SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

e-ISSN 2147-835X | Period Bimonthly | Founded: 1997 | Publisher Sakarya University |
http://www.saujs.sakarya.edu.tr/en/

Title: Using Physical Parameters for Phase Prediction of Multi-Component Alloys by the Help of TensorFlow Machine Learning with Limited Data

Authors: Kağan ŞARLAR

Recieved: 2020-12-14 15:57:44

Accepted: 2021-04-28 14:07:42

Article Type: Research Article

Volume: 25
Issue: 3
Month: June
Year: 2021
Pages: 766-773

How to cite
Kağan ŞARLAR; (2021), Using Physical Parameters for Phase Prediction of Multi-Component Alloys by the Help of TensorFlow Machine Learning with Limited Data. Sakarya University Journal of Science, 25(3), 766-773, DOI: https://doi.org/10.16984/saufenbilder.840548
Access link
http://www.saujs.sakarya.edu.tr/en/pub/issue/62736/840548

Sakarya University Journal of Science 25(3), 766-773, 2021

Using Physical Parameters for Phase Prediction of Multi-Component Alloys by the Help of TensorFlow Machine Learning with Limited Data

Kağan ŞARLAR*[1]

Abstract

In recent years developing new material and compounds have become more important because of the community's needs. Material scientist and physicist great effort make significant changes in daily life. But nowadays it is important to make these changes in a short time. In this point of view, artificial intelligence and machine learning gives the scientist a great opportunity to predict the properties of new compounds before produced in the laboratory. In this study, the valence electron concentration (VEC), atomic size difference (δ), enthalpy of mixing (ΔH_{mix}), the entropy of mixing (ΔS_{mix}) and electronegativity difference ($\Delta\chi$) values are calculated for each alloy and a dataset has been created. We use gradient boosted trees machine learning method with TensorFlow artificial intelligence program to explore phase selection using an experimental dataset consisting of 118 multi-component alloy system. We divide the whole dataset into two portions with training and evaluate dataset. The training dataset contains 73 and evaluate dataset contains 45 multi-component alloy systems. We also show three of the predicted multi-component alloy system to examine which physical values are used predominantly during prediction. We look at the Receiver Operating Characteristic (ROC) of the results, which will give us a better idea of the tradeoff between the true positive rate and false positive rate. It has been observed that this learning method predicts the structure correctly in 95% of the results with limited data.

Keywords: Multicomponent alloy, Machine learning, TensorFlow, Artificial intelligence.

*Corresponding author: kagansarlar@kmu.edu.tr
[1]Physics Department, Kamil Ozdag Faculty of Sciences, Karamanoglu Mehmetbey University, Yunus Emre Campus, 70100 Karaman,Turkey
Email:kagansarlar@kmu.edu.tr
ORCID: https://orcid.org/0000-0002-8871-2357

1. INTRODUCTION

Since ancient times, humanity has discovered and developed new materials for the development of society [1]. In recent years, these materials cannot meet the needs of developing technology. Therefore, the development of new generation materials is of great importance for industrial and technological applications [2,3]. Multi-component alloys offer a wide range of functional materials to be produced [4]. But scientists have to struggle with the rapid change of needs and long laboratory studies. Today, machine learning is used extensively in many other areas such as fraud prevention, risk analysis, better customer insight, and more complex solutions such as improving medical science [5]. Artificial Intelligence is the general name of the technology that tries to imitate human intelligence. Machine learning is also the application of computational methods that support this technology [6]. While machine learning imitates human intelligence, unlike other applications, it is a set of algorithms which do not need rules that we will interpret and define [7,8]. If we explain how to learn machine learning by comparing it with human learning, when a person learns new information, he/she does not need someone else to load this information into his/her brain. Similarly, in machine learning, the rules are not given beforehand (loading information to the brain), only information-containing datasets are given, and machine learning performs the desired task by understanding these datasets. [6]. TensorFlow is a free and open-source software library for data flow and differentiable programming in a variety of tasks [9]. It is a symbolic math library and is also used for machine learning applications such as neural networks. It is used for both research and production at Google. TensorFlow has been developed by the Google Brain team for internal Google use. It was released for free as open-source under the Apache License 2.0 on November 9, 2015.

