Using Support Vector Machines to Enhance the Performance of X-Ray Diffraction Data Analysis in Crystalline Materials Cubic Structure Identification

Mohammad Syukur

Faculty of Mathematics and Natural Sciences, University of Sumatera Utara, 20155 Medan, Sumut, Indonesia

Summary
Crystalline materials cubic structure identification is very important in crystallography and material science research. For a long time researchers in the field have used manual approach in matching the result data from X-Ray Diffraction (XRD) method with the known fingerprint. These manual matching processes are complicated and sometimes are tedious because the diffracted data are complex and may have more than one fingerprint inside. This paper proposes the use of support vector machines to enhance the performance of the matching process between the diffracted data of crystalline material and the fingerprints. It is demonstrated, through experiments, that support vector machines gives more accurate and reliable identification results compared to the use of neural network.

Key words:
Support Vector Machine, X-Ray Diffraction Data Analysis, Cubic Structure Identification, Material Sciences, Artificial Intelligence Application.

1. Background

Crystallography is one of the areas of research in physics that deals with the scientific study of crystals. It has always been one of the most challenging research fields since eighteenth century. The significant discovery of X-ray by Röntgen in 1895 [8] had yielded to a new way of doing crystallography research. Since then X-ray diffraction method has been proposed and applied to many different sub-area of crystallography such as identification of crystalline phases, qualitative and quantitative analysis of mixtures and minor constituents, distinction between crystalline and amorphous states, side-chain packing of protein structures, identification of crystalline material, etc. The last area is the focus of this paper.

X-ray diffraction data interpretation for most crystalline materials is a very complex and difficult task. This is due to the condition that different crystalline material may contain more than one cubic structures component type and after being diffracted using X-ray diffraction method, the diffracted data are complex. Hence, the data can be very ambiguous and is not easy to track and understand.

Numerous artificial intelligence techniques and application have been applied and developed to solve the problems in various domains. In our previous work [5], an attempted has been made to use neural network to perform automatic cubic structure identification on the crystalline materials. Though it was a success, there's still left room for improvement. This paper proposes the use of support vector machine (SVM) to enhance the performance of crystalline materials cubic structure identification. The result of using SVM is compared with the result using neural network.

Support vector machines have been proven to be a powerful method to solve identification and classification problems. It includes gene identification [10], paraphrase identification [11], protein classification using X-ray crystallography [9], and many more. The main intent of this paper is to showcase the superior results on the use of support vector machine over neural network in crystalline materials cubic structure identification.

2. Cubic Structure Identification using X-Ray Diffraction Data

In principle, there are four cubic structures type for crystalline materials, the Simple Cubic (SC), Body Centered Cubic (BCC), Face Centered Cubic (FCC) and Diamond [1]. In our previous work [2], a formula has been proposed to calculate the fingerprints for these four cubic structures. The formula utilizes the Miller index (h,k,l) [6]. The proposed formula can be written as follows:

$$\sin^2\theta = \frac{\lambda^2(h^2+k^2+l^2)}{4a^2} \quad (1)$$

Since the wavelength of the incoming X-ray (λ) and lattice constant (a) are both constants, we can eliminate these

Manuscript received July 5, 2007
Manuscript revised July 25, 2007

quantities from Eq. 1 and derive the ratio of two $sin^2\theta$ as follows:

$$R_{m,n} = \frac{\sin^2\theta_m}{\sin^2\theta_n} = \frac{h_m^2 + k_m^2 + l_m^2}{h_n^2 + k_n^2 + l_n^2} \quad (2)$$

where θ_m and θ_n are the diffracting angles for two peak associated with the diffracting planes $\{h_m,k_m,l_m\}$ and $\{h_n,k_n,l_n\}$ respectively for ratio $R_{m,n}$.

