PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Implementation of multi-task learning neural network architectures for robust industrial optical sensing

Francesca Venturini, Umberto Michelucci, Michael Baumgartner

Francesca Venturini, Umberto Michelucci, Michael Baumgartner, "Implementation of multi-task learning neural network architectures for robust industrial optical sensing," Proc. SPIE 11782, Optical Measurement Systems for Industrial Inspection XII, 117822H (20 June 2021); doi: 10.1117/12.2593469

SPIE.

Event: SPIE Optical Metrology, 2021, Online Only

Implementation of multi-task learning neural network architectures for robust industrial optical sensing

Francesca Venturini[a,b], Umberto Michelucci[b], and Michael Baumgartner[a]

[a]Institute of Applied Mathematics and Physics, Zurich University of Applied Sciences, Technikumstrasse 9, 8401 Winterthur, Switzerland
[b]TOELT LLC; Birchlenstr. 25, 8600 Dübendorf, Switzerland

ABSTRACT

The simultaneous determination of multiple physical or chemical parameters can be very advantageous in many sensor applications. In some cases, it is unavoidable because the parameters of interest display cross sensitivities or depend on multiple quantities varying simultaneously. One notable example is the determination of oxygen partial pressure via luminescence quenching. The measuring principle is based on the measurement of the luminescence of a specific molecule, whose intensity and decay time are reduced due to collisions with oxygen molecules. Since both the luminescence and the quenching phenomena are strongly temperature-dependent, this type of sensor needs continuous monitoring of the temperature. This is typically achieved by adding temperature sensors and employing a multi-parametric model (Stern–Volmer equation), whose parameters are all temperature-dependent. As a result, the incorrect measurement of the temperature of the indicator is a major source of error. In this work a new approach based on multi-task learning (MTL) artificial neural networks (ANN) was successfully implemented to achieve robust sensing for industrial applications. These were integrated in a sensor that not only does not need the separate detection of temperature but even exploits the intrinsic cross-interferences of the sensing principle to predict simultaneously oxygen partial pressure and temperature. A detailed analysis of the robustness of the method was performed to demonstrate its potential for industrial applications. This type of sensor could in the future significantly simplify the design of the sensor and at the same time increase its performance.

Keywords: oxygen monitoring, luminescence quenching, artificial neural networks, multi-task learning, error limited accuracy

1. INTRODUCTION

The determination of molecular oxygen concentration is of great interest in numerous areas since oxygen plays an important role in respiration and metabolic processes in living organisms. The applications are not limited to research fields but include applications in many industries: food packaging, beverage production and bottling, agriculture for plant respiration and soil aeration, biopharma, aquaculture, to mention only few.

One of the most used optical methods, successfully industrialized for several years, is based on luminescence quenching: the intensity and decay time of a specific luminophore are reduced due to collisions with molecular oxygen.[1] Among the most frequently used luminophores are porphyrin and metalloporphyrin. Among these, PtTFPP and PdTFPP are particularly attractive because they display a strong phosphorescence, good photo-stability, are strongly quenched by oxygen, and have long lifetimes. These characteristics are important for the realization of an industrial sensor because they directly impact on the signal-to-noise ratio and the long time stability. The long lifetimes, additionally, make the realization of the excitation and interrogation electronics simpler. However, both the luminescence itself and its quenching by oxygen are temperature dependent, which requires continuous and fast determination of the temperature of the luminophore. This is most frequently achieved with a separate sensor, for example, a resistance temperature detector, whose response is then used to correct the calculated oxygen concentration. This is a challenging task for practical implementation and can be

Further author information: (Send correspondence to F.V.)
F.V.: E-mail: francesca.venturini@zhaw.ch

Optical Measurement Systems for Industrial Inspection XII, edited by Peter Lehmann, Wolfgang Osten, Armando Albertazzi Gonçalves, Proc. of SPIE Vol. 11782, 117822H
 · doi: 10.1117/12.2593469

a significant source of error in oxygen determination. Therefore, to achieve a robust industrial oxygen sensor, it is essential to simultaneous determination both oxygen and temperature.

