

Machine learning for determining the size distribution of metallic and pharmaceutical submicron particles based on scattered light extinction and indicatrices

© V.V. Rovenko, T.A. Semenov, E.O. Epifanov, G.V. Mishakov, N.V. Minaev,
A.A. Garmatina, E.I. Mareev

Kurchatov Complex of Crystallography and Photonics NRC „Kurchatov Institute“, Moscow, Russia
E-mail: rovenko.vladimir@physics.msu.ru

Received January 12, 2026

Revised April 2, 2026

Accepted April 7, 2026

A method for determining the size distribution of submicron particles (from 1 to 4000 nm) is proposed, based on training a convolutional neural network using scattered light indicatrices and extinctions. The training was carried out on optical response matrices synthesized within the Mie theory for metallic (gold) and pharmacological (ibuprofen) particles, taking into account their complex refractive indices. The proposed neural network model allows for reconstructing the parameters of the lognormal particle size distribution in less than 1 s on a graphics processor, providing an average discrepancy with scanning electron microscopy data at the level of 30 %, which is comparable to the accuracy of manual analysis with a significant reduction in time costs.

Keywords: Mie theory, submicron particles, light scattering, extinction, scattering indicatrix, machine learning.

DOI: 10.61011/TPL.2026.07.63832.20621

Rapid and accurate determination of submicron particle size distribution is a critical task for nanotechnology, materials science, pharmacology and environmental monitoring [1]. In particular, size control of metallic (e.g., gold) nanoparticles is necessary to ensure their functionality in catalysis, sensorics, and biomedicine [2], and in the case of pharmacological submicron particles (such as ibuprofen), size directly affects the bioavailability, solubility, and efficacy of drugs [3]. Although electron microscopy remains one of the main methods of analysis, it is associated with labor-intensive sample preparation and significant time costs. A promising alternative is light scattering-based methods, which allow for non-destructive and rapid *in situ* analysis. However, extracting particle size information from experimentally measured scattered light phase functions and extinctions is a complex inverse problem. Conventional approach based on a „manual“ selection of the direct problem solution is very time-consuming, especially for polydisperse systems. It should also be noted that the development of modern data collection techniques makes it possible to achieve a spectral data collection rate of 100–1000 Hz, which makes it impossible to process them in „manual“ mode. In this view, the development of autonomous determination of the nano- and microparticles sizes using optical scattering data, including neural network approaches, is actively being developed [4–7]. For example, in [5,6] the possibility of determining the size of monodisperse nanoparticles from extinction spectra using neural networks was demonstrated, however, the problem of reconstructing the complete particle size distribution in polydisperse systems is somewhat less studied. In [7] a method for estimating distribution parameters based on dynamic light scattering data is proposed, but its

application is limited to the submicron range and requires a priori information about the distribution shape. Paper [8] describes a neural network method for reverse engineering of arbitrary scatterers based on spherical harmonic expansion, which allows predicting multipole coefficients and structural parameters of single particles with given electromagnetic properties. This method is focused on the design of meta-atoms and does not solve the problem of determining the size distribution in polydisperse assemblies. The given study delves into these methods by proposing a hybrid technique that combines rigorous Mie theory-based calculations with deep learning to reconstruct lognormal particle size distribution parameters from both extinction spectra and scattering indicatrices. Extinction spectra are sensitive to particle size over a wide range of wavelengths, and scattering indicatrices provide information on scattering anisotropy, which significantly expands the technical capabilities for monitoring the production of nano- and microparticles.

The method was tested in applied problems related to the generation of gold nanoparticles by laser ablation and the formation of ibuprofen particles by method of rapid expansion of supercritical solutions (a description of the experimental setup is given in [9]). To automate the particle sizing process, a hybrid method was developed that combined rigorous calculations based on Mie theory with machine learning algorithms. In the first stage, the newly developed software is used to solve the direct problem of light scattering and generate extensive databases linking the particle distribution parameters with their optical response. For gold particles, we used dispersion relations from paper [10], the refractive index values for ibuprofen were measured by us on a refractometer DR-M2/1550 (Atago).