The fingerprint for each crystalline material's cubic structure is calculated by taking 10 peaks from the X-ray diffraction data and calculates the two peaks combinations from that 10 peaks. That is, each fingerprint is actually contains 45 values of the quadratic sinus ratio from Eq. 2 since C_2^{10} equal to 45. Table 1, 2, 3 and 4 depict the fingerprint for Face Centered Cubic (FCC), Diamond, Body Centered Cubic (BCC) and Simple Cubic (SC) respectively.

Table 1: Face Centered Cubic (FCC) Fingerprint

$R_{1,2}$ = 0.750	$R_{2,3}$ = 0.500	$R_{3,5}$ = 0.666	$R_{4,8}$ = 0.550	$R_{6,8}$ = 0.800
$R_{1,3}$ = 0.375	$R_{2,4}$ = 0.364	$R_{3,6}$ = 0.500	$R_{1,9}$ = 0.458	$R_{6,9}$ = 0.666
$R_{1,4}$ = 0.273	$R_{2,5}$ = 0.333	$R_{3,7}$ = 0.421	$R_{1,10}$ = 0.407	$R_{6,10}$ = 0.593
$R_{1,5}$ = 0.250	$R_{2,6}$ = 0.250	$R_{3,8}$ = 0.400	$R_{5,6}$ = 0.750	$R_{7,8}$ = 0.950
$R_{1,6}$ = 0.187	$R_{2,7}$ = 0.210	$R_{3,9}$ = 0.333	$R_{5,7}$ = 0.632	$R_{7,9}$ = 0.792
$R_{1,7}$ = 0.158	$R_{2,8}$ = 0.200	$R_{3,10}$ = 0.296	$R_{5,8}$ = 0.600	$R_{7,10}$ = 0.704
$R_{1,8}$ = 0.150	$R_{2,9}$ = 0.166	$R_{4,5}$ = 0.916	$R_{5,9}$ = 0.500	$R_{8,9}$ = 0.833
$R_{1,9}$ = 0.125	$R_{2,10}$ = 0.148	$R_{4,6}$ = 0.687	$R_{5,10}$ = 0.444	$R_{8,10}$ = 0.741
$R_{1,10}$ = 0.111	$R_{3,4}$ = 0.727	$R_{4,7}$ = 0.579	$R_{6,7}$ = 0.842	$R_{9,10}$ = 0.888

Table 2: Diamond Fingerprint

$R_{1,2}$ = 0.375	$R_{2,3}$ = 0.727	$R_{3,5}$ = 0.579	$R_{4,8}$ = 0.500	$R_{6,8}$ = 0.750
$R_{1,3}$ = 0.273	$R_{2,4}$ = 0.500	$R_{3,6}$ = 0.458	$R_{1,9}$ = 0.457	$R_{6,9}$ = 0.686
$R_{1,4}$ = 0.187	$R_{2,5}$ = 0.421	$R_{3,7}$ = 0.407	$R_{1,10}$ = 0.400	$R_{6,10}$ = 0.600
$R_{1,5}$ = 0.158	$R_{2,6}$ = 0.333	$R_{3,8}$ = 0.344	$R_{5,6}$ = 0.792	$R_{7,8}$ = 0.844
$R_{1,6}$ = 0.125	$R_{2,7}$ = 0.296	$R_{3,9}$ = 0.314	$R_{5,7}$ = 0.704	$R_{7,9}$ = 0.771
$R_{1,7}$ = 0.111	$R_{2,8}$ = 0.250	$R_{3,10}$ = 0.275	$R_{5,8}$ = 0.594	$R_{7,10}$ = 0.675
$R_{1,8}$ = 0.094	$R_{2,9}$ = 0.228	$R_{4,5}$ = 0.842	$R_{5,9}$ = 0.543	$R_{8,9}$ = 0.914
$R_{1,9}$ = 0.086	$R_{2,10}$ = 0.200	$R_{4,6}$ = 0.666	$R_{5,10}$ = 0.457	$R_{8,10}$ = 0.800
$R_{1,10}$ = 0.075	$R_{3,4}$ = 0.687	$R_{4,7}$ = 0.593	$R_{6,7}$ = 0.889	$R_{9,10}$ = 0.875