Density functional theory (DFT) calculations are a technique used to predict the phase for multi-component alloys. However, to simulate the DFT method, computers must operate for a very long time and have large simulation cells for each alloy. Also, the uncertainties in treating the d orbitals of transition-metal atoms that are often components of multi-component alloys make DFT calculations impractical [10].

Parametric approaches are widely used to predict the phase of multi-component alloys. Guo et al. proposed that the phase selection of multi-component is determined by parameters such as the valence electron concentration (VEC), structure factor (δ), enthalpy of mixing (ΔH_{mix}), the entropy of mixing (ΔS_{mix}) and electronegativity difference ($\Delta\chi$). Unlike the parametric approaches, machine learning (ML) provides a significant tool that offers insight from given data of relevant properties of alloys. Machine learning models, which give fast and accurate results, have been used increasingly in material design in recent years [11-14].

In this study, the valence electron concentration (VEC), structure factor (δ), enthalpy of mixing (ΔH_{mix}), the entropy of mixing (ΔS_{mix}) and electronegativity difference ($\Delta\chi$) values are calculated for each alloy and a dataset has been created. This dataset was trained by machine learning with Gradient Boosted Trees and it was predicted which amorphous or solid solution phase structure would be formed. Data collection is a long process. In this study, it was also aimed to make better estimation with fewer data by using the TensorFlow library with Gradient boosted trees method.

2. METHODS

2.1. Data collection and analysis

We used the data from Guo's phase estimation article [15]. The given dataset contains 118 entries. However, it should not be forgotten that we aim to make a phase estimation with a small amount of data. In Guo's study different physical parameters are used for predicting the alloys' phase amorphous or not. In this study phase estimation was made based on the five physical properties given below as input. Guo et al. [16-18] have used three parameters which are atomic size difference (δ), mixing enthalpy (ΔH_{mix}) and

mixing entropy (ΔS_{mix}). The δ, ΔH_{mix} and ΔS_{mix} which are defined as follows:

$$\delta = 100\sqrt{\sum_{i=1}^{N} c_i(1-r_i/\bar{r})^2} \quad (1)$$

$$\Delta H_{mix} = \sum_{i=1, i\neq j}^{n} \Omega_{ij} c_i c_j \quad (2)$$

$$\Delta S_{mix} = -R\sum_{i=1}^{N} c_i \ln c_i \quad (3)$$

where *N* is the number of components of the alloy, c_i is the atomic percentage of element *i* in the alloy. $\bar{r}(=\sum_{i=1}^{n} c_i r_i)$, is the average atomic radius. r_i is the atomic radius which found for all elements in ref [15]. $\Omega_{ij}(=\Delta H^{mix}{}_{AB})$ is the regular melt-inter action parameter between i^{th} and j^{th} elements, $\Delta H^{mix}{}_{AB}$ is the heat of mixing [19], and *R* is the gas constant.

Besides these values, valence electron concentration and electronegativity difference were added to the data to diversify learning. Electronegativity difference is defined as follows:

$$\Delta\chi = \sqrt{\sum_{i=1}^{n} C_i(\chi_i - \bar{\chi})^2} \quad (4)$$

where $\bar{\chi} = \sum_{i=1}^{n} C_i \chi_i$, χ_i is the Pauling electronegativity for the *i*th element.

Valance electron concentration (VEC) is calculated by

$$VEC = \sum_{i=1}^{n} C_i (VEC)_i \quad (5)$$

where $(VEC)_i$ is the VEC for the *i*th element.

Before transferring the data of 118 multi-component alloy to machine learning, we divide the whole dataset into two portions with training and evaluate dataset. The training dataset contains 73 and evaluate dataset contains 45 multi-component alloy systems. And statistical data analysis Pearson correlation coefficients were performed overall 118 multi-component alloy system. Pearson's correlation was calculated for each pair of features. It has been observed that there is no strong correlation (P~1.0) between any two features. As a result, we suggest that these five features can be used in machine learning. The data in this work can also be found at https://github.com/kaansarlar/Gradientboostedtrees-article-data.

