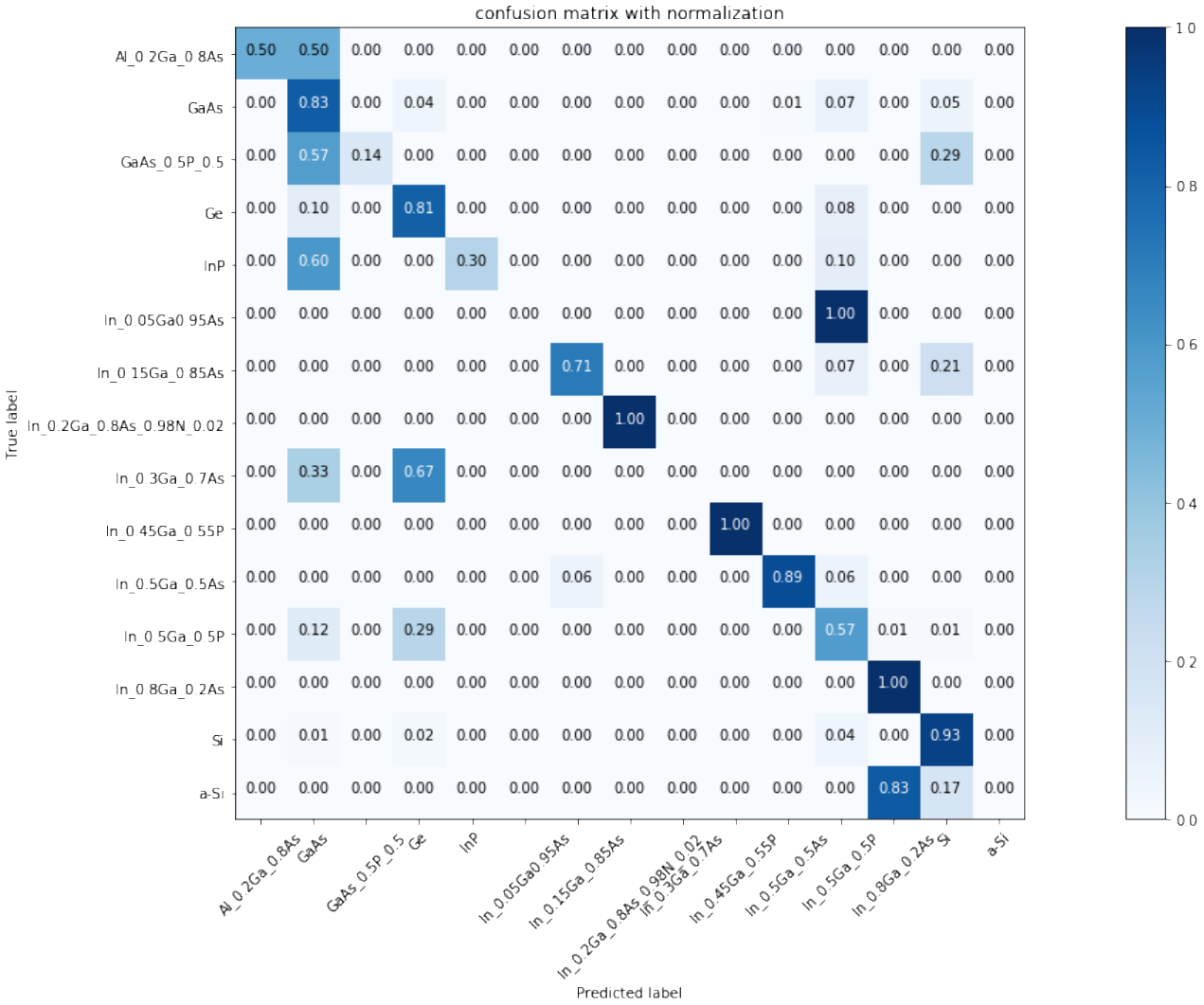



Figure 4b: The confusion matrix for the single junction solar cell neural network model. Number of sampled IV curves from each of 15 given single junction materials: Al_0.2Ga_0.8As (11), GaAs (1055), GaAs_0.5P_0.5 (12), Ge (323), InP (27), In_0.05Ga0.95As (5), In_0.15Ga_0.85As (34), In_0.2Ga_0.8As_0.98N_0.02 (9), In_0.3Ga_0.7As (12), In_0.45Ga_0.55P (4), In_0.5Ga_0.5As (67), In_0.5Ga_0.5P (435), In_0.8Ga_0.2As (73), Si (1156), a-Si (22)

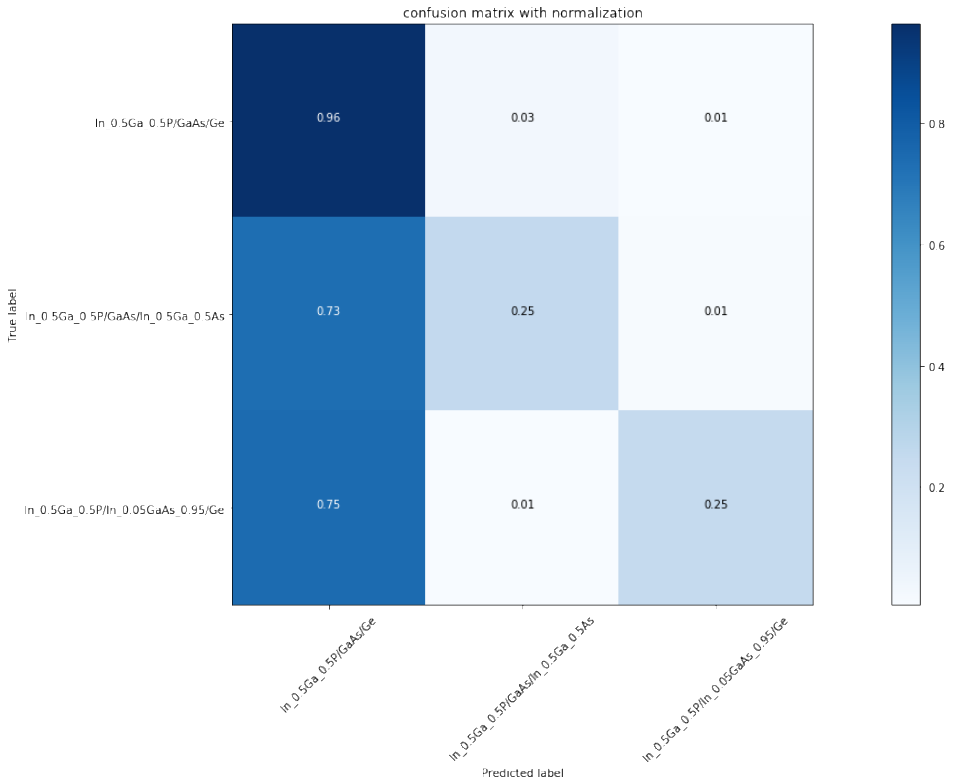



Figure 4c: Confusion matrix for the triple junction solar cell neural network model. Number of sampled IV curves from each of 3 given materials: In_0.5Ga_0.5P/GaAs/Ge (2588), In_0.5Ga_0.5P/GaAs/In_0.5Ga_0.5As (294), In_0.5Ga_0.5P/In_0.05GaAs_0.95/Ge (565)

**Conclusion:** Trained neural network models were used to predict the material composition of solar cells from electrical parameters generated from IV curves. Of the three solar cell materials classes considered, this neural network approach was found to be best at correctly picking non-alloy compound/elemental solar cells, achieving a sample-weighted 91% accuracy. The prediction accuracy for the single junction (81%) and triple junction (78%) models did not fare as well. As material sample counts were often low and were not equally distributed across material types in these cases, improved prediction performance may require larger amounts of more evenly distributed sampled data.