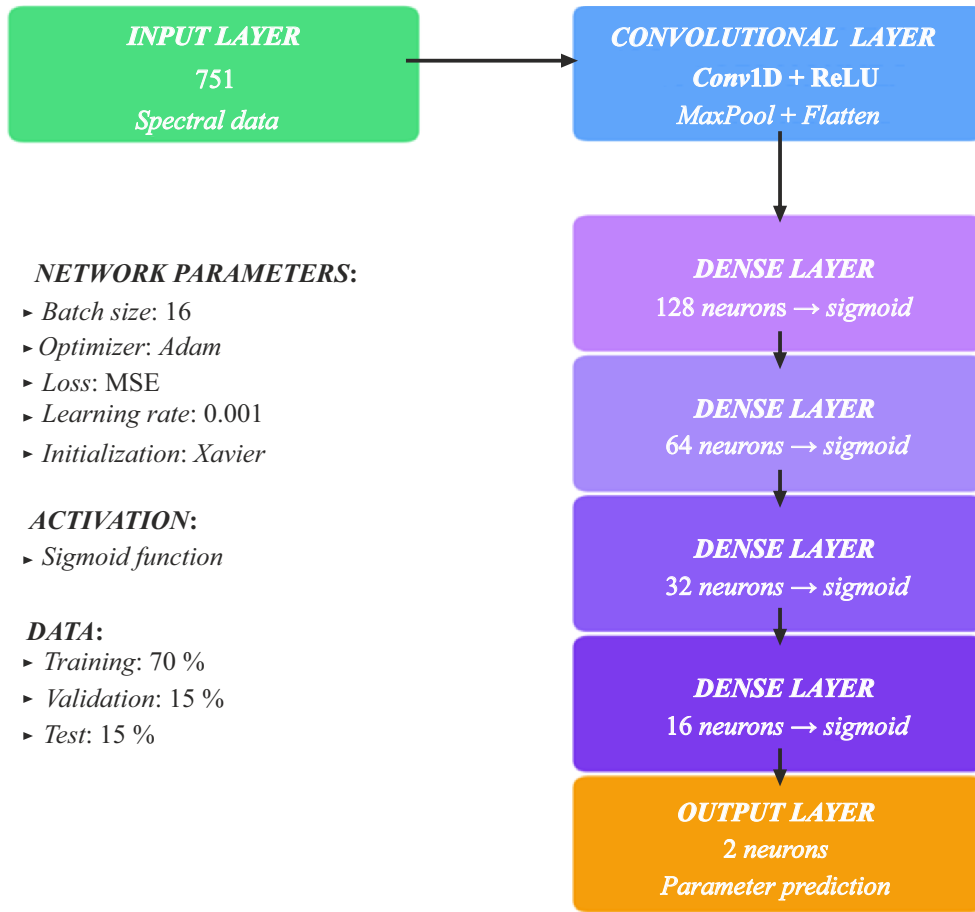


Figure 1. Architecture of neural network.

At the second stage, the neural network trained on this data solves the inverse problem — it reconstructs the particle size distribution based on experimentally measured scattering indicatrices or extinction spectra. The key advantage of the approach lies in the ability of the neural network model to approximate complex nonlinear dependencies, which provides its significant superiority over classical minimization methods based on linear approximation. Thus, the proposed approach expands the possibilities for precise control of nanoparticle sizes in various technological processes.

The software package performs database calculations based on Mie theory for spherical particles. The calculations consider classical problem of scattering of a planar monochromatic electromagnetic wave with wavelength λ_0 by a spherical particle of radius a with complex refractive index $m = p + ik$, located in a non-absorbing medium with refractive index n_m [11,12]. A dimensionless parameter of x size is used:

$$x = k_m a = \frac{2\pi n_m}{\lambda_0} a, \quad (1)$$

where k_m — is the wavenumber in the surrounding environment. For an ensemble of particles with a given size distribution $p(a)$, the optical response matrix is calculated as a convolution of the Mie problem solution for monodisperse

particles. It is assumed that the system is rarefied and interparticle interaction can be neglected, which is similar to the experimental conditions (low concentration of particles in the liquid or gas phase):