2.2. Gradient Boosted Trees

Gradient Boosted Trees is a machine learning technique, for regression and classification problems [20]. Boosting is a method of transforming a weak learner into a strong learner [21]. It does this gradually with iterations. Gradient Boosted Trees are calculated in four steps. First, the loss function is defined as $L(y_i, F(x))$. y_i is observed value and $F(x)$ is the predicted value. The loss function is a measure of how well the coefficients of the model fit the underlying data. In step two, the constant variable is determined. $F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$ In the formula, the value in sigma is the loss function. γ is the predicted value. The loss function will be summed up for all observations and the minimum state of its value will be found. For this, the loss function is derivatized, the values added up and set to zero. As a result, there is the initial leaf. This value is equal to the average of all values in the target variable. Step three takes place in 4 stages and is a cycle that will be applied to all trees. Errors are calculated based on the previous estimate.

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad i = 1,.....,n$$ In the formula, r means residual. i observation number, m denotes the number of the established tree. It is a parenthetically differentiated loss function. When looked outside the parentheses, it is seen that the value of F (x) represents the output of the previous tree. With this formula, residues are calculated for all observations. Step four, a

decision tree is created for residuals and the value of each leaf is found. $\gamma_{jm} = \arg\min_{\gamma} \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma) \; j = 1 \ldots J_m$. The formula minimizes the error for each leaf. In the same way, the derivation of the loss function is taken and the values are summed and equalled to the class. The number that comes out is the value of the leaf. Estimates are made for each observation. $F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$ when the formula is examined, it is seen that F (x) = results of the previous tree + learning rate * new tree. The cycle will continue like this. The flow chart of gradient boosted trees is shown in Figure 1.

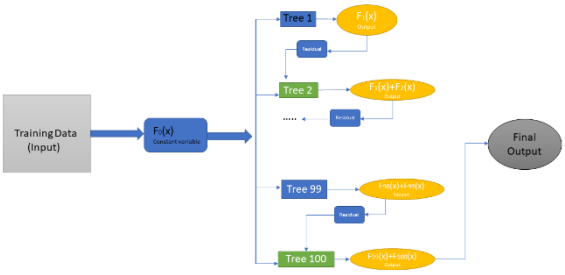

Figure 1 Flow chart of the gradient boosted trees

3. RESULTS AND DISCUSSION

All code is written in python and TensorFlow's estimator is used. First, the data is divided into two parts. Firstly, 73 data were taught to the model as train data. The model was trained on the train data set. Then, the trained model was shown in the Evaluate data set with 45 data, and it was asked to predict whether the alloys were amorphous or solid solutions.

Table 1 Results of the model after processing the trained data series.

auc	0.951428
auc_precision_recall	0.987701
average_loss	0.377974
label/mean	0.777778
loss	0.377974
precision	0.968750
prediction/mean	0.691700
recall	0.885714
global_step	100.000000

Table 1 shows the models' result after training. Here auc is the percentage of correct number of classifications, auc_precision_recall is the percentage of relevant instances, among the retrieved instances, that have been retrieved over the total amount of relevant instances, average_loss is the average value of loss function given the current batches, the label/mean is the mean of the value labels, the loss is current value of the loss, precision is how many of the values we guess as positive are actually positive, prediction/mean is the value of the corresponding predictions. (two classes could give you a value between 0 and 1), recall computes the recall of the predictions with respect to the labels, global_step is the number of iterations. The models' prediction accuracy is 95 %. In figure 2. we examined which physical values are used predominantly during the prediction of the alloy $Al_{0.8}CrCu_{1.5}FeMnNi$ in solid solution form. Here, alloy $Al_{0.8}CrCu_{1.5}FeMnNi$ is chosen at random. The prediction of our model for this alloy is 97%

for the solid solution and 3% for amorphous structure. To avoid any errors in the program, the d symbol is used instead of the Δ symbol. As seen in this prediction, the entropy of mixing and enthalpy of mixing contributed more than other physical parameters.

