

Extracting the Size Distribution of Gold Nanoparticles from UV-Visible Spectrum: an Artificial Neural Network Model

Jaeyoung Kim, Seyong Choi, Kiduk Kim, Kisang Byun, and Joonkyung Jang*

Department of Nanoenergy Engineering, Pusan National University, Busan 46241, Korea.

*E-mail: jkjang@pusan.ac.kr

(Received July 7, 2025; Accepted August 9, 2025)

ABSTRACT. A nondestructive detection of the size distribution of nanoparticles (NPs) is desired in various applications. The conventional method utilizes the dynamic light scattering from which the sizes of NPs are drawn by using the Stokes-Einstein relation. We propose a novel method to draw the size distribution of NPs from the UV-visible spectrum. Our method utilizes an artificial neural network (ANN) model trained against an extensive dataset generated by Mie theory. The promising performance of the present ANN model is demonstrated.

Key words: Gold nanoparticles, Size distribution, Mie theory, Artificial neural network, UV-visible spectrum

INTRODUCTION

Metallic nanoparticles (NPs) find wide applications in fuel cells, solar energy storage, growth catalysts and biomedicines, just to name a few.^{1,2} In particular, the NPs of gold have been extensively utilized in the diagnostics, treatment, catalysis, optical sensing, and biomarkers and building blocks in nanotechnology.³ Various properties of a single NP, such as the optical and electrical properties, sensitively depend on its size.^{3,4}

More often than not, it is an ensemble of NPs of different sizes that is utilized in experiment. It is therefore important to determine the particle size distribution (PSD) of NPs. An ex-situ method, such as the transmission electron⁵ or scanning electron⁶ microscopy, can be used to measure the sizes of NPs. Preferably however, a non-destructive in-situ detection of the size distribution is called for. At present, a popular in-situ method is the dynamic light scattering (DLS)⁷⁻⁹ which extracts the PSD by utilizing the Stokes-Einstein relation. DLS technique can be used to measure particle size for particles with a monodisperse size distribution. However, it becomes challenging to measure for particles with polydisperse or large-size distribution because big particles can screen smaller objects.^{8,10,11}

UV-visible spectroscopy is widely used in the study of nanomaterials as a diagnostic of nanoparticle formation. It is a very useful tool employed for the characterization, estimation of the sizes of NPs, concentration, and aggregation level. UV-visible spectrometers can be found in many laboratories, the analysis does not change the sample, and the time needed for registration of the spectrum is very

short.^{4,12-15} Several attempts have been made to derive the PSD from a UV-visible extinction spectrum. Kim et al. assumed that the size distribution of gold NPs follows a Gaussian distribution and demonstrated that PSD can be derived solely from UV-visible spectra under this assumption. Calculations considering PSD were more experimentally accurate than calculations that ignored size distribution.³ Haiss et al. demonstrated the predictability of gold NPs sizes using the mean-free-path-corrected Mie theory.⁴ Doak et al. also utilized Mie theory to predict the size distribution of gold NPs from UV-visible spectra. In this study gold NPs were considered to have a broad size distribution, which resulted in better agreement with experimental data compared to theoretical calculations that only considered a single particle size.² Karlik et al. showed the multilayered perceptron is efficient for the prediction of the diameter of gold nanospheres. The principal component dimension reduction technique was used during the preprocessing of UV-visible spectra. The training set consisted of 25 spectra calculated according to the Mie theory for nanospheres with diameters from 2 to 50 nm. The approach was validated for theoretical spectra of particles with 5, 7, 10, 15, and 30 nm diameter.¹³

Herein, we propose a novel alternative method to extract the PSD from the UV-visible extinction spectrum. The present method employs an artificial neural network (ANN) model of the PSD for a colloidal gold NPs. By utilizing the extensive dataset of UV-visible spectra generated from the exact Mie theory, we trained the ANN model to predict the PSD solely from the UV-visible spectra. The present ANN model is favorably tested against a finite set of experi-

mental spectra collected from the literature.

