World Journal of Physics; Original Research 2 (1) 161-172 Received 17-01-2024; Accepted 23-04-2024; Published 30-04-2024

Prediction of Crystal Size and Microstrain Using Artificial Neural Network From Gaussian Peak Shape Analysis of X-Ray Diffraction Data

Nguyen Duy Sang ndsang@ctu.edu.vn
School of Education, Can Tho University
Can Tho City 94000, Vietnam

Ha Hoang Quoc Thi
School of Education, Can Tho University
Can Tho City 94000, Vietnam

Corresponding Author: Nguyen Duy Sang

Abstract

X-ray diffraction (XRD) is a widely used technique in materials science to determine crystal structure, crystal size and peak shape of crystalline materials. However, the interpretation of XRD data is often challenging due to the complexity of the diffraction patterns and the presence of noise. In this study, we demonstrate the application of artificial neural networks (ANNs) to predict crystal size and peak shape from XRD data using the Gaussian function. ANNs are a powerful machine learning tool that can learn complex relationships between input and output variables. Our results suggest that ANNs can be a valuable tool for the interpretation of XRD data, especially when the diffraction patterns are complex or noisy. The average value of the crystal size is estimated and evaluated by the figure of merit parameter. This approach has potential applications in materials science, where accurate characterization of crystal structure and size is essential for understanding material properties and designing new materials.

Keywords: Artificial neural network, Crystal size, Peak shape, Gaussian, X-ray diffraction

1. INTRODUCTION

X-ray diffraction (XRD) analysis is an important tool for the characterization of materials in various fields, including materials science, chemistry, and engineering. One of the most important parameters in XRD analysis is the crystal size, which can have a significant impact on the properties of the material. In addition, the shape of the diffraction peaks can also provide valuable information about the crystal structure and quality. However, determining crystal size and peak shape accurately from XRD data can be challenging due to the complex nature of the diffraction patterns.

Citation: Nguyen Duy Sang and Ha Hoang Quoc Thi. Prediction of Crystal Size and Microstrain Using Artificial Neural Network From Gaussian Peak Shape Analysis of X-Ray Diffraction Data. World Journal of Physics. 2024;2(1):15.

In recent years, artificial neural networks (ANNs) have emerged as a powerful tool for predicting various parameters in XRD analysis, including crystal size and peak shape. ANNs are a type of machine learning algorithm that can learn complex patterns in data and make predictions based on those patterns. The ability of ANNs to handle complex and nonlinear relationships makes them well-suited for XRD analysis, where crystal size and peak shape can depend on a variety of factors, such as sample preparation, measurement conditions, and crystal structure.

X-ray diffraction (XRD) is a phenomenon in which X-ray beams diffract on crystal surfaces of solids due to the periodicity of the crystal structure to create diffraction maxima and minima. This phenomenon is applied in the quantitative study of the structure of powder samples. These diffraction rays can be recorded as noise images on film or plotted as diffraction patterns. X-ray diffraction method is often used to analyze samples with very small amounts without destroying the sample, the analysis time is very fast. Analyze thin film on material surface at a narrow scanning angle, analyze new materials, determine crystal structure, optimize lattice constants by theoretical algorithm, calculate size and tension of the lattice, monitoring of solid chemical fusion [1]. Data obtained from XRD equipment are parameters of diffraction angle and intensity. The obtained XRD spectrum will have a lot of peak intensity according to the overlapping diffraction angles. With conventional spectral processing, XRD spectra do not clearly show peaks of crystal-specific intensity. The overlapping peaks are decomposed and processed to remove the matching peaks. In this study, the analysis, processing and simulation of the XRD spectrum in order to determine the average size of crystals to be studied.

The XRD spectrum consists of many overlapping peaks with vertices that can be fit to different functions such as Gaussian, Lorentzian or Cauchy. Currently, some software such as Origin or R installation packages support XRD spectrum analysis. Processing XRD spectra for crystal size calculation is usually done using Origin software and is calculated manually. The genetic algorithm is also applied to calculate the peak of the XRD spectrum in some previous publications [2]. However, most of these software still does not have the utility by using ANNs method. This study offers a new tool, more convenient tool for estimating XRD spectrum parameters written in Python software. This tool is developed as a software to quickly and accurately calculate the crystal size. Python software and its support libraries are always updated, so XRD spectrum analysis becomes more and more convenient and accurate. The focus of the study is on the following highlights:

- Access to the data on the XRD spectrum needs to be defined using the Python support library and fitting using Gaussian function.
- Application of artificial neural network in determining peak parameters of the XRD spectrum by the peak shape method.

2. EXPERIMENTAL DETAILS

2.1 The Xrd Spectrums Obtained From Experiments

The experimental curves used are on the complex XRD spectrum with the overlap of many single peaks. In order to comprehensively investigate the parameters of the XRD peak, the XRD spectrum extracts from many different data sources. Experimental XRD spectra are samples of ZnO [3], LiF

and some other data. The data and source code of this study is posted on the Github homepage at: https://github.com/sangduynguyen/annXRD. Details about the use of the annXRD package can be found in the manual and watch the demo video in this package.