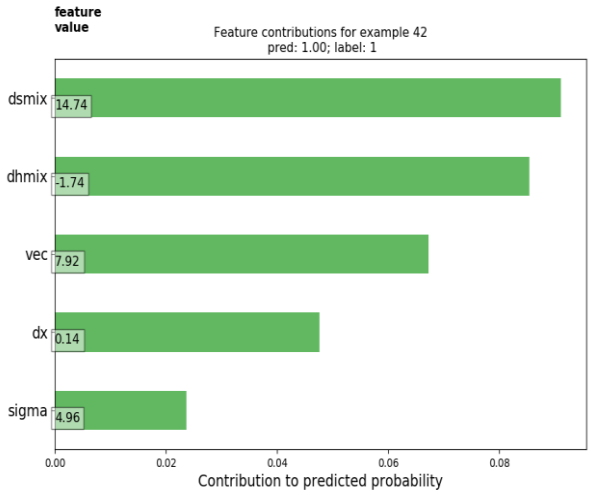

Figure 2 Contribution of physical parameters in the prediction of the phase structure of $Al_{0.8}CrCu_{1.5}FeMnNi$.

Figure 3 shows the prediction results of the alloy $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10}Be_{22.5}$ using the gradient boosted trees model. The prediction of our model for this alloy is 2% for the solid solution and 98% for amorphous structure.

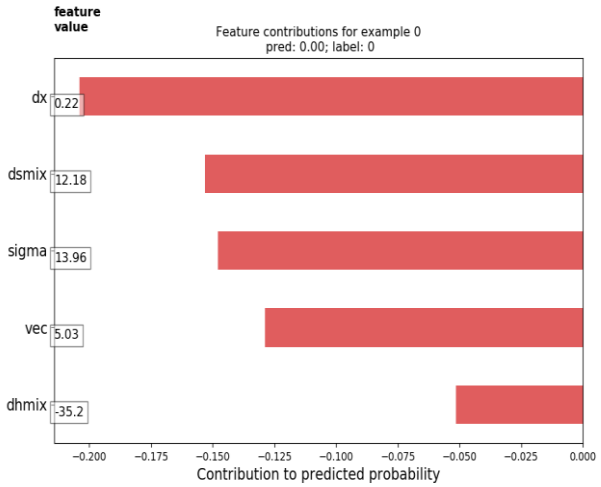

Figure 3 Contribution of physical parameters in the prediction of the phase structure of $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10}Be_{22.5}$.

As seen in this prediction, electronegativity difference and entropy of mixing contributed more than other physical parameters. When Figure 2 and Figure 3 were compared, the model perceived positive contributions of physical properties as the solid solution and negative contributions as the amorphous phase. Not all predictions we make for multi-alloy systems have such high predictions. Figures 4 shows an example for the amorphous phase.

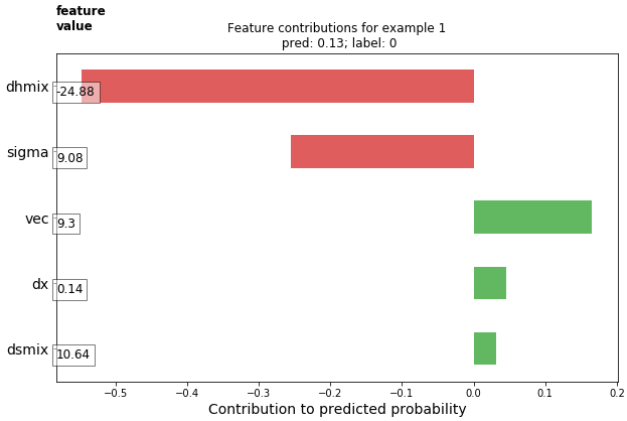

Figure 4 Contribution of physical parameters in the prediction of the phase structure of $Pd_{40}Cu_{30}Ni_{10}P_{20}$.

In Figure 4 the calculated probabilities for the $Pd_{40}Cu_{30}Ni_{10}P_{20}$. alloy is 87% for the amorphous phase and 13% for the solid solution, respectively.