COMPUTATIONAL METHODS

Preparation of the UV-Visible Spectra for Training the ANN Model

In order to train the present ANN model, we prepared the UV-visible spectra of gold NPs whose PSDs are known. We generated the discretized distribution of the radii of NPs, $P(r)$, which ranged from $r = 1$ to 200 nm with an interval of 5 nm (40 nodes in the output layer). $P(r)$ was sampled from the random numbers obeying a Gaussian distribution with a mean μ and standard deviation σ . The mean radius of NP, μ , was varied from 5 nm to 150 nm with an interval of 1 nm. For a given μ , we have chosen various σ s ranging from 0.05 μ to 0.125 μ , in order to consider diverse PSDs. To each Gaussian $P(r)$ generated as above, we added a small random noise uniform sampled from the range between $-0.2P(r)$ and $0.2P(r)$. This addition of random noises was done to mimic the small noises in the experimental size distribution, and the other was to double the number of data for training the present ANN model.

For a given $P(r)$ generated as above, the extinction spectrum $Q_{ext}(\lambda)$ at a specific wavelength of λ was calculated as

$$Q_{ext}(\lambda) = \Delta r \sum_{i=1}^N P(r_i) Q_{ext}(\lambda, r_i), \quad (1)$$

where $Q_{ext}(\lambda, r_i)$ is the extinction coefficient for a specific radius of NP. We discretized the range of radii by using the bin of $\Delta r = 0.5$ nm and $N = 200$. λ was varied from 400 to 700 nm with an increment 1 nm, in order to include the peak of a typical extinction spectrum (located between 500 and 600 nm). The absorbance, A , is related to $Q_{ext}(\lambda)$ through the following relation,

$$A = \frac{\pi r^2 Q_{ext} l N}{\ln 10}, \quad (2)$$

where the path length of light l varies from equipment to equipment and it is difficult to know exactly the number density N of the samples.² To solve these, the value of spectrum at each wavelength $Q_{ext}(\lambda)$ was divided by the difference between the maximum and minimum of the spectrum.

The size-dependent extinction coefficient $Q_{ext}(\lambda, r)$ of a gold NP was calculated by using Mie theory,¹⁶

$$Q_{ext}(r, \lambda) = \frac{1}{x^2} \sum_{n=1}^{\infty} (2n+1) \text{Re}(a_n + b_n), \quad (3)$$

where λ is the wavelength of an incident light. a_n and b_n

are the scattering coefficients given by

$$a_n = \frac{m \psi_n(mx) \psi_n'(x) - \psi_n(x) \psi_n'(mx)}{m \psi_n(mx) \zeta_n'(x) - \zeta_n(x) \psi_n'(mx)} \quad (4)$$

and

$$b_n = \frac{\psi_n(mx) \psi_n'(x) - m \psi_n(x) \psi_n'(mx)}{\psi_n(mx) \zeta_n'(x) - m \zeta_n(x) \psi_n'(mx)} \quad (5)$$

In equations (4) and (5), we introduced a dimensionless size parameter $x = 2\pi n_m r / \lambda$. Also, ψ_n and ζ_n are the Ricatti-Bessel functions given by

$$\psi_n(x) = \left(\frac{x\pi}{2}\right)^{\frac{1}{2}} J_{n+1/2}(x) \quad (6)$$

and

$$\zeta_n(x) = \left(\frac{x\pi}{2}\right)^{\frac{1}{2}} H_{j+1/2}^{(1)}(x) \quad (7)$$

where $J_{n+1/2}(x)$ and $H_{j+1/2}^{(1)}(x)$ are the Bessel j and primary Hankel functions, respectively.¹⁷ m is the relative refraction coefficient defines as $m = n_p / n_m$ where n_m and n_p the real refractive index of the solvent and complex refractive index of the NP, respectively.

We took the solvent to be water, and it was assumed that the refractive index of the solvent did not depend on the wavelength of light, therefore using $n_m = 1.333$. The refractive index of gold, taken to be dependent on the frequency of an incident light, is taken from the data reported by Johnson and Christy.¹⁸ We used a cubic spline interpolation¹⁹ to calculate the values of refractive indices intermediate between the discrete raw data reported by Johnson and Christy. We generated total of 1,752 spectra by using an in-house Python code.