In previous work it was demonstrated that it is possible to apply multi-task learning (MTL) artificial neural network (ANN) architectures to predict both oxygen concentration and temperature,[2] making the oxygen determination temperature immune. The advantage of this type of approach is that not only the sensor does not need the separate detection of temperature, but it even exploits the intrinsic cross-interferences of the measuring approach (luminescence quenching).

In this work, the robustness of approach and the accuracy of the sensor are investigated in detail to demonstrate that the application of ANN to industrial sensor can enable the realization of sensors with simpler concept and construction, but higher accuracy and reliability.

2. MATERIALS AND METHODS

2.1 Experimental

The luminophore used for oxygen detection was PtTFPP, commercially available as Oxygen Sensor Spot (PSt3, PreSens GmbH, Regensburg, Germany). The excitation light was provided by a 405 nm LED, filtered by a short pass filter, and focused on the surface of the sample with a collimation lens. The luminescence focused by a lens was collected by a photodiode after filtering with a band pass filter. The frequency modulation for the excitation light and luminescence phase shift determination were performed with a commercial lock-in amplifier. The details of the experimental setup are described in a previous paper.[3]

For the realization of the oxygen sensor based on ANN, a large amount of data under varying oxygen concentration, temperature, and modulation frequency conditions were carried out. The phase shift was measured for 50 modulation frequencies between 200 Hz and 15 kHz, at 21 oxygen concentrations between 0% air and 100% air and at 9 temperatures between 5 °C and 45 °C and each measurement was repeated 20 times, for a total 189'000 phase shift values. This was performed by a fully automated program for both instrument control and the automatic data acquisition. The automated acquisition procedure is explained in detail in a previous work.[2]

The availability of a large set of experimental data was an essential prerequisite for the training and test of the ANN. Additionally, it allowed to test the the classical approach based on an analytical model when applied to measurements under varying conditions of oxygen concentration, temperature, and modulation frequency, as described in Section 3.1.

2.2 Artificial Neural Network Model

The proposed approach uses an ANN with a specifically designed architecture instead of describing the response of the sensor through an analytical model, like a Stern-Volmer two-site model.[1] Thanks to the MTL architecture,[4,5] the ANN predicts the oxygen concentration without any information about the temperature. The MTL architecture used in this work (Figure 1) is characterized by a set of common hidden layers, whose output is then the input of multiple branches of task-specific hidden layers, that learn to predict the oxygen concentration and temperature at the same time. The details of the architecture and the parameters for the training are described in detail in a previous authors' work.[6]

To assess the performance of the ANN, the dataset was split in two parts:[7,8] a training set (80% of the available observations) indicated with S_{train} for the training of the ANN, and a development set (20% of the available observations) indicated with S_{dev}, for the validation. The results shown in this work were obtained by calculating the metrics on S_{dev}.

The metrics to evaluate the performance of the sensor are the absolute error (AE) and its average over the a specific set of observations, the mean absolute error (MAE). The AE for the j^{th} observation is indicated with $AE^{[j]}$, wile the MAE evaluated on a dataset S will be indicated with $MAE(S)$. A subscript will indicate for which quantity ($[O2]$ or T) it is calculated. For example, for the oxygen concentration $[O_2]$ and the j^{th} observation the $AE^{[j]}_{[O_2]}$ is