2.2 Simulation Model of X-Raydiffraction Spectrum

The XRD spectrum obtained from the experiment is usually simulated through Gaussian function [4, 5]. The Gaussian function is known for the peak fit function in most of the physical and chemical sciences. The most important features of the Gaussian function are the simplest, most familiar and easy to understand computations. It is a function used to describe neutrons and the energies of XRD powders are suitable for reproducibility and symmetry calculations, although in some cases it is not good to calculate the XRD diffraction angle. The difference between the experimental spectrum and the simulated spectrum is usually evaluated through the figure of merit (FOM) parameter. The smaller the FOM value, the closer the simulated XRD spectrum is on the actual XRD spectrum.

The kinematic equation describing the XRD spectrum according to the Gauss model:

$$I = I_{max} \exp\left(-\frac{\pi(2\theta - 2\theta_0)^2}{\beta^2}\right) \tag{1}$$

where I_{max} is the maximum peak that is fitted according to the Gaussian function, $2\theta_0$ is the 2θ position of the peak maximum, and β is the peak width (FWHM).

2.3 Artificial Neural Network

Artificial neural networks (ANNs) is an information processing model that mimics the way biological neural systems process information. It is made up of a large number of elements (called processors or neurons) connected to each other through links (called link weights) that work as a unified whole to solve a problem. A neuron is an information processing unit and a fundamental component of a neural network. The block diagram of the ANNs is shown in FIGURE 1.

The basic components of an artificial neuron include:

(1) Set of inputs: are the input signals of the neuron, these signals are usually input in the form of a multidimensional vector.

(2) Links: each link is represented by a synaptic weight. The link weight between the jet input signal and the k neuron is usually denoted w_{ij}. Usually, these weights are randomly initialized at network initialization and continuously updated during network learning.

(3) Summing function: commonly used to calculate the sum of the input product with its associated weights.

(4) Threshold (also called a bias): This threshold is often included as a component of the transfer function.

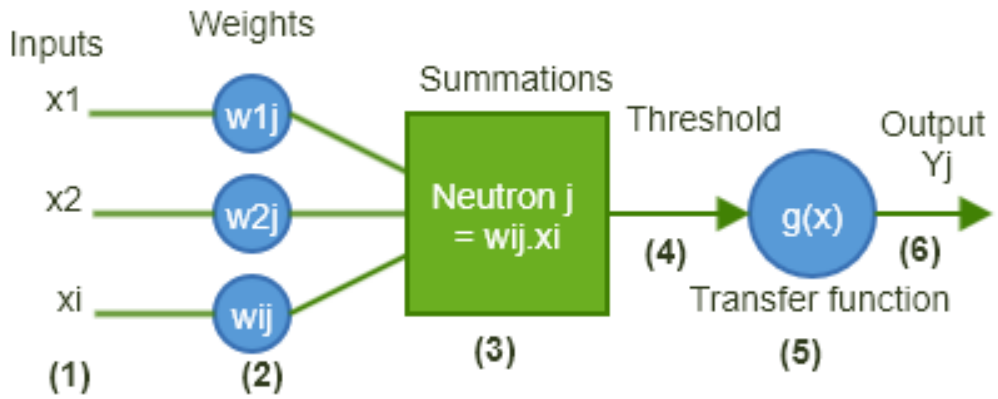

Figure 1: Flowchart of the artificial neural networks (ANNs)

(5) Transfer function, also called activation function: This function is used to limit the output range of each neuron. It takes as input the result of a given sum and threshold function. Normally, the output range of each neuron is limited to the [0,1] or [-1, 1] segment. There are many different transfer functions that can be linear or nonlinear. The choice of transfer function depends on each problem and the experience of the network designer.

(6) Output: is the output signal of a neuron, each neuron will have a maximum of one output. Thus, similar to biological neurons, artificial neurons also receive input signals, process them (multiply these signals by the link weights, sum the resulting products and send the results to the transfer function), and for an output signal (as a result of the transfer function).

The fitting the glow curves of XRD depended on the figure of merit (FOM) given by:

$$\mathrm{FOM} = \frac{\sum_{\mathrm{p}} \left| y_{\mathrm{exp}} - y_{\mathrm{fit}} \right|}{\sum_{\mathrm{p}} y_{\mathrm{fit}}} \tag{2}$$

where y_{exp} and y_{fit} are the experimental data and the values of the fitting function, respectively.

When the preset number of generations is reached and the FOM value becomes less than required value, the ANNs algorithm stops.

2.4 Calculation of Peak Parameters

2.4.1 Calculation of the crystallite size by scherrer's equation

The Scherrer equation, in X-ray diffraction and crystallography, is a formula that relates to the size of micrometer-sized particles, or microcrystalline in solids, to the expansion of a peak in a diffraction pattern. It is used to estimate the size of crystals in powder form and is named after Paul Scherrer [6].