Table 3: Body Centered Cubic (BCC) Fingerprint

$R_{1,2}$ = 0.500	$R_{2,3}$ = 0.666	$R_{3,5}$ = 0.600	$R_{1,8}$ = 0.500	$R_{6,8}$ = 0.750
$R_{1,3}$ = 0.333	$R_{2,4}$ = 0.500	$R_{3,6}$ = 0.500	$R_{1,9}$ = 0.444	$R_{6,9}$ = 0.666
$R_{1,4}$ = 0.250	$R_{2,5}$ = 0.400	$R_{3,7}$ = 0.428	$R_{1,10}$ = 0.400	$R_{6,10}$ = 0.600
$R_{1,5}$ = 0.200	$R_{2,6}$ = 0.333	$R_{3,8}$ = 0.375	$R_{5,6}$ = 0.833	$R_{7,8}$ = 0.875
$R_{1,6}$ = 0.166	$R_{2,7}$ = 0.286	$R_{3,9}$ = 0.333	$R_{5,7}$ = 0.714	$R_{7,9}$ = 0.777
$R_{1,7}$ = 0.143	$R_{2,8}$ = 0.250	$R_{3,10}$ = 0.300	$R_{5,8}$ = 0.625	$R_{7,10}$ = 0.700
$R_{1,8}$ = 0.125	$R_{2,9}$ = 0.222	$R_{1,5}$ = 0.800	$R_{5,9}$ = 0.555	$R_{8,9}$ = 0.888
$R_{1,9}$ = 0.111	$R_{2,10}$ = 0.200	$R_{1,6}$ = 0.666	$R_{5,10}$ = 0.500	$R_{8,10}$ = 0.800
$R_{1,10}$ = 0.100	$R_{3,4}$ = 0.750	$R_{1,7}$ = 0.571	$R_{6,7}$ = 0.857	$R_{9,10}$ = 0.900

Table 4: Simple Cubic (SC) Fingerprint

$R_{1,2}$ = 0.500	$R_{2,3}$ = 0.666	$R_{3,5}$ = 0.600	$R_{4,8}$ = 0.444	$R_{6,8}$ = 0.666
$R_{1,3}$ = 0.333	$R_{2,4}$ = 0.500	$R_{3,6}$ = 0.500	$R_{1,9}$ = 0.400	$R_{6,9}$ = 0.600
$R_{1,4}$ = 0.250	$R_{2,5}$ = 0.400	$R_{3,7}$ = 0.375	$R_{1,10}$ = 0.364	$R_{6,10}$ = 0.545
$R_{1,5}$ = 0.200	$R_{2,6}$ = 0.333	$R_{3,8}$ = 0.333	$R_{5,6}$ = 0.833	$R_{7,8}$ = 0.888
$R_{1,6}$ = 0.166	$R_{2,7}$ = 0.250	$R_{3,9}$ = 0.300	$R_{5,7}$ = 0.625	$R_{7,9}$ = 0.800
$R_{1,7}$ = 0.125	$R_{2,8}$ = 0.222	$R_{3,10}$ = 0.273	$R_{5,8}$ = 0.555	$R_{7,10}$ = 0.727
$R_{1,8}$ = 0.111	$R_{2,9}$ = 0.200	$R_{4,5}$ = 0.800	$R_{5,9}$ = 0.500	$R_{8,9}$ = 0.900
$R_{1,9}$ = 0.100	$R_{2,10}$ = 0.182	$R_{4,6}$ = 0.666	$R_{5,10}$ = 0.454	$R_{8,10}$ = 0.818
$R_{1,10}$ = 0.091	$R_{3,4}$ = 0.750	$R_{4,7}$ = 0.500	$R_{6,7}$ = 0.750	$R_{9,10}$ = 0.909

3. Support Vector Machine

Support Vector Machine (SVM) can be regarded as an excellent statistical learning performance and superior classification performance. Simply put, the support vector machine that we used in this paper can be summarized as follows: it divides two specified training samples which belong to two different categories through constructing an optimal separating hyperplane either in the original space or in the mapped higher dimensional space [4].