The accuracy of the phase estimation results of the $Al_{0.8}CrCu_{1.5}FeMnNi$, $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10}Be_{22.5}$, $Pd_{40}Cu_{30}Ni_{10}P_{20}$ alloy systems shown as an example out of 45 samples. In Table-2, the experimental results of these analysed alloy systems and machine learning phase predictions are compared.

Table 2 Comparison of machine learning results with experimental results.

Alloy System	**Experimental Phase**	**Machine Learning estimation (Solid-State Phase)**	**Machine Learning estimation (Amorphous Phase)**	**References**

$Al_{0.8}CrCu_{1.5}FeMnNi$	Solid-State	97 %	3 %	[15]
$Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10}Be_{22.5}$	Amorphous Phase	2 %	98 %	[15]
$Pd_{40}Cu_{30}Ni_{10}P_{20}$	Amorphous Phase	13 %	87 %	[15]

The Receiver Operating Characteristic (ROC) curve arises from the ratio of sensitivity to precision in cases where the discrimination threshold differs in binary classification systems. ROC can be simply expressed as the fraction of true positives versus false positives. As with any classification process, the methods strive to strike the balance between precision (the ability to eliminate false positives) and sensitivity (the ability to detect correct positives). In figure 5 we can also look at the ROC of the results, which will give us a better idea of the tradeoff between the true positive rate and false-positive rate.

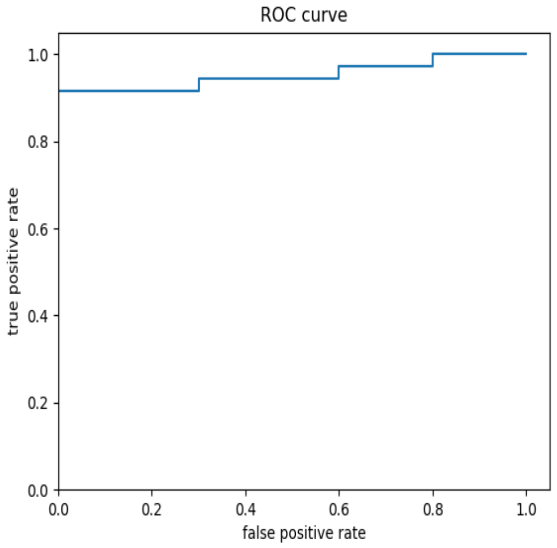

Figure 5 Receiver Operating Characteristic curve

4. CONCLUSION

In this study, the phase prediction of multi-component alloys has been tried to be made with limited data. The machine learning method chosen for prediction is Gradient boosted trees. The accuracy of the method is 95%. When the contribution of each physical property is examined, it is seen that the additive amounts for each alloy are different. With this method, it can be predicted in which phase group the alloy to be produced will be, before starting the experimental work, with a highly accurate prediction. Thus, a preliminary study can be done before the experiment is carried out. We expect this work to guide the design and phase prediction of new multi-component alloys.

Acknowledgments

The author would like to thank the SAUJS editors and reviewers who reviewed the study.

Funding

The author received no financial support for the research, authorship, and/or publication of this paper.

The Declaration of Conflict of Interest/ Common Interest

No conflict of interest or common interest has been declared by the author.

The Declaration of Ethics Committee Approval

This study does not require ethics committee permission or any special permission.

The Declaration of Research and Publication Ethics

The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1] P. F. Pagoria, G. S. Lee, A. R. Mitchell, and R. D. Schmidt, 'A review of energetic materials synthesis', *Thermochim. Acta*, vol. 384, no. 1–2, pp. 187–204, 2002.

[2] M. H. Phan and S. C. Yu, 'Review of the magnetocaloric effect in manganite materials', *J. Magn. Magn. Mater.*, vol. 308, no. 2, pp. 325–340, 2007.

[3] M. N. Gueye, A. Carella, J. Faure-Vincent, R. Demadrille, and J. P. Simonato, 'Progress in understanding structure and transport properties of PEDOT-based materials: A critical review', *Prog. Mater. Sci.*, vol. 108, no. November 2019, p. 100616, 2020.