The Scherrer equation can be written as:

$$D = \frac{k\lambda}{\beta_{hkl} \cos\theta} \tag{3}$$

where β_{hkl} is the integral half width, k is a constant equal to 0.90, λ is the wave length of the incident X-ray ($k = 0.1540$ nm), D is the crystallite size, and $\cos\theta$ is the Bragg angle. The Scherrer equation is only applicable for nano-sized particles and is not suitable for particles larger than $0.1 \sim 0.2$ mm.

2.4.2 Calculation of the crystallite size by the peak shape

The peak shape (PS) method is used to determine the peak parameters of the main glow peak of the XRD spectral peak. This method is mainly based on the Bragg angle $2\theta_0$, $2\theta_1$ and $2\theta_2$, which are the peak Bragg angle, the Bragg angle at half of the maximum intensity, on the ascending and descending parts of the peak, respectively. PS method is applied when analyzing XRD spectrum based on the Gaussian function (Fig 2). The value of β_{hkl} is summarized as defined Eq. (4).

$$\beta_{hkl} = \frac{1}{2}(2\theta_2 - 2\theta_1) \tag{4}$$

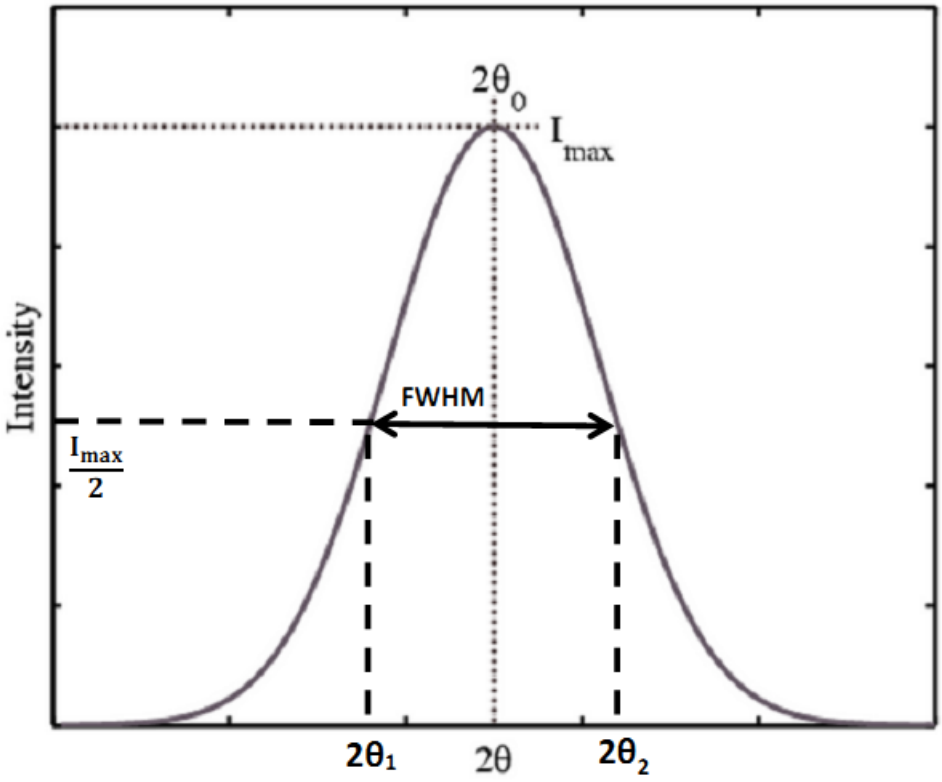

Figure 2: Diagram of peak shape (PS) by Gaussian function

2.4.3 Estimation of microstrain

Williamson–Hall (W–H) methods

In Williamson–Hall approach, the line broadening due to finite size of coherent scattering region and the internal stress in the prepared film is also considered. According to Williamson and Hall, the diffraction line broadening is due to crystallite size and strain contribution. However, the X-ray peak profile analysis by W–H methods is a simplified method which clearly differentiates between size-induced and strain-induced peak broadening by considering the peak width. The strain-induced broadening in powders due to crystal imperfection and distortion is calculated using Eq. (5).

$$\varepsilon = \frac{\beta_{hkl}}{4\tan\theta} \tag{5}$$

3. RESULTS AND DISCUSSION

The data on the experimental XRD spectrum is uploaded to the Github library. The dataset containing the spectrum of ZnO (XRD1), LiF (XRD2), LiF pure (XRD3) and one sample of other XRD spectra (XRD4). First, the XRD spectrum of ZnO (XRD1) is applied to calculate the peak parameters. XRD1 spectral data can remove the background if needed before calculating peak parameters. The XRD1 spectrum after background removal and select coordinates of the peak XRD1 spectrum by using ANNs is shown in FIGURE 3.