The basic idea of constructing this optimal separating hyperplane is to guarantee maximum distance between each training sample and the separating hyperplane. The SVM algorithm and its learning procedure can be comprehended as follows:

- If the data are linearly separable in input space, a binary classification task is taken into account. Let $\{(x_i, y_i)\}(1 \le i \le N)$ be a linearly separable set, where

$x_i \in R^d, y_i \in \{-1,1\}$ and y_i are labels of categories. The general expression of the linear discrimination function in *d*-dimension space is defined as $g(x) = w \cdot x + b$, and the corresponding equation of separating hyperplane is $w \cdot x + b = 0$ where w is normal to the hyperplane, $|b|/\|w\|$ is the perpendicular distance from the hyperplane to the origin, and $\|w\|$ is the Euclidean norm of w. Further normalize $g(x)$ and let all the x_i meet $|g(x)| \geq 1$ will results to the samples which are closest to optimal separating hyperplane that meet $|g(x)| = 1$. Hence, the separating interval is equal to $2/\|w\|$ and solving the optimal separating hyperplane itself is equivalent to minimizing $\|w\|$. The objective function used for this is as follows:

$$\min \Phi(w) = \frac{1}{2}\|w\|^2 \tag{3}$$

subject to the constraints:

$$y_i(w \cdot x_i + b) \geq 1, i = 1,..., N \tag{4}$$

The constraints in Eq. 4 can be replaced by the Lagrange multipliers of the Lagrangian algorithm to further solved the constraints problem where $w = \sum_i \alpha_i y_i x_i$ and x_i are the samples only appearing in the separating interval planes. These samples are called support vectors and its classification function is defined as follows:

$$f(x) = \text{sgn}\left(\sum_i \alpha_i y_i x_i \cdot x + b\right) \tag{5}$$

- If data are not linearly separable in the input space, simply use the following objective function:

$$\min \Phi(w, \xi) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{N} \xi_i\right) \tag{6}$$

Where, ξ is slack variable and C is a penalty factor. Simultaneously, through a non-linear transform $\Phi(\cdot)$, the input space is mapped into a higher dimensional space called feature space in which optimal separating hyperplane can be solved. In addition, the inner product calculation is changed into $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ where $K(x_i, x_j)$ kernel function is defined as inner product in Hilbert space. Thus the final classification function can be represented as follows:

$$f(x) = \text{sgn}\left(\sum_i \alpha_i y_i K(x, x_i) \cdot x + b\right) \tag{7}$$

4. Experiments

As mentioned earlier, the main intent of this paper is to showcase the superior results on the use of support vector machine over neural network in crystalline materials cubic structure identification. In our previous work [5], we had applied a back-propagation neural network to identify the cubic structure of three samples, Aluminium (Al), Silicon (Si), and a mixture of Al and Si. In this experiment, more samples are used in addition to that three samples such as Iron (Fe), Tungsten (W), Sodiumchloride (NaCl), and CopperZinc (CuZn). From the new samples, a mixture sample is also created from three samples Al, Si, and Fe to evaluate the performance of support vector machine in identifying a more complex pattern data.

The experiments are carried out using the same equipments and environment settings as with our previous work which is using a Philips' X-Ray device diffractometer control PW1710. The device was also using PW1729 series of X-Ray generator, anode Cu tube, and PCAPD (PW1877) software version 3 integrated with our proposed application installed. In the overall experiments, we use the same tube voltage and also the same tube current, i.e. 30kV and 20mA respectively.