$$M_{base}(z) = \sum_a T_{scat}(x, z) \cdot p(x). \quad (2)$$

For the extinction spectra z is itself a radiation wavelength $z = \lambda$, and $T_{scat}(x, z) = Q_{ext}(x, \lambda)$ — is the extinction efficiency factor. For scattering indicatrices, z represents the scattering angle $z = \theta$, whereas $T_{scat}(x, z) = S(x, \theta)$ is the scattering amplitude function. For comparison with the real experiment, the relative magnitude of forward and backward scattering $Ratio(\theta) = M_{for}(\theta)/M_{back}(\pi - \theta)$ was calculated. As $p(a)$ the lognormal distribution was chosen, which has proven itself well for describing such systems [9,13] and is justified by the physics of laser ablation processes and rapid expansion of supercritical solutions:

$$p(a) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln a - \mu)^2}{2\sigma^2}\right), \quad a > 0,$$

$$\mu = \ln\left(\frac{M^2}{\sqrt{V + M^2}}\right), \quad \sigma = \sqrt{\ln\left(\frac{V}{M^2} + 1\right)}. \quad (3)$$

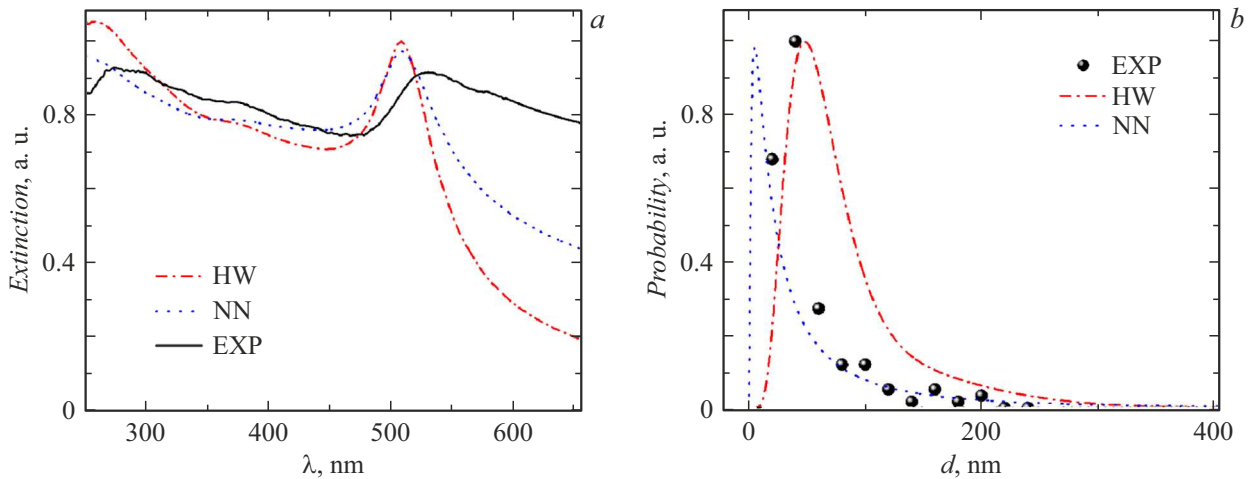


Figure 2. The particle size distributions for a sample of gold nanoparticles obtained by scanning electron microscopy (EXP), „handwork“ (HW) selection from optical data, and predicted by a neural network (NN), compared. *a* — experimental and simulated extinction spectra; *b* — probability particle size distributions. The observed discrepancy is about 30% and is primarily due to the deviation of the particle shape from spherical one.