Artificial Neural Network Model

We divided the dataset of extinction spectra into the training, validation, and testing sets by randomly taking 60, 20, and 20% of 1,752 spectra in total. We used the validation set to determine the presence of an overfitting. The present ANN model, illustrated in Fig. 1, is made of three hidden layers. The input layer consists of 301 nodes each of which contains the discretized spectrum of the solution of gold NPs where the wavelength of an incident light ranges from 400 to 700 nm with a 1 nm interval. Each hidden layer consists of 256 nodes. The output is discretized PSD of nanoparticles radii, with each node corresponding to the fraction of particles falling into a specific size bin. We used the ReLu function for an activation function, except for the output layer where the identity function was used.

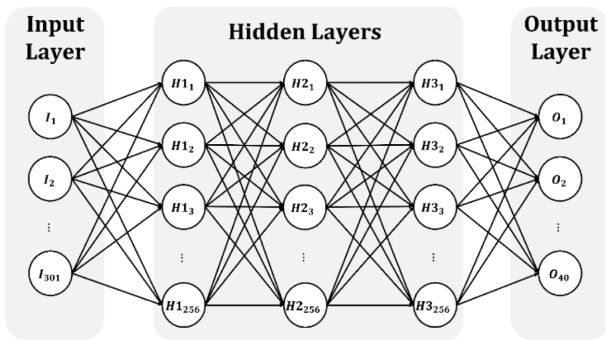


Figure 1. Scheme for the present ANN model to extract the size distribution of NPs from an input extinction spectrum. Each of 301 nodes in the input layer contains the extinction cross-section value for the wavelength of an incident light. Each of 3 hidden layers consists of 256 nodes. Each of 40 nodes in the output layer gives discretized PSD of gold NPs.

We trained the present ANN model by using the back-propagation algorithm and found the optimized values of weights. We employed an Adam optimizer with a learning-rate of 10^{-6} to minimize cost function defined as the mean-squared error.²⁰ In order to prevent overfitting, we applied a dropout method where the nodes in the hidden layers were dropped out with a rate of 0.7. The present ANN model was trained in 200,000 steps, and the cost values were saved for every 10,000 steps.

We calculated the R^2 value, root-mean-square error (RMSE), and mean-absolute error (MAE) as the performance indicators of the ANN model. R^2 was defined as $R^2 = \sum_{i=1}^n (\hat{d}_i - \bar{d}) / \sum_{i=1}^n (d_i - \bar{d})^2$, where i and n represent the index and the total number of data points, respectively. $\hat{d}_i$ and d_i are the i th value of PSD, $P(r_i)$, from the ANN model and the Gaussian distribution with a random noise, respectively. $\bar{d}$ is the mean value of d_i ($= \frac{1}{n} \sum_{i=1}^n d_i$).

The MAE and RMSE were defined as $MAE = \frac{1}{n} \sum_{i=1}^n |d_i - \hat{d}_i|$

and $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - \hat{d}_i)^2}$. All the ANN methods were implemented by using TensorFlow package.²¹

Experimental Data

We further tested the present ANN model against 41 experimental data. The experimental UV-visible spectra were obtained from the nanoComposix²² and Cytodiagnosics²³ web pages, and from the paper reported by López-Muñoz, G.A., et al.²⁴ After that, the graphic images of the experimental spectra were converted to digital numbers by using Engauge Digitizer.²⁵ Some experimental data are

given size distributions as mean and standard deviation of diameters, mean diameters were calculated from the predicted size distributions to compare the ANN model with experimental value.

RESULTS AND DISCUSSION

Fig. 2 shows the variation in the cost function as the training progresses in 200,000 steps. If the error of prediction using the training set is larger than the error of the prediction using the validation set, the learning is stopped. In this study, no learning interruption occurred. The blue line and orange line show the training and validation cost respectively. Training and validation costs decreased from 75.9811 and 75.3383 at 10,000th step to 13.2096 and 8.3396 at 200,000th step, respectively. Since the training cost decreased smoothly during training, it can be seen training of ANN finished well. In addition, the validation cost decreased as smoothly as training cost, and showed lower values during training, so it can be considered that there was no overfitting.