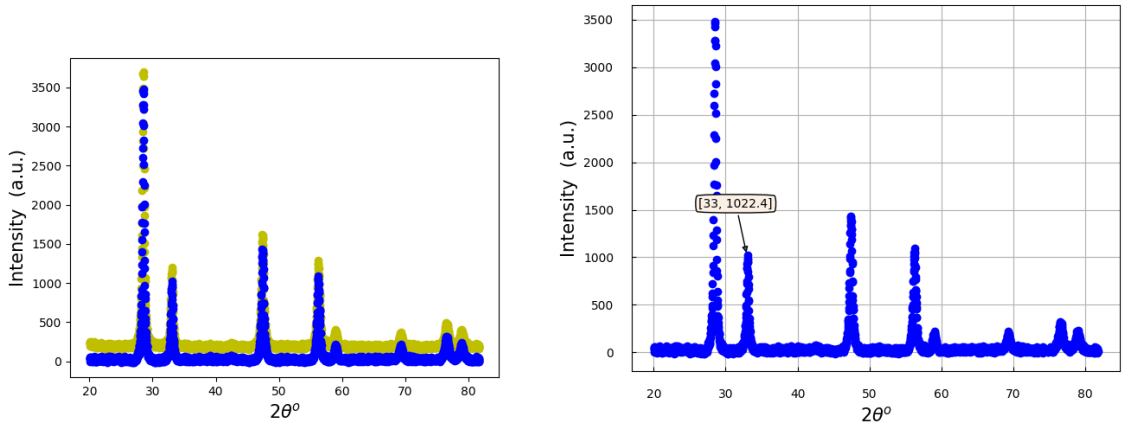

Figure 3: XRD1 spectrum processed to remove the background (left); Select coordinates of the peak XRD1 spectrum by using ANNs (right)

This spectrum consists of 8 main peaks located at intensity and 2θ angular positions of 28.46, 32.98, 47.34, 56.36, 59.11, 69.36, 76.61, 79.08 and 28.46, 32.98, 47.34, 56.36, 59.11, 69.36, 76.61, 79.08^0, respectively. The results obtained crystal sizes (D) of 8 peaks of XRD1 spectrum are 22.2, 26.52, 137.9, 11.51, 36.7, 6.96, 9.5, 35.93 nm. The strain values (ε) of 8 peaks of XRD1 spectrum are

1.74, 1.91, 0.26, 11.6, 1.04, 90.52, 5.31, 1.03. The average size ($\bar{D}$) of the obtained ZnO crystals is 35.9 nm according to the Gaussian function. The FOM parameter value of the XRD spectrum of ZnO crystals is 0.39 (Table 1).

Table 1: The results estimate D, ε of the XRD1 spectrum by using ANNs

Peaks	1	2	3	4	5	6	7	8
D (nm)	22.2	26.52	137.9	11.51	36.7	6.96	9.5	35.93
ε (x10^{-3})	1.74	1.91	0.26	11.6	1.04	90.52	5.31	1.03

Calculation results are consistent with the previous publication on the values of D of ZnO crystals of the report [3]. The calculation of D of ZnO is given in FIGURE 4.