The SVMlight package [3] was used to construct the support vector machine (SVM) classifiers. For a given set of binary-labeled training examples, SVM maps the input space into a higher dimensional feature space and seeks a hyperplane in the feature space to separate the positive data instances from the negative ones. Different values for the γ and C parameters were tested to optimize the performance of SVM to identify the cubic structures from the presented crystalline material's sample.

Since the fingerprint training dataset was imbalanced, the cost factor was set to 5.8 for giving more weight to training errors on positive examples than errors on negative ones. In the end, after experimenting several different value for C and γ, we use γ = 0.095 and C = 0.55 as the SVM parameter in this experiment. While the rest of the parameters were set to their default values as specified in SVMlight package.

As for the neural network, we kept the same setting from our previous work [5] for comparison wise. That is, the neural network was constructed with 8 nodes in the hidden layer, 5 in the first hidden layer and 3 in the second hidden layer. It uses standard iterative gradient algorithm for back-propagation training.

5. Analysis & Results

After feeding all the samples into the X-Ray diffractometer, the resulted diffraction data for each sample and its ratio $R_{m,n}$ values are shown on Table 5, 6, 7, 8 and 9 for Iron (Fe), Tungsten (W), Sodiumchloride (NaCl), CopperZinc (CuZn), and the mixture of Al, Si, and Fe respectively. While the $R_{m,n}$ ratio values for Al, Si and mixture of Al and Si are available at [5].

Table 5: The resulted $R_{m,n}$ ratio values for the sample of Iron (Fe)

$R_{1,2}$ = 0.500	$R_{1,5}$ = 0.200	$R_{2,4}$ = 0.500	$R_{3,4}$ = 0.752	$R_{4,5}$ = 0.799
$R_{1,3}$ = 0.332	$R_{1,6}$ = 0.166	$R_{2,5}$ = 0.399	$R_{3,5}$ = 0.601	$R_{4,6}$ = 0.666
$R_{1,4}$ = 0.250	$R_{2,3}$ = 0.664	$R_{2,6}$ = 0.333	$R_{3,6}$ = 0.501	$R_{5,6}$ = 0.833

Table 6: The resulted $R_{m,n}$ ratio values for the sample of Tungsten (W)

$R_{1,2}$ = 0.500	$R_{2,3}$ = 0.666	$R_{3,5}$ = 0.600	$R_{4,8}$ = 0.500
$R_{1,3}$ = 0.333	$R_{2,4}$ = 0.500	$R_{3,6}$ = 0.500	$R_{5,6}$ = 0.833
$R_{1,4}$ = 0.250	$R_{2,5}$ = 0.400	$R_{3,7}$ = 0.428	$R_{5,7}$ = 0.714
$R_{1,5}$ = 0.200	$R_{2,6}$ = 0.333	$R_{3,8}$ = 0.375	$R_{5,8}$ = 0.625
$R_{1,6}$ = 0.166	$R_{2,7}$ = 0.286	$R_{4,5}$ = 0.800	$R_{6,7}$ = 0.857
$R_{1,7}$ = 0.143	$R_{2,8}$ = 0.250	$R_{4,6}$ = 0.666	$R_{6,8}$ = 0.750
$R_{1,8}$ = 0.125	$R_{3,4}$ = 0.750	$R_{4,7}$ = 0.571	$R_{7,8}$ = 0.875

Table 7: The resulted $R_{m,n}$ ratio values for the sample of Sodiumchloride (NaCl)