As illustrated in Fig. 3, the present ANN model trained well against the theoretical calculation. The ANN model gave the R^2 values of 0.7949 and 0.7945 for the training and test sets, respectively (Table 1). The nearly identical values of R^2 for training and test sets show the absence of overfitting. The RMSEs of the ANN model were 2.810 and 2.7639 for the training and test sets, respectively (Table 1). The MAEs were 0.5708 and 0.5740 for the training and test sets, respectively. However, discrepancies were observed at high fraction values. These higher fractions typically occurred in sharp PSDs and corresponded to smaller particle sizes (~ 30 nm), since the standard deviation σ was expressed as a ratio to the mean diameter during PSD generation. The discrepancies in these regions indicate lower

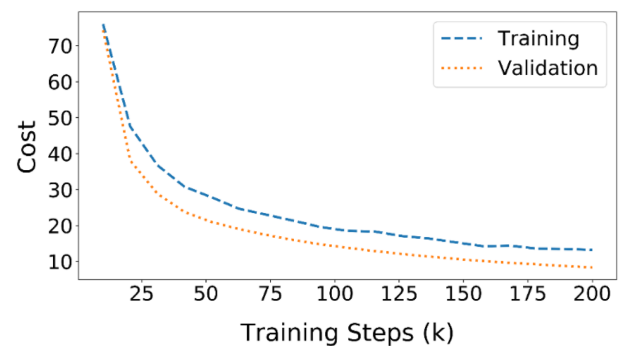


Figure 2. Change in the cost function with the progression of training. Drawn as the broken and dotted lines are the variations in cost for training and validation cost, respectively.

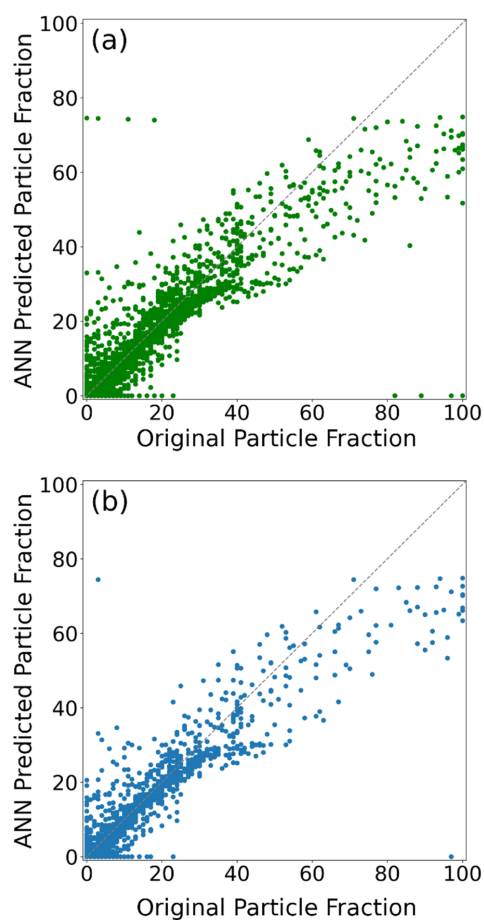


Figure 3. Performance of the present ANN model vs. the original data for the training (a) and test (b) datasets. Each data point indicates the fraction of particles falling into a specific size bin. The training and test datasets consisted of 1050 and 351 data, respectively. As a reference, the dotted lines represent the exact match between the ANN and theoretical values.

Table 1. Training results of the present artificial neural network model

	RMSE	MAE	R2 Value
Train	2.801	0.5708	0.7949
Test	2.7639	0.5740	0.7945
Train-Test	0.0371	0.0032	0.0004

prediction accuracy for smaller particles, which may be attributed to minimal variation in the extinction spectra within this small size range. This problem can be solved by making the intervals of size distribution smaller than 5 nm in a small size section, but in very small cases the prediction may not be as good.

We investigated whether the ANN model predicts well in the experiment. The RMSE and MAE were 7.944 nm

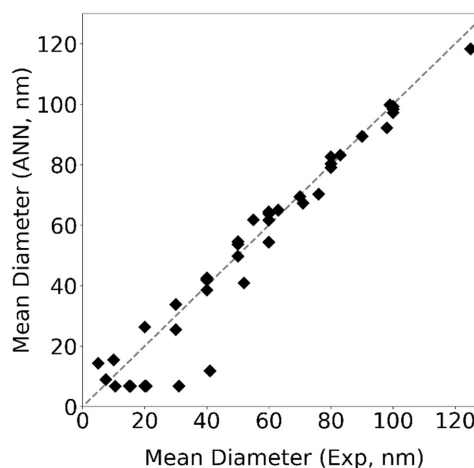


Figure 4. Mean diameters of experimental NPs plotted vs. those predicted from the ANN model, drawn as diamonds. As a reference, the dotted line represents the exact match between the ANN and theoretical values. The R^2 value for the mean diameter is 0.932.