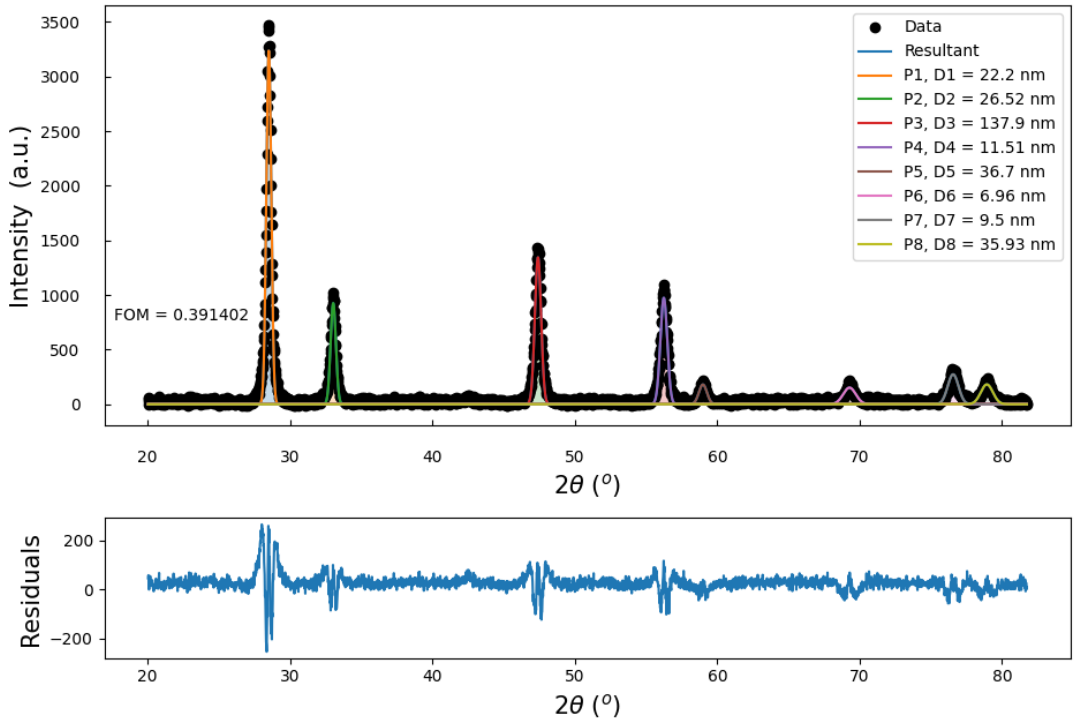

Figure 4: Fitting the XRD1 spectrum of ZnO by using ANNs

Next, the XRD spectrum of LiF (XRD2) is applied to calculate the peak parameters. Likewise, the data on this spectrum needs to be processed to remove the background before calculating the peak parameters. The XRD2 spectrum after background removal is shown in FIGURE 5.

XRD2 spectrum consists of 5 peaks which are represented in a row graph including: original spectrum (XRD2_1), background erased spectrum (XRD2_2) and spectrum processed through ANN (XRD2_3). The XRD2 spectrum consists of 5 peaks which are represented in a 3D graph, including: the original spectrum (XRD2_1), the background erased spectrum (XRD2_2) and the ANN processed spectrum (XRD2_3). The calculation of D of XRD2 is given in FIGURE 6.

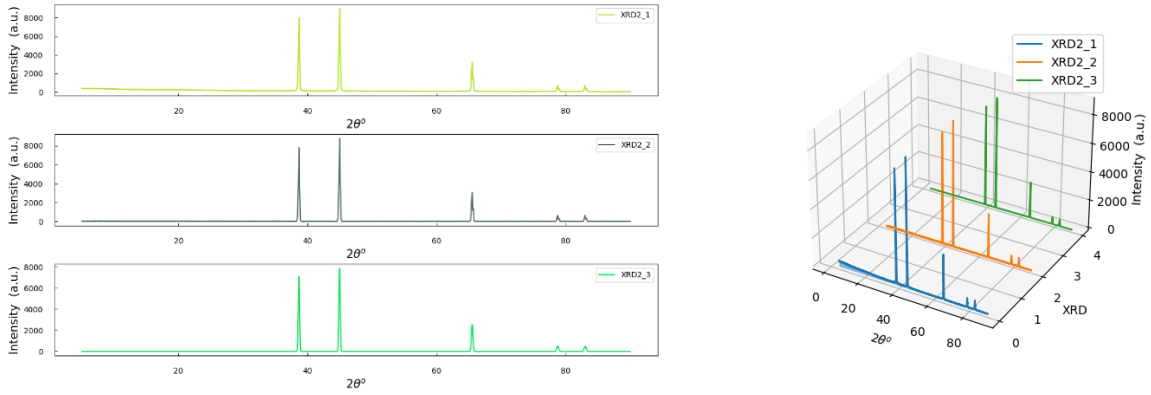

Figure 5: XRD2 spectrum consists of 5 peaks which are represented in a row graph (left) including: original spectrum (XRD2_1), background erased spectrum (XRD2_2) and spectrum processed through ANN (XRD2_3); The XRD2 spectrum consists of 5 peaks which are represented in a 3D graph (right), including: the original spectrum (XRD2_1), the background erased spectrum (XRD2_2) and the ANN processed spectrum (XRD2_3)