$R_{1,2}$ = 0.750	$R_{2,3}$ = 0.500	$R_{3,5}$ = 0.666	$R_{4,8}$ = 0.550	$R_{6,8}$ = 0.800
$R_{1,3}$ = 0.375	$R_{2,4}$ = 0.363	$R_{3,6}$ = 0.500	$R_{4,9}$ = 0.458	$R_{6,9}$ = 0.666
$R_{1,4}$ = 0.273	$R_{2,5}$ = 0.333	$R_{3,7}$ = 0.421	$R_{4,10}$ = 0.407	$R_{6,10}$ = 0.593
$R_{1,5}$ = 0.250	$R_{2,6}$ = 0.250	$R_{3,8}$ = 0.400	$R_{5,6}$ = 0.750	$R_{7,8}$ = 0.950
$R_{1,6}$ = 0.187	$R_{2,7}$ = 0.210	$R_{3,9}$ = 0.333	$R_{5,7}$ = 0.632	$R_{7,9}$ = 0.792
$R_{1,7}$ = 0.158	$R_{2,8}$ = 0.200	$R_{3,10}$ = 0.296	$R_{5,8}$ = 0.600	$R_{7,10}$ = 0.704
$R_{1,8}$ = 0.150	$R_{2,9}$ = 0.166	$R_{4,5}$ = 0.916	$R_{5,9}$ = 0.500	$R_{8,9}$ = 0.833
$R_{1,9}$ = 0.125	$R_{2,10}$ = 0.148	$R_{4,6}$ = 0.687	$R_{5,10}$ = 0.444	$R_{8,10}$ = 0.741
$R_{1,10}$ = 0.111	$R_{3,4}$ = 0.727	$R_{4,7}$ = 0.578	$R_{6,7}$ = 0.842	$R_{9,10}$ = 0.888

Table 8: The resulted $R_{m,n}$ ratio values for the sample of CopperZinc (CuZn)

$R_{1,2}$ = 0.500	$R_{2,3}$ = 0.667	$R_{3,5}$ = 0.600	$R_{4,8}$ = 0.444	$R_{6,8}$ = 0.666
$R_{1,3}$ = 0.333	$R_{2,4}$ = 0.501	$R_{3,6}$ = 0.500	$R_{4,9}$ = 0.401	$R_{6,9}$ = 0.601
$R_{1,4}$ = 0.250	$R_{2,5}$ = 0.401	$R_{3,7}$ = 0.376	$R_{4,10}$ = 0.363	$R_{6,}10$ = 0.546
$R_{1,5}$ = 0.200	$R_{2,6}$ = 0.334	$R_{3,8}$ = 0.333	$R_{5,6}$ = 0.833	$R_{7,8}$ = 0.887
$R_{1,6}$ = 0.166	$R_{2,7}$ = 0.251	$R_{3,9}$ = 0.301	$R_{5,7}$ = 0.626	$R_{7,9}$ = 0.800
$R_{1,7}$ = 0.125	$R_{2,8}$ = 0.223	$R_{3,10}$ = 0.273	$R_{5,8}$ = 0.555	$R_{7,}10$ = 0.726
$R_{1,8}$ = 0.111	$R_{2,9}$ = 0.201	$R_{4,5}$ = 0.800	$R_{5,9}$ = 0.501	$R_{8,9}$ = 0.900
$R_{1,9}$ = 0.100	$R_{2,}10$ = 0.182	$R_{4,6}$ = 0.666	$R_{5,}10$ = 0.454	$R_{8,10}$ = 0.817
$R_{1,10}$ = 0.091	$R_{3,4}$ = 0.750	$R_{4,7}$ = 0.501	$R_{6,7}$ = 0.751	$R_{9,10}$ = 0.908

Table 9: The resulted $R_{m,n}$ ratio values for the mixture sample of Aluminium (Al), Silicon (Si), and Iron (Fe)