$$AE^{[j]}_{[O_2]} = |[O_2]^{[j]}_{pred} - [O_2]^{[j]}_{meas}|. \tag{1}$$

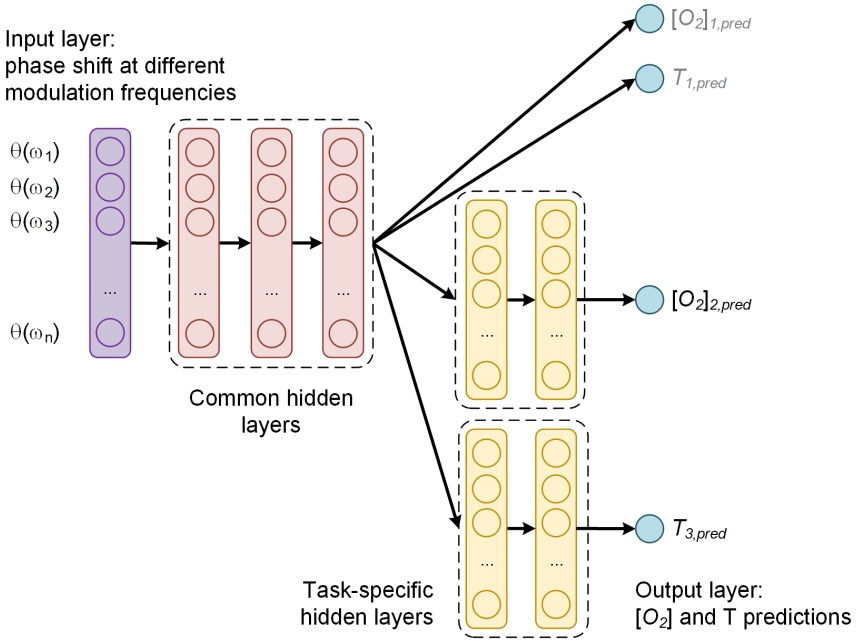

Figure 1. Architecture of the multi-task learning neural network. The input are the normalized phase shit at different modulation frequencies. The common hidden layers generate a shared representation, that is the input to task-specific branches. The output are the oxygen and temperature predictions. The predictions of the common hidden layer were disregarded in this work.

where $[O_2]^{[j]}_{pred}$ and $[O_2]^{[j]}_{meas}$ are the predicted and expected oxygen concentration for observation j respectively. The $MAE_{[O_2]}(S_{dev})$ is defined as

$$MAE_{[O_2]}(S_{dev}) = \frac{1}{|S_{dev}|} \sum_{j \in S_{dev}} |[O_2]^{[j]}_{pred} - [O_2]^{[j]}_{real}| \tag{2}$$

where $|S_{dev}|$ is the size (or cardinality) of the training dataset. The AE_T and MAE_T are similarly defined.

When working with ANNs, it is important to carefully investigate how the MAE depends on the particular choice of the training and test data. To be usable in industrial applications, a sensor needs to achieve high reproducibility in the prediction accuracy. The average and standard deviation of the MAE, calculated for large number of dataset splits, can be used as an indicator of the robustness of the ANN predictions. From a practical point of view, this is unfortunately very time consuming. The training of the described ANN for one single dataset split with an optimal set of hyperparameters (10^5 epochs, mini-batch size of 32) requires ca. 4 hours on a 2.3 GHz 8-Core IntelCore i9 with 32 GB 2667 MHz DDR4 memory. This problem can be addressed by using the resampling technique bootstrap with a reduced number of dataset splits to have enough MAE values to study.[9] The method is described in detail for completeness in Appendix A in Algorithm 1. The Algorithm 1 describes the steps necessary for the evaluation of average and standard deviation of $MAE_{[O_2]}$ and MAE_T over multiple dataset splits and multiple samples generated via bootstrap on the validation dataset obtained in each split.

Finally, using the error limited accuracy[2] (ELA), the minimum absolute error value for which all observations are predicted within the given error was evaluated. In other words, the minimum value of the absolute error (indicated with $\overline{AE}$) for which the network predicts all the observations correctly is calculated for all splits. This value ($\overline{AE}$) can be interpreted as the biggest error expected in the sensor predictions.

3. RESULTS AND DISCUSSION

3.1 Limitations of the Stern-Volmer relationship and of the two-site model

The conventional approach to describe the relationship between the oxygen concentration and the luminescence intensity or decay is the Stern-Volmer relationship. When the the indicator is embedded in a matrix or substrate, however, the Stern-Volmer linear behaviour is rarely observed.[1,10] The deviations are attributed to heterogeneities of the matrix or to the presence of static quenching. The most widely used analytical model to describe the deviation from the linear behaviour for oxygen quenching is the two-site model.[11–14] The idea behind the model is that the indicator is embedded in at least two environments, and therefore quenched at different rates.