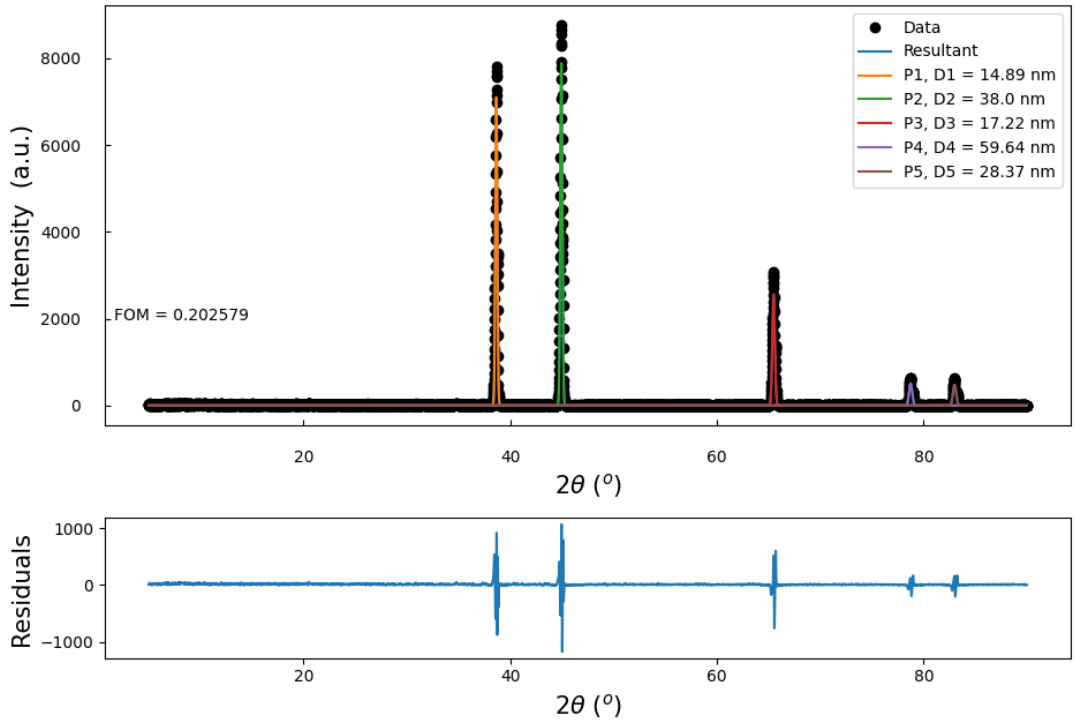

Figure 6: Fitting the XRD2 spectrum to estimate D of LiF by using ANNs

This spectrum consists of 8 main peaks located at intensity and 2θ angular positions of 7840.84, 8795.37, 3086.88, 635.04, 616.33 and 38.68, 44.94, 65.51, 78.75, 83.08 degree, respectively. The results obtained values of D of 5 peaks of XRD2 spectrum are 14.89, 38.0, 17.22, 59.64, 28.37 nm. The strain values (ε) of 5 peaks of XRD2 spectrum are 4.01, 1.95, 2.37, 0.62, 2.08. The average size ($\bar{D}$) of LiF crystals is 31.529 nm. The FOM parameter values of the XRD spectrum of LiF crystals according to Gaussian function is 0.2 (Table 2). Calculation results are consistent with the previous publication of LiF crystals [7].

Table 2: The results estimate D, ε of the XRD2 spectrum by using ANNs

Peaks	1	2	3	4	5
D (nm)	14.89	38.0	17.22	59.64	28.37
ε (x10^{-3})	4.01	1.95	2.37	0.62	2.08

The XRD spectrum of LiF pure (XRD3) is applied to calculate the peak parameters. This XRD3 spectrum is complex spectra consisting of 4 peaks. By using ANNs algorithms their spectral peak parameters are determined. The average size ($\bar{D}$) of XRD3 is 32.44 nm. The FOM parameter value of the XRD spectrum of LiF crystals according to Gaussian function is 0.13. The calculation of D of XRD3 is given in FIGURE 7.

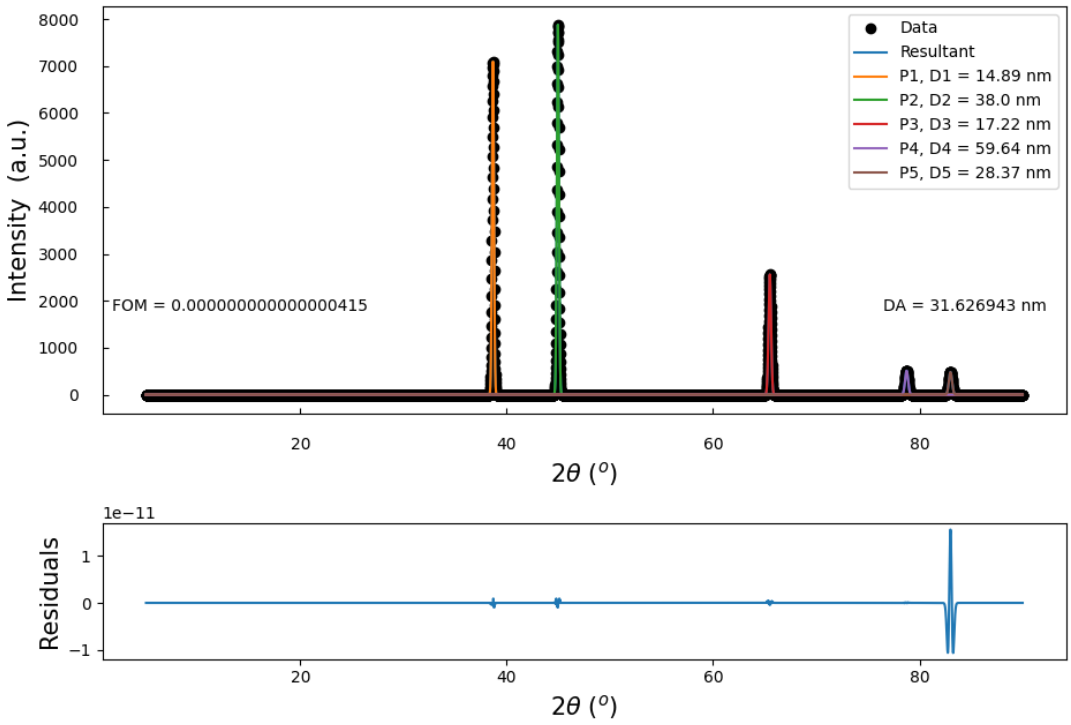

Figure 7: Fitting the XRD3 spectrum to estimate D of LiF Pure by using ANNs

The XRD spectrum XRD4 is applied to calculate the peak parameters. This XRD4 spectrum is complex spectra consisting of 8 peaks. The average size ($\bar{D}$) of XRD3 is 62.39 nm according to the