[4] R. Shan, J. Han, J. Gu, H. Yuan, B. Luo, and Y. Chen, 'A review of recent developments in catalytic applications of biochar-based materials', *Resour. Conserv. Recycl.*, vol. 162, no. June, p. 105036, 2020.

[5] E. Glikson and A. W. Woolley, 'Human trust in artificial intelligence: Review of empirical research', *Acad. Manag. Ann.*, vol. 14, no. 2, pp. 627–660, 2020.

[6] C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, and S. P. Ong, 'A Critical Review of Machine Learning of Energy Materials', *Adv. Energy Mater.*, vol. 10, no. 8, pp. 1–36, 2020.

[7] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, 'Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review', *Chaos, Solitons and Fractals*, vol. 139, p. 110059, 2020.

[8] I. Z. A. D. P. No and W. Naudé, 'DISCUSSION PAPER SERIES Artificial Intelligence against COVID-19 : An Early Review Artificial Intelligence against COVID-19 : An Early Review', no. 13110, 2020.

[9] 'Abadi, Mart'in, Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J, "Tensorflow: A system for large-scale machine learning." In 12th $USENIX$ Symposium on Operating Systems Design and Implementation ($OSDI$ 16) pp. 265–283, 2016

[10] R. Feng, P. K. Liaw, M. C. Gao, and M. Widom, 'First-principles prediction of high-entropy-alloy stability', *npj Comput. Mater.*, vol. 3, no. 1, p. 50, Dec. 2017.

[11] M. Rupp, A. Tkatchenko, K. R. Müller, and O. A. Von Lilienfeld, 'Fast and accurate modeling of molecular atomization energies with machine learning', *Phys. Rev. Lett.*, vol. 108, no. 5, pp. 1–5, 2012.

[12] Y. M. Zhang, S. Yang, and J. R. G. Evans, 'Revisiting Hume-Rothery's Rules with artificial neural networks', *Acta Mater.*, vol. 56, no. 5, pp. 1094–1105, 2008.

[13] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, 'A general-purpose machine learning framework for predicting properties of inorganic materials', *npj Comput. Mater.*, vol. 2, no. July, pp. 1–7, 2016.

[14] B. Meredig and C. Wolverton, 'A hybrid computational-experimental approach for automated crystal structure solution', *Nat. Mater.*, vol. 12, no. 2, pp. 123–127, 2013.

[15] S. Guo and C. T. Liu, 'Phase stability in high entropy alloys: Formation of solid-solution phase or amorphous phase', *Prog. Nat. Sci. Mater. Int.*, vol. 21, no. 6, pp. 433–446, 2011.

[16] S. Guo, C. Ng, J. Lu, and C. T. Liu, 'Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys', *J. Appl. Phys.*, vol. 109, no. 10, 2011.

[17] S. Guo, 'Phase selection rules for cast high entropy alloys: an overview', *Mater. Sci. Technol.*, vol. 31, no. 10, pp. 1223–1230,

2015.

[18] Z. Wang, S. Guo, and C. T. Liu, 'Phase Selection in High-Entropy Alloys: From Nonequilibrium to Equilibrium', *JOM*, vol. 66, no. 10, pp. 1966–1972, Oct. 2014.

[19] A. Takeuchi and A. Inoue, 'Classification of Bulk Metallic Glasses by Atomic Size Difference, Heat of Mixing and Period of Constituent Elements and Its Application to Characterization of the Main Alloying Element', *Mater. Trans.*, vol. 46, no. 12, pp. 2817–2829, 2006.

[20] G. Ke *et al.*, 'LightGBM: A highly efficient gradient boosting decision tree', *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 3147–3155, 2017.

[21] A. J. Ferreira and M. A. T. Figueiredo, 'Boosting Algorithms: A Review of Methods, Theory, and Applications', in *Ensemble Machine Learning: Methods and Applications*, C. Zhang and Y. Ma, Eds. Boston, MA: Springer US, 2012, pp. 35–85.