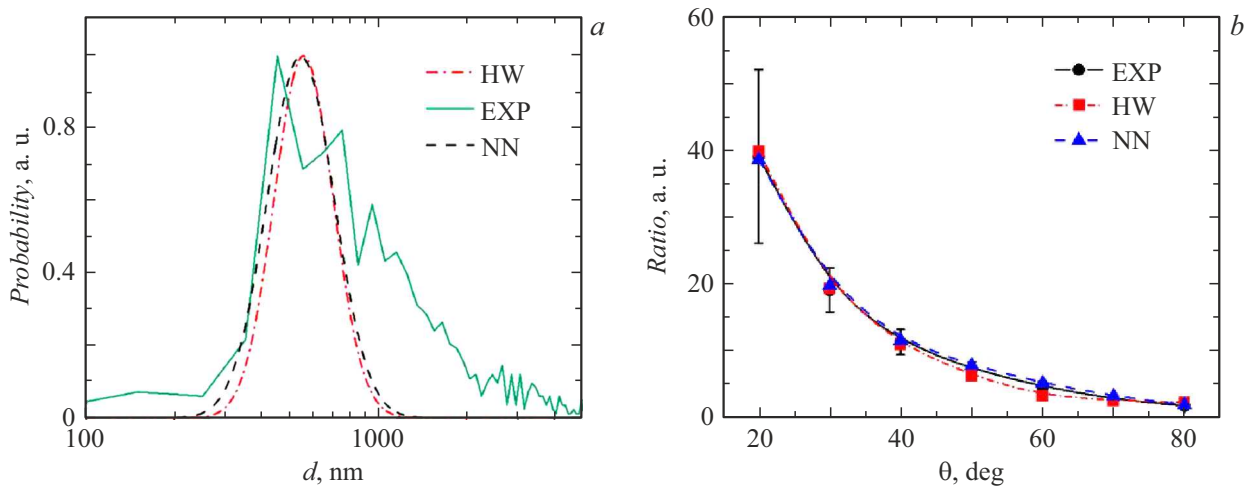


Figure 3. The particle size distributions for a sample of ibuprofen particles (RESS method) obtained by scanning electron microscopy (EXP), handwork selection (HW) and predicted by a neural network (NN), compared. *a* — probability particle size distributions; *b* — experimental and model scattering indicatrices. The accuracy of reconstruction is comparable to Fig. 2.

The parameters μ and σ are related through an average arithmetic size M and variance V . The database formation algorithm consists of the following stages. First, a parallel calculation of $T_{scat}(x, z)$ is performed for all specified parameters and the obtained results are saved. The obtained results are then used to calculate the optical response matrices using formula (2). To augment the data, Gaussian noise is added to the obtained results to increase the model's robustness to experimental errors. As a result, an extensive database is formed that establishes a correspondence between the parameters of the log-normal distribution (μ , σ) and the synthesized optical response matrices. For each material (gold, ibuprofen), a separate database was generated taking into account its optical constants.

To solve the inverse problem of determining the particle distribution parameters (μ , σ) based on light scattering

data, a convolutional neural network was developed and trained with its architecture shown in Fig. 1. The model deployed using the PyTorch library solves the task of multidimensional regression [14]. Its purpose is to predict the parameters of the lognormal distribution (μ , σ) from the input extinction spectrum or scattering indicatrix. The network architecture includes one convolution layer and four fully connected layers with sizes of 128, 64, 32 and 16 neurons. The output layer contains 2 neurons (μ and σ). Further increasing the number of layers does not result in higher model accuracy; reducing the number of fully connected layers or abandoning the use of a convolutional layer leads to a significant (more than twofold) drop in the quality of reconstruction. Training was performed on Nvidia RTX 3070 graphics accelerator using Adam (learning rate = 0.001) optimizer and MSE loss function.