$R_{1,2}$ = 0.551	$R_{2,3}$ = 0.750	$R_{3,4}$ = 0.898	$R_{4,5}$ = 0.727	$R_{5,6}$ = 0.758
$R_{1,3}$ = 0.416	$R_{2,4}$ = 0.679	$R_{3,5}$ = 0.654	$R_{4,6}$ = 0.553	$R_{5,7}$ = 0.687
$R_{1,4}$ = 0.375	$R_{2,5}$ = 0.495	$R_{3,6}$ = 0.500	$R_{4,7}$ = 0.500	$R_{5,8}$ = 0.579
$R_{1,5}$ = 0.273	$R_{2,6}$ = 0.375	$R_{3,7}$ = 0.448	$R_{4,8}$ = 0.421	$R_{5,9}$ = 0.554
$R_{1,6}$ = 0.207	$R_{2,7}$ = 0.339	$R_{3,8}$ = 0.378	$R_{4,9}$ = 0.404	$R_{5,10}$ = 0.509
$R_{1,7}$ = 0.187	$R_{2,8}$ = 0.286	$R_{3,9}$ = 0.363	$R_{4,10}$ = 0.371	$R_{5,11}$ = 0.458
$R_{1,8}$ = 0.158	$R_{2,9}$ = 0.275	$R_{3,10}$ = 0.333	$R_{4,11}$ = 0.333	$R_{5,12}$ = 0.407
$R_{1,9}$ = 0.151	$R_{2,10}$ = 0.250	$R_{3,11}$ = 0.299	$R_{4,12}$ = 0.296	$R_{5,13}$ = 0.381
$R_{1,10}$ = 0.139	$R_{2,11}$ = 0.226	$R_{3,12}$ = 0.266	$R_{4,13}$ = 0.277	$R_{5,14}$ = 0.343
$R_{1,}11$ = 0.125	$R_{2,12}$ = 0.201	$R_{3,13}$ = 0.250	$R_{4,14}$ = 0.250	$R_{5,15}$ = 0.321
$R_{1,12}$ = 0.111	$R_{2,13}$ = 0.187	$R_{3,14}$ = 0.224	$R_{4,15}$ = 0.234	$R_{5,16}$ = 0.313
$R_{1,13}$ = 0.104	$R_{2,14}$ = 0.169	$R_{3,15}$ = 0.210	$R_{4,16}$ = 0.228	$R_{5,17}$ = 0.305
$R_{1,14}$ = 0.093	$R_{2,15}$ = 0.158	$R_{3,16}$ = 0.205	$R_{4,17}$ = 0.222	$R_{5,18}$ = 0.275
$R_{1,15}$ = 0.087	$R_{2,16}$ = 0.155	$R_{3,17}$ = 0.200	$R_{4,18}$ = 0.200	$R_{5,19}$ = 0.250
$R_{1,16}$ = 0.085	$R_{2,17}$ = 0.151	$R_{3,18}$ = 0.179	$R_{4,19}$ = 0.185	$R_{10,11}$ = 0.897
$R_{1,17}$ = 0.083	$R_{2,18}$ = 0.135	$R_{3,19}$ = 0.166	$R_{9,10}$ = 0.914	$R_{10,12}$ = 0.798
$R_{1,18}$ = 0.075	$R_{2,19}$ = 0.125	$R_{8,9}$ = 0.960	$R_{9,11}$ = 0.823	$R_{10,13}$ = 0.750
$R_{1,19}$ = 0.069	$R_{7,8}$ = 0.842	$R_{8,10}$ = 0.881	$R_{9,12}$ = 0.732	$R_{10,14}$ = 0.673
$R_{6,7}$ = 0.904	$R_{7,9}$ = 0.809	$R_{8,11}$ = 0.792	$R_{9,13}$ = 0.687	$R_{10,15}$ = 0.630
$R_{6,8}$ = 0.762	$R_{7,10}$ = 0.742	$R_{8,12}$ = 0.704	$R_{9,14}$ = 0.617	$R_{10,16}$ = 0.615