For industrial applications, the measurement of the luminescence decay time in the frequency domain is the preferred method due to its higher reliability and robustness.[15] In this case, the excitation light is modulated; the luminescence light is also modulated, but shows a phase shift with respect to the excitation light. The oxygen concentration can be then determined from the phase shift as[3]

$$\frac{\tan\theta_0}{\tan\theta} = \left(\frac{f}{1+K_{SV1}\cdot[O_2]} + \frac{1-f}{1+K_{SV2}\cdot[O_2]}\right)^{-1} \tag{3}$$

where θ_0 and θ, respectively, are the phase shifts in the absence and presence of oxygen, f and $1-f$ are the fractions of the total emission for each component under unquenched conditions, and K_{SV1} and K_{SV2} are the associated Stern–Volmer constants for each component.

The complexity of the implementation of this approach is best illustrated when considering that the quantities f, K_{SV1}, and K_{SV2} are not constants, but they actually depend on both the angular modulation frequency ω and the temperature. This is illustrated in Figure 2, where the above-mentioned parameters are plotted for selected modulation frequencies and temperatures. The parameters were obtained via standard nonlinear fitting procedures with the model of Eq. (3) from the data at constant temperature and modulation frequency.

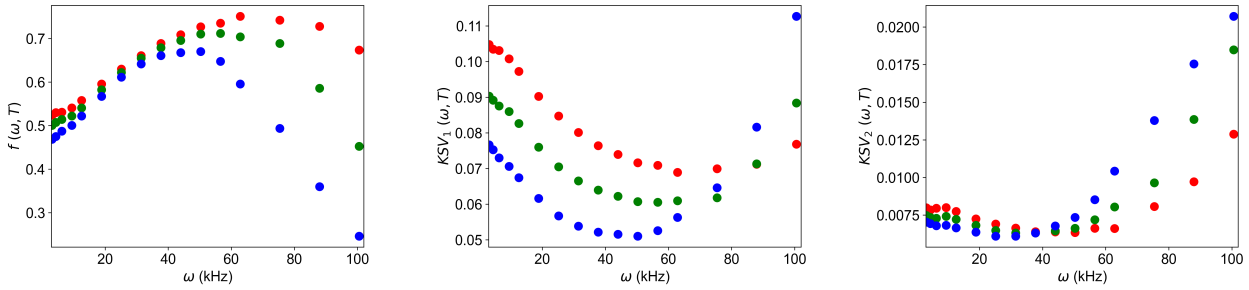

Figure 2. Dependence of the parameters of f, K_{SV1}, and K_{SV2} from the modulation frequency obtained by fitting the measurement at 5 °C (blue), 25 °C (green), and 45 °C (red).

Although the temperature dependence of K_{SV1} and K_{SV2} might be physically justified due to, for example to an increased permeability for the oxygen molecule at higher temperatures, the dependence of the parameters from the modulation frequency has no physical explanation and is rather an artefact of the model. Alternatives to the two-site models have been proposed, like the non-linearity solubility model[16] or the description of the Stern-Volmer relationship in terms of thermodynamic parameters.[17] Both approaches, however, do not simplify the description of the behaviour, but rather offer different physically-based parametrizations.

3.2 Performance of the ANN model

The robustness of the predictions was tested by splitting the development dataset six times and creating 1000 bootstrap samples each time, and by calculating the $MAE_{[O_2]}$ and MAE_T for each of the 6000 bootstrap samples. Table 1 shows the average and standard deviation of the 6000 values of $MAE_{[O_2]}$ and MAE_T. The results reported here were obtained with a mini-batch size of 32. The results are shown in Figure 3 and summarized in Table 1.

Parameter	$\langle MAE \rangle$	$\sigma(MAE)$
$[O_2]$	0.18 (% air)	0.03 (% air)
T	0.20 (°C)	0.03 (°C)

Table 1. Average $\langle MAE \rangle$ and standard deviation σ of $MAE_{[O_2]}$ and MAE_T of 6000 bootstrap samples obtained by splitting the development dataset six times and creating 1000 bootstrap samples each time.