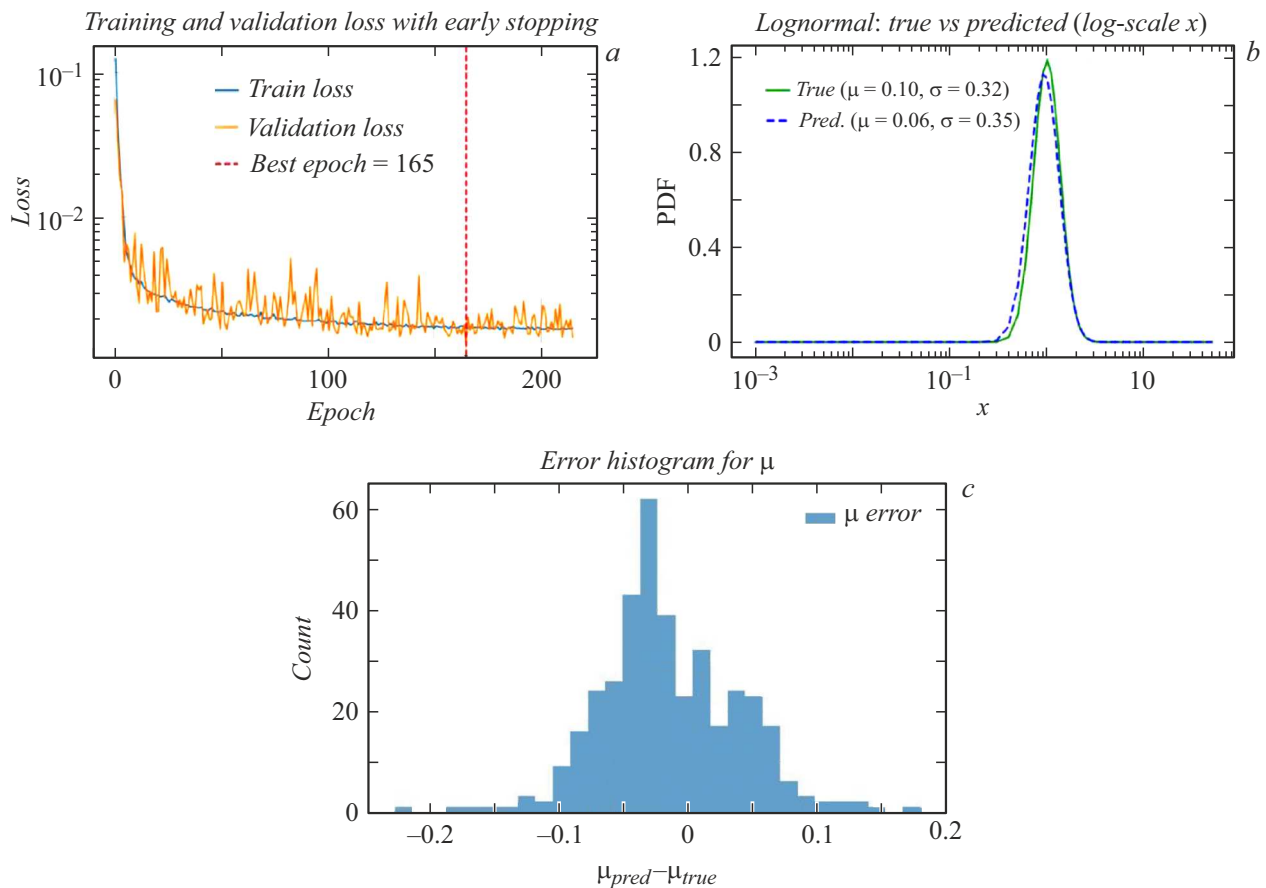


Figure 4. Loss function values for the training and validation sets and comparison of the predicted distribution with the model one. *a* — losses for the training and validation sets depending on the number of epochs with early stopping; *b* — probability distributions of particle sizes; *c* — error histogram for parameter μ .

The batch size was 64. No significant influence of batch size and training speed on the final result was found. The input data are pre-normalized and represented as one-dimensional vectors of fixed length. For the extinction spectra: 751 points which stands for the wavelength range of 250–1000 nm with a 1 nm step. Experimental indicatrices were measured at eight angular positions and interpolated to 200 points for feeding into the neural network; indicatrices calculated with high angular resolution (1°) were used in the training set.

The learning convergence was controlled by the dynamics of the loss function on the training and validation samples with the training stop function. To validate the method, databases containing 100,000 unique distributions for each type of experiment were generated. Figures 2, *b* and 3, *a* show a comparison of the distributions obtained using scanning electron microscopy (SEM) with those reconstructed by classical „manual“ selection and the proposed neural network model for samples of gold nanoparticles (average size ~ 80 nm, obtained by laser ablation in water) and ibuprofen particles (average size ~ 200 nm, obtained by the RESS method). Figures 2, *a* and 3, *b* show the result of modeling the extinction spectrum for gold and the scattering indicatrix for ibuprofen based on neural network data on the

distribution of particles, as well as its comparison with the experiment. The average discrepancy between the predicted and reference distributions was about 30 %, which is primarily explained by deviation of the shape of real particles from the spherical one used in Mie theory. The loss function graph (Fig. 4, *a*) shows that it took about 165 epochs to reach convergence. The neuronnetwork model demonstrated low error of parameters prediction ($MSE \sim 2.5 \cdot 10^{-3}$) on the test sample (Fig. 4, *a, b*), the accuracy of sizes reconstruction (within 3 %, Fig. 4, *b, c*) weakly depends on the particles sizes (within 1 nm–1.5 μ m). Additional testing on synthetic monodisperse distributions showed that the network correctly reconstructs the size with an error of less than 5% for particles in the specified sensitivity range.

Thus, the study provides a universal, accurate and automated approach for determining the size distribution of submicron particles in real time. The developed complex method that combines rigorous calculations based on Mie theory with neural network data analysis, enables to overcome the key limitations of traditional methods, such as time-consuming „handwork“ selection and linear approximation.