Table 2. RMSE, MAE, R2 values between the calculated mean diameters based on predicted size distribution and the mean diameters from experiment

	RMSE (nm)	MAE (nm)	R2 Value
Total	7.944	5.280	0.932
Size < 45 nm	11.277	8.366	0.393
Size > 45 nm	3.602	2.865	0.960

and 5.280 nm, respectively. Table 2 shows error and correlation value between the mean diameter from experiment and the ANN model, and Fig. 4 shows the correlation between them. In Fig. 4, x and y axis represent the mean diameters from experiment and the ANN model, respectively. Overall, they show high correlation, and especially when the average size is 45 nm or more it can be seen that the prediction is very good. However, in the case below size 45 nm it has relatively high RMSE and MAE values although high correlation coefficient.

Table 2 shows the predicted average size by ANN and absolute error and relative error between predicted and experimental values. It can be considered as follows that the prediction results in the section below the size of 45 nm are not good. Note the present ANN model showed lower prediction accuracy for smaller particle sizes. Additionally, in the previous reported study by Haiss et al., when the size was small (below 25 nm), the peak position of the experimental surface plasmon resonance (SPR) was smaller than the theoretical calculation.⁴ Since the ANN model used in this study predicts the size distribution of gold

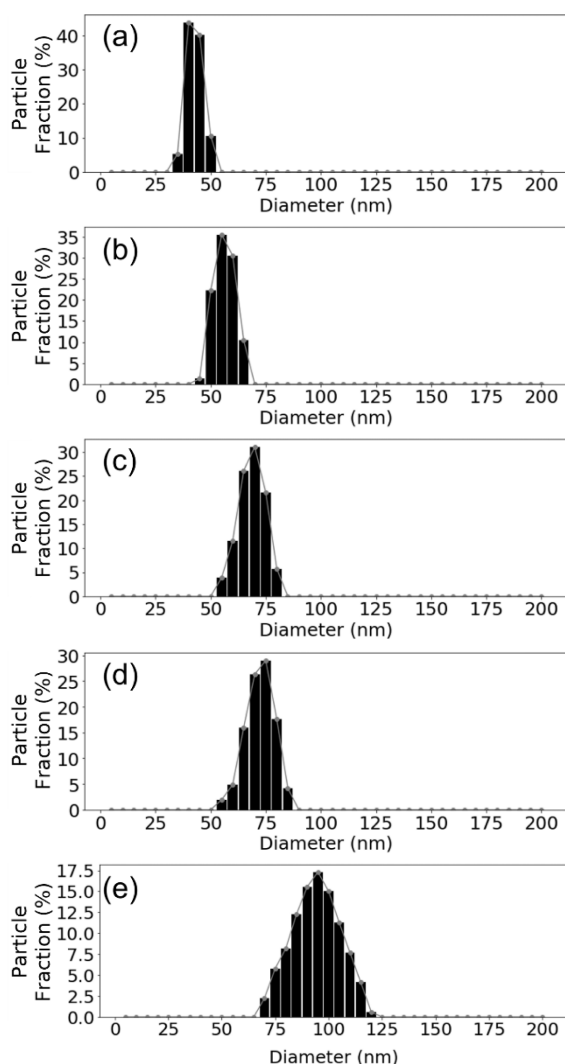


Figure 5. Examples of predicted PSD using ANN model which matches well with experimental data. The experimentally measured mean diameter and standard deviation for each sample were 52 nm and 5 nm(a), 60 nm and 6 nm(b), 71 nm and 8 nm(c), 80 nm and 5 nm(d), and 98 nm and 11 nm(e).

nanoparticles using only theoretical UV-visible extinction spectra, the difference in the shape between experimental and theoretical spectra can be considered as the reason. It seems that this can be solved by training the experimental spectrum in the case of small sizes.