The statistical variation of $\langle MAE_{[O_2]} \rangle$ and $\langle MAE_T \rangle$ was estimated by creating 30000 bootstrap samples of the size of $|S_{dev}|$ from the 6000 samples available. The detailed steps are described in 1 in Appendix A. Figure 3 shows the resulting distributions of $\langle MAE_{[O_2]} \rangle$ and $\langle MAE_T \rangle$. For each distribution, the kernel density estimation of the distributions was calculated.[2] As expected the $\langle MAE_{[O_2]} \rangle$ and $\langle MAE_T \rangle$ distributions approximate a Gaussian shape.[9] The results indicate that average of the $\langle MAE_{[O_2]} \rangle$ is below 0.2 (% air) with a standard deviation of only 0.03 (% air) and that the average of the $\langle MAE_T \rangle$ is 0.2 °C, thus demonstrating that such a sensor can be used confidently since the results are extremely stable and the standard deviations have very small values.

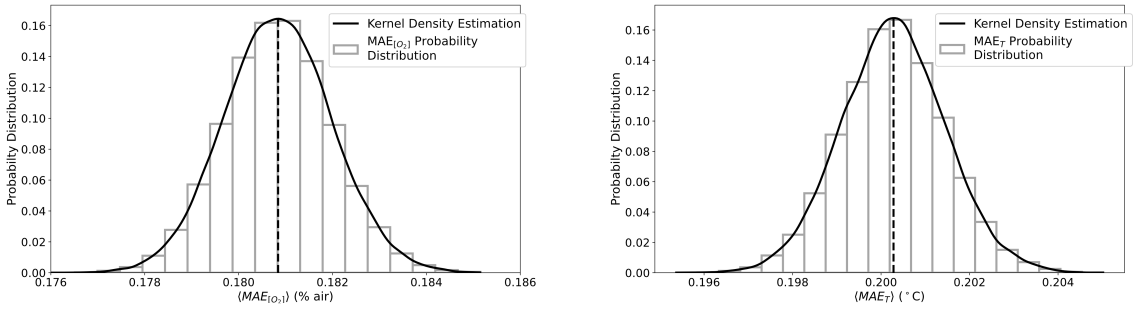

Figure 3. The two plots show the distributions of $\langle MAE_{[O_2]} \rangle$ and of $\langle MAE_T \rangle$. The distributions have a Gaussian shape. The continuous lines are plots of the kernel density estimation of the distributions.

The average and standard deviation of the MAE, however, do not tell the entire story. For practical industrial applications, the maximum AE to be expected by a trained ANN on a specific dataset is perhaps the most important quantity. This can be calculated through the ELA as the value of the absolute error value for which all observations are predicted within this given absolute error. These values, $\overline{AE}_{[O_2]}$ and $\overline{AE}_T$, obtained for the 6 splits are summarized in Table 2. The average and standard deviation of these results can be easily calculated and gives $\langle \overline{AE}_{[O_2]} \rangle = 1.22$ % air, $\langle \overline{AE}_T \rangle = 2.05$ °C, $\sigma(\overline{AE}_{[O_2]}) = 0.31$ % air and $\sigma(\overline{AE}_T) = 0.95$°C.

Split	$\overline{AE}_{[O_2]}$ (% air)	$\overline{AE}_T$ (°C)
1	1.29	2.30
2	1.81	2.98
3	0.97	1.15
4	1.05	1.18
5	1.29	2.13
6	0.88	1.18

Table 2. Values of the absolute errors for which all the oxygen concentration and temperature are predicted within these absolute errors, $\overline{AE}_{[O_2]}$ and $\overline{AE}_T$, for each split.

This indicates that the maximum error to be expected for the prediction will be below 2 % air and 3 °C for the trained ANN. These results are slightly higher that previously reported[2] due to the multiple splits performed here. It is not excluded that a longer training would allow even lower errors.