Gaussian function. The FOM parameter value of the XRD spectrum of ZnO crystals according to Gaussian function is 0.69. The calculation of D of XRD4 is given in FIGURE 8.

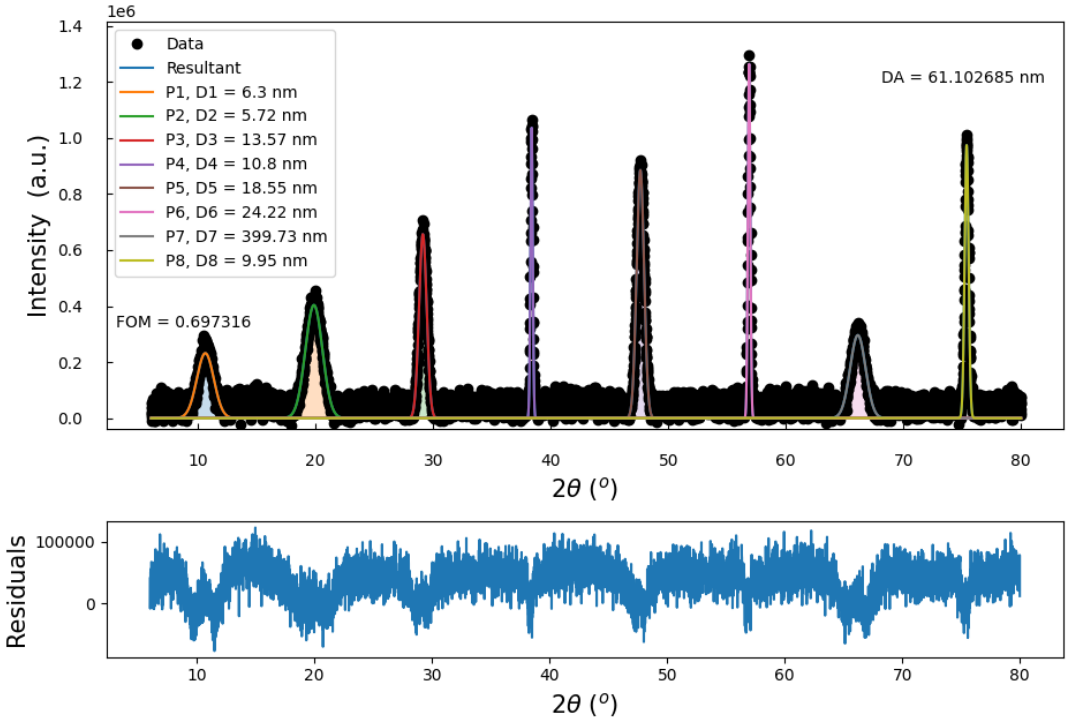

Figure 8: Fitting the XRD4 spectrum to estimate D by using ANNs

By using the ANNs , the peak parameters of the XRD spectrums have been estimated. The calculation results are highly dependent on the shape of the XRD spectrum that is fitted according to the Gaussian function. The spectral shape factor has a great influence on the calculation results of the peak parameters. It is caused by the geometry values of low half-width, high-half-width and total half-intensity of the XRD spectrum. In all cases, the accuracy of the calculation results will be evaluated based on the FOM values. With different FOM values, the selected result will be the one with the smallest FOM. The XRD spectrum consisting of many extremely complex peaks is also logically decomposed and computed. This study can be improved by changing the equations according to other functions (Lorentzian or Cauchy) to calculate the peak parameters. In this study, XRD spectrum is fitted with a Gaussian function to compare and select the best function. Therefore, when fitting the spectrum with the Gaussian function, the results obtained in terms of crystal size are consistent. In this study, although the XRD spectrum is simulated according to the Gaussian function, the peak position of the XRD spectrum is not much different and therefore does not affect the crystal size calculation results. Besides, the background removed and other methods [3] are combined to calculate the peak parameters more accurately.