$R_{6,9}$ = 0.727	$R_{7,11}$ = 0.666	$R_{8,13}$ = 0.660	$R_{9,15}$ = 0.579	$R_{10,17}$ = 0.600
$R_{6,10}$ = 0.666	$R_{7,12}$ = 0.592	$R_{8,14}$ = 0.593	$R_{9,16}$ = 0.564	$R_{10,18}$ = 0.538
$R_{6,11}$ = 0.601	$R_{7,13}$ = 0.556	$R_{8,15}$ = 0.555	$R_{9,17}$ = 0.550	$R_{10,19}$ = 0.500
$R_{6,12}$ = 0.536	$R_{7,14}$ = 0.500	$R_{8,16}$ = 0.542	$R_{9,18}$ = 0.494	
$R_{6,13}$ = 0.500	$R_{7,15}$ = 0.468	$R_{8,17}$ = 0.529	$R_{9,}$19 = 0.458	
$R_{6,14}$ = 0.452	$R_{7,16}$ = 0.458	$R_{8,18}$ = 0.475	$R_{15,16}$ = 0.976	
$R_{6,15}$ = 0.421	$R_{7,17}$ = 0.445	$R_{8,19}$ = 0.440	$R_{15,17}$ = 0.950	
$R_{6,16}$ = 0.413	$R_{7,18}$ = 0.400	$R_{13,14}$ = 0.899	$R_{15,18}$ = 0.854	
$R_{6,17}$ = 0.400	$R_{7,19}$ = 0.371	$R_{13,15}$ = 0.842	$R_{15,19}$ = 0.792	
$R_{6,18}$ = 0.364	$R_{12,13}$ = 0.938	$R_{13,16}$ = 0.821	$R_{16,17}$ = 0.975	
$R_{6,19}$ = 0.335	$R_{12,14}$ = 0.842	$R_{13,17}$ = 0.800	$R_{16,18}$ = 0.875	
$R_{11,12}$ = 0.888	$R_{12,15}$ = 0.789	$R_{13,18}$ = 0.719	$R_{16,19}$ = 0.812	
$R_{11,13}$ = 0.834	$R_{12,16}$ = 0.771	$R_{13,19}$ = 0.666	$R_{17,18}$ = 0.897	
$R_{11,14}$ = 0.750	$R_{12,17}$ = 0.752	$R_{14,15}$ = 0.936	$R_{17,19}$ = 0.833	
$R_{11,15}$ = 0.702	$R_{12,18}$ = 0.675	$R_{14,16}$ = 0.914	$R_{18,19}$ = 0.927	
$R_{11,16}$ = 0.686	$R_{12,19}$ = 0.625	$R_{14,17}$ = 0.891		
$R_{11,17}$ = 0.667		$R_{14,18}$ = 0.800		
$R_{11,18}$ = 0.600		$R_{14,19}$ = 0.742		
$R_{11,19}$ = 0.555				

Table 10: SVM and Neural Network Comparison Results

Sample Name	SVM		Neural Network	
	Type	Accuracy	Type	Accuracy
Al + Si	FCC	91%	FCC	86%
	Diamond	92%	Diamond	84%
Al + Si + Fe	FCC	90%	FCC	78%
	Diamond	82%	Diamond	67%
	BCC	73%	SC	51%
Al	FCC	93%	FCC	92%
Si	Diamond	92%	Diamond	91%
Fe	BCC	89%	BCC	78%
W	BCC	91%	BCC	89%
NaCl	FCC	94%	FCC	92%
CuZn	SC	90%	SC	87%

The last step is to pass the resulted data for each crystalline material sample into the trained SVM and neural network. Table 10 shows the comparison results of SVM and neural network.

From result outlined on Table 10, we can clearly see that support vector machine (SVM) outperform neural network in terms of the accuracy of identifying the cubic structure type from each crystalline material sample. Both SVM and neural network successfully detect the correct cubic structure type for single component crystalline material, but neural network suffers and failed to detect the correct cubic structure type for multi component crystalline material especially the one with three cubic structure components inside.

In the experiment with the mixture sample of Aluminium (Al), Silicon (