4. CONCLUSIONS

In this work, a sensor based on luminescence quenching and ANNs is described which can predict the oxygen concentration and temperature simultaneously. The sensor uses an ANN with MTL architecture which takes

as input the normalized phase shift of the luminescence for multiple modulation frequencies and returns as output the oxygen concentration and the temperature. The performance of the sensor is investigated in detail to demonstrate its potential for industrial applications. In particular, since the errors in the prediction measured as absolute error (AE) and mean absolute error (MAE) are influenced by the particular choice of the training and development datasets, the reproducibility of a the prediction of both parameters was evaluated by performing multiple splits, applying bootstrapping to the respective development datasets. For the resulting models, the statistical distributions of the average of the MAE was calculated, to have an indication of the average error in the prediction, indicated as $\langle MAE \rangle$. Additionally, to evaluate the worst possible performance of the sensor, the error limited accuracy (ELA) was computed and the value of the AE for which all the predictions are below this specific error ($\overline{AE}$) was determined. The results over the 6000 bootstrap samples show that the $\langle MAE \rangle$ is 0.18 % air for the oxygen prediction and 0.2 °C for the temperature predictions respectively. $\overline{AE}$ remains below 1.8 % air for the oxygen prediction and below 3 °C for the temperature. These results show how the ANN is well able to predict both the oxygen concentration and the temperature simultaneously, although the latter with slightly lower accuracy, since Pt-TFPP has a stronger dependence on oxygen than on the temperature. This demonstrates how the use of this type of sensor architecture has the potential of revolutionizing the sensor design.

APPENDIX A. PSEUDO-CODE TO PERFORM THE STATISTICAL ANALYSIS

The pseudo-code to perform the statistical analysis of the performance of the ANN model shown in 3 is described in the Algorithm 1. This describes the evaluation of average and standard deviation of $\langle MAE_{[O_2]} \rangle$ and of

Algorithm 1: Pseudo-code for the evaluation of average and standard deviation of $MAE_{[O_2]}$ and MAE_T over multiple data splits and multiple samples generated via bootstrap on the validation dataset obtained in each split. In this work the following values were used: $N_s = 6$, $N_b = 1000$ and N_{b_2}=30000.

Result: Average and standard deviation of $MAE_{[O_2]}$ and MAE_T.

Create two empty lists L_{O2} and L_T;
for $i = 1, \ldots, N_s$ **do**
  Split the data in two parts: training set $S_{T,i}$ (80% of the data) and development set $S_{dev,i}$ (20% of the data) using a random seed equal to i;
  Train the ANN on $S_{T,i}$ and save the best model obtained in 10^6 epochs;
  for $j = 1, \ldots, N_b$ **do**
    Generate a new dataset $S_{dev,i,j}$ choosing $|S_{dev,i}|$ elements from $S_{dev,i}$ *with repetitions*. $|S_{dev,i}|$ is the cardinality of the set $|S_{dev,i}|$, or in other words its size;
    Calculate $MAE_{[O_2]}(S_{dev,i,j})$ and $MAE_T(S_{dev,i,j})$;
    add $MAE_{[O_2]}(S_{dev,i,j})$ to the list L_{O2} and $MAE_T(S_{dev,i,j})$ to the list L_T;
  end
end
for $i = 1, \ldots, N_{b_2}$ **do**
  Generate two new datasets $L_{O2,i}$ and $L_{T,i}$ by choosing $|S_{dev}|$ elements from L_{O2} and L_T respectively with repetitions;
  Evaluate $\langle MAE_{[O_2]} \rangle = \frac{1}{N_{b_2}} \sum_{i=1}^{N_{b_2}} \langle MAE_{[O_2]} \rangle_i$ where $\langle MAE_{[O_2]} \rangle_i = \frac{1}{|S_{dev}|} \sum_{j=1}^{|S_{dev}|} L_{O2,i}^{[j]}$ and $L_{O2,i}^{[j]}$ is the j^{th} element in the list $L_{O2,i}^{[j]}$;
  Evaluate $\sigma(MAE_{[O_2]}) = \sum_{j=1}^{N_{b_2}} \left(\frac{\langle MAE_{[O_2]} \rangle_i - \langle MAE_{[O_2]} \rangle}{N_{b_2}} \right)^{1/2}$;
  Evaluate the equivalent quantities for the temperature;
end