The major benefit of this method is the ability of the convolutional neural network to identify complex nonlinear relationships between the optical response and distribution parameters, which ensures high prediction accuracy based on synthetic data ($\text{MSE} \sim 2.5 \cdot 10^{-3}$) and fast convergence (about 165 epochs). Testing the method on real experimental data proved that this method is valid in practice. A 30% observed discrepancy with SEM data is natural and is primarily associated with deviation of the shape of the studied particles from spherical, which is a fundamental limitation of the optical model used, and not of the neural network approach itself. The obtained results demonstrate the potential of the method for effective particle size control in various technological processes.

Acknowledgments

The particle sizes were studied using the equipment of the the Shared Center for „Structural Diagnostics of Materials“ at the Kurchatov Complex of Crystallography and Photonics of the National Research Center „Kurchatov Institute“.

Funding

The study on reconstruction of nanoparticle sizes was supported by grant № 23-79-10188 provided by the Russian Science Foundation.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] A.J. Hickey, C.-C. Huang, A. Misra, V. Balasubramanian, B.M. Moudgil, A.D. Rosato, L.M. Tavares, Q. Zhang, W.-N. Wang, *KONA Powder Part. J.*, **42**, 2026016 (2025). DOI: 10.14356/kona.2026016
- [2] S. Chakraborty, S. Mal, A. Halder, S. Das, K.K. Sen, A. Mukherjee, P. Roy, *Part. Sci. Technol.*, **42** (1), 145 (2024). DOI: 10.1080/02726351.2023.2211954
- [3] S. Jacob, F.S. Kather, S.H. Boddu, M. Attimarad, A.B. Nair, *Pharmaceutics*, **17** (1), 136 (2025). DOI: 10.3390/pharmaceutics17010136
- [4] T. Klinavicius, N. Khinevich, A. Tamuleviciene, L. Vidal, S. Tamulevicius, T. Tamulevicius, *J. Phys. Chem. C*, **128** (23), 9662 (2024). DOI: 10.1021/acs.jpcc.4c02459
- [5] A. Otsuki, H. Jang, *ChemEng.*, **6** (6), 92 (2022). DOI: 10.3390/chemengineering6060092
- [6] Y. Ren, J. Mao, H. Zhao, C. Zhou, X. Gong, Z. Rao, Q. Wang, Y. Zhang, *Adv. Meteorol.*, **2020**, 5074192 (2020). DOI: 10.1155/2020/5074192
- [7] M.A. Habitzreuter, M.C. Barbosa, *Phys. Rev. E*, **107** (3), 034601 (2023). DOI: 10.1103/PhysRevE.107.034601
- [8] J. Li, Z. Zhang, Y. Xu, S. Fu, J. Yang, Y. Wang, A.S. Shalin, Y. Qin, *Phys. Rev. Appl.*, **23** (1), 014001 (2025). DOI: 10.1103/PhysRevApplied.23.014001
- [9] E.O. Epifanov, N.M. Asharchuk, D.N. Hmelenin, I.V. Trofimov, A.O. Rybaltovskij, G.V. Mishakov, T.A. Semenov, N.V. Minaev, E.I. Mareev, *Russ. J. Phys. Chem. B*, **18** (8), 1948 (2024). DOI: 10.1134/S1990793124030047
- [10] P.B. Johnson, R.W. Christy, *Phys. Rev. B*, **6** (12), 4370 (1972). DOI: 10.1103/PhysRevB.6.4370
- [11] L.B. Scaffardi, J.O. Tocho, *Nanotechnology*, **17**, 1309 (2006). DOI: 10.1088/0957-4484/17/5/024
- [12] C.F. Bohren, D.R. Huffman, *Absorption and scattering of light by small particle* (Wiley-VCH, 1983), p. 102, 103.
- [13] H. Bagheri, H. Hashemipour, V. Ghalandari, S. Ghader, *Particuology*, **57**, 201 (2021). DOI: 10.1016/j.partic.2020.12.011
- [14] S.S.A. Zaidi, M.S. Ansari, A. Aslam, N. Kanwal, M. Asghar, B. Lee, *Comput. Biol. Med.*, **126**, 103514 (2022). DOI: 10.48550/arXiv.2104.11892

Translated by T.Zorina