CONCLUSION

In this study, Artificial neural network (ANN) that predicts size distributions of colloidal gold nanoparticles using UV-visible extinction spectrum is introduced. Extinction spectra of colloidal gold nanoparticles which considering size distribution were calculated by using Mie theory. ANN

was trained using calculated extinction spectra as input factors and size distribution of gold nanoparticles as output factors. For comparison with the experiment, prediction size distributions using 41 experimental data were attempted. The average sizes were calculated from the predicted size distribution and compared with experimentally measured values. Overall, it showed high correlation, especially when the average size was more than 45 nm. However, in the case of size smaller than 45 nm, they showed a high error value although the high correlation coefficient. This can be seen since small particles shows differences in the shape of UV-visible extinction spectrum of theory and experiment, and the size distributions were gathered at 5 nm intervals. If the prediction in small sized section is successfully done, it will lead to the development of a non-destructive measurement method that effectively predicts the particle size using UV-visible extinction spectrum.

Acknowledgments. This work was supported by a 2-Year Research Grant of Pusan National University

REFERENCES

1. De Vyt, A.; Gijbels, R.; Davock, H.; Van Roost, C.; Geuens, I. *Journal of Analytical Atomic Spectrometry* **1999**, *14*, 499.
2. Doak, J.; Gupta, R. K.; Manivannan, K.; Ghosh, K.; Kahol, P. K. *Physica E: Low-dimensional Systems and Nanostructures* **2010**, *42*, 1605.
3. Kim, H.; Jang, J.; Leton Chandra, S. *Journal of the Korean Chemical Society* **2007**, *51*, 407.
4. Haiss, W.; Thanh, N.T.K.; Aveyard, J.; Fernig, D. G. *Anal. Chem.* **2007**, *79*, 4215.
5. Reimer, L. *Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*; Springer: 2013.
6. Reimer, L. *Measurement Science and Technology* **2000**, *11*, 1826.
7. Berne, B. J.; Pecora, R. *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics*; Courier Corporation: 2000.
8. Zanetti-Ramos, B. G.; Fritzen-Garcia, M. B.; de Oliveira, C. S.; Pasa, A. A.; Soldi, V.; Borsali, R.; Creczynski-Pasa, T. B. *Materials Science and Engineering: C* **2009**, *29*, 638.
9. Jans, H.; Liu, X.; Austin, L.; Maes, G.; Huo, Q. *Analytical Chemistry* **2009**, *81*, 9425.
10. Khlebtsov, B. N.; Khlebtsov, N. G. *Colloid Journal* **2011**, *73*, 118.
11. Tomaszewska, E.; Soliwoda, K.; Kadziola, K.; Tkacz-Szczesna, B.; Celichowski, G.; Cichowski, M.; Szmaja, W.; Grobelny, J. *Journal of Nanomaterials* **2013**, *1*, 13081.
12. Martínez, J. C.; Chequer, N. A.; González, J. L.; Cordova, T. *Nanosci. Nanotechnol* **2012**, *2*, 184.

13. Karlik, B.; Yilmaz, M. F.; Ozdemir, M.; Yavuz, C. T.; Danisman, Y. *Plasmonics* **2021**, *16*, 147.
14. Amendola, V.; Meneghetti, M. *The Journal of Physical Chemistry C* **2009**, *113*, 4277.
15. Oldenburg, S. J.; Averitt, R. D.; Westcott, S. L.; Halas, N. J. *Chemical Physics Letters* **1998**, *288*, 243.
16. Bohren, C. F.; Huffman, D. R. *Absorption and Scattering of Light by Small Particles*; John Wiley & Sons: 2008.
17. Abramowitz, M.; Stegun, I. A. *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*; US Government printing office: 1964.
18. Johnson, P. B.; Christy, R.-W. *Physical Review B* **1972**, *6*, 4370.
19. McKinley, S.; Levine, M. *College of the Redwoods* **1998**, *45*, 1049.
20. Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. *Nature* **1986**, *232*, 533.
21. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mane, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viegas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. *arXiv preprint arXiv:1603.04467* **2016**.
22. nanoComposix, <https://nanocomposix.com/>.
23. Cytodiagnosics, <https://cytodiagnosics.com>.
24. López-Muñoz, G. A.; Pescador-Rojas, J. A.; Ortega-Lopez, J.; Salazar, J. S.; Balderas-López, J. A. *J. Nanoscale Research Letters*. **2012**, *7*, 667.
25. Mitchell, M.; Muftakhidinov, B.; Winchen, T.; Jędrzejewski-Szmek, J.; Badger G. Zenodo. *Engauge Digitizer Software* 2016.