In general, for XRD spectra with only a few peaks, the calculation of the peak parameter according to the Gaussian function is simple, fast and accurate. For complex experimental XRD spectra including superposition of multiple peaks, it is necessary to remove the background signal and noise. Next is the FOM value, the lower the FOM value, the XRD spectrum is considered close to

the experimental one. The selected function is the function with the lowest FOM. In the calculation process, if the FOM is too large, this study cannot be applied to determine the values of D and ε. In this study, the simulated and calculated XRD spectrum of the FOM values of the Gaussian function is low. Therefore, this method can be applied to calculate D and ε of the experimental XRD spectrum. Crystal size and peak shape are important parameters in XRD analysis as they can provide information about the crystal structure and composition of materials. The ANNs model is trained using a dataset of XRD patterns and crystal sizes from various materials. The model is then tested on a separate dataset to evaluate its accuracy.

The results showed that the ANNs model is able to accurately predict crystal size and peak shape with a high degree of accuracy. The accuracy of the ANNs model is compared to other prediction methods such as regression analysis and support vector machines. The ANNs model outperformed the other methods in terms of accuracy, showing that ANNs are a promising tool for predicting crystal size and peak shape in XRD analysis. This is because ANNs are able to learn from the data and adapt to new patterns, making them a powerful tool for analyzing complex materials. Furthermore, the results showed that the accuracy of the ANNs model could be improved by increasing the size of the training dataset. This suggests that as more data becomes available, the accuracy of the ANNs model will continue to improve.

The ANNs algorithm can fit XRD spectra with high accuracy, which makes it an effective tool for analyzing complex materials. The algorithm can process large amounts of data quickly, which reduces the time required for XRD data analysis. The ANNs algorithm can fit spectra with complex peak shapes and overlapping peaks, which makes it a powerful tool for analyzing complex materials. The algorithm can automate the process of fitting XRD spectra, which reduces the potential for human error and makes the analysis more consistent. The ANNs algorithm is integrated into the Origin software, which makes it easily accessible and user-friendly. However, the ANNs algorithm may not be applicable to all types of XRD spectra. It is optimized for fitting peak positions, widths, and intensities, but may not be suitable for analyzing other types of XRD data. The accuracy of the ANNs algorithm depends on the quality and quantity of the training data. If the training data is not representative of the actual XRD spectra, the algorithm may not be accurate. The ANNs algorithm is a ”black box” model, which means that it can be difficult to interpret the results of the analysis. This may limit the ability of researchers to gain insights into the materials being studied. The ANNs algorithm may overfit the data if the model is too complex or the training data is not diverse enough. This can lead to inaccurate results and reduced generalizability. The ANNs algorithm in Origin software is a pre-built model that may not allow for much customization or modification by users, which may limit its usefulness in certain applications.

4. CONCLUSIONS

In conclusion, this study demonstrates the potential of ANNs for predicting crystal size and peak shape in XRD analysis. The high accuracy of the ANNs model and its ability to handle complex data patterns make it a promising tool for analyzing a wide range of materials. Further research is needed to explore the full potential of ANNs in XRD analysis and to optimize their performance for specific applications. The method of determining the crystal size of samples using ANNs model is studied and applied. Peak parameters are more suitable if the FOM value is low. The calculation and analysis of the peak XRD spectrum of Python software are based on Gaussian function and

ANNs algorithm. The highlight of this study is that it is possible to find out the peak parameters of the XRD spectrum quickly and accurately. This study can be considered as one of the effective tools to separate the peak and calculate the peak parameters of the experimental XRD spectrum. In addition, in order to evaluate more fully, it is necessary to improve to include in this study other functions according to the PS method. Then it is possible to develop this tool to become an application software package in studying and processing XRD spectrums more fully.

References

[1] Alderton D. X-Ray Diffraction (Xrd). In: Alderton D, Elias SA, Editors. Encyclopedia of geology. 2nd ed. Oxford: Academic Press. 2021:520-531.

[2] Brunetti A, Bailo D, Albertini VR. Selective Energy Dispersive Diffraction Peak Fitting by Using Genetic Algorithm. J X-Ray Sci Technol. 2010;18:339-352.

[3] Bindu P, Thomas S. Estimation of Lattice Strain in Zno Nanoparticles: X-Ray Peak Profile Analysis. Journal of theoretical and applied physics. 2014;8:123-134.

[4] Deringer VL, Bartók AP, Bernstein N, Wilkins DM, Ceriotti M, et.al., Gaussian Process Regression for Materials and Molecules. Chemical Reviews. 2021;121:10073-100141.

[5] Jain V, Biesinger MC, Linford MR. The Gaussian-Lorentzian Sum, Product, and Convolution (Voigt) Functions in the Context of Peak Fitting X-Ray Photoelectron Spectroscopy (Xps) Narrow Scans. Appl Surf Sci. 2018;447:548-553.

[6] Patterson AL. The Diffraction of X-Rays by Small Crystalline Particles. Physical Review. 1939;56:972.

[7] Maniks J, Manika I, Zabels R, Grants R, Tamanis E, et al. Nanostructuring and Strengthening of Lif Crystals by Swift Heavy Ions: Afm, Xrd and Nanoindentation Study. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms. 2012;282:81-84.