$\langle MAE_T \rangle$. $MAE_{[O_2]}(S)$ indicates the mean absolute error of the oxygen concentration's prediction evaluated on a generic dataset S. $MAE_T(S)$ indicates the mean absolute error of the temperature's prediction evaluated on a generic dataset S. N_s indicates the number of splits performed, N_b and N_{b_2} the number of bootstrap samples. The average is indicated with $\langle \cdot \rangle$ and the standard deviation with $\sigma(\cdot)$. The formula for $\langle MAE_T \rangle$ and $\sigma(MAE_T)$ are not reported for simplicity but are defined similarly to those reported for the oxygen concentration by substituting $[O_2]$ with T.

REFERENCES

[1] Wang, X.-d. and Wolfbeis, O. S., "Optical methods for sensing and imaging oxygen: materials, spectroscopies and applications," *Chemical Society Reviews* **43**(10), 3666–3761 (2014).

[2] Venturini, F., Michelucci, U., and Baumgartner, M., "Dual oxygen and temperature luminescence learning sensor with parallel inference," *Sensors* **20**(17), 4886 (2020).

[3] Michelucci, U., Baumgartner, M., and Venturini, F., "Optical oxygen sensing with artificial intelligence," *Sensors* **19**(4), 777 (2019).

[4] Caruana, R., "Multitask learning," *Machine learning* **28**(1), 41–75 (1997).

[5] Baxter, J., "A model of inductive bias learning," *Journal of artificial intelligence research* **12**, 149–198 (2000).

[6] Michelucci, U. and Venturini, F., "Multi-task learning for multi-dimensional regression: application to luminescence sensing," *Applied Sciences* **9**(22), 4748 (2019).

[7] Michelucci, U., [*Applied Deep Learning - A Case-Based Approach to Understanding Deep Neural Networks*], APRESS Media, LLC (2018).

[8] Huang, J. Z., "An introduction to statistical learning: With applications in r by gareth james, trevor hastie, robert tibshirani, daniela witten," (2014).

[9] Michelucci, U. and Venturini, F., "Estimating neural network's performance with bootstrap: A tutorial," *Machine Learning and Knowledge Extraction* **3**(2), 357–373 (2021).

[10] Quaranta, M., Borisov, S. M., and Klimant, I., "Indicators for optical oxygen sensors," *Bioanalytical reviews* **4**(2-4), 115–157 (2012).

[11] Carraway, E., Demas, J. N., and DeGraff, B., "Luminescence quenching mechanism for microheterogeneous systems," *Analytical Chemistry* **63**(4), 332–336 (1991).

[12] Demas, J. N., DeGraff, B., and Xu, W., "Modeling of luminescence quenching-based sensors: comparison of multisite and nonlinear gas solubility models," *Analytical Chemistry* **67**(8), 1377–1380 (1995).

[13] Hartmann, P., Leiner, M. J., and Lippitsch, M. E., "Response characteristics of luminescent oxygen sensors," *Sensors and Actuators B: Chemical* **29**(1-3), 251–257 (1995).

[14] Kneas, K. A., Demas, J., DeGraff, B., and Periasamy, A., "Fluorescence microscopy study of heterogeneity in polymer-supported luminescence-based oxygen sensors," *Microscopy and Microanalysis* **6**(6), 551–561 (2000).

[15] Wei, Y., Jiao, Y., An, D., Li, D., Li, W., and Wei, Q., "Review of dissolved oxygen detection technology: From laboratory analysis to online intelligent detection," *Sensors* **19**(18), 3995 (2019).

[16] Chatni, M. R., Li, G., and Porterfield, D. M., "Frequency-domain fluorescence lifetime optrode system design and instrumentation without a concurrent reference light-emitting diode," *Applied optics* **48**(29), 5528–5536 (2009).

[17] Badocco, D., Mondin, A., and Pastore, P., "Determination of thermodynamic parameters from light intensity signals obtained from oxygen optical sensors," *Sensors and Actuators B: Chemical* **163**(1), 165–170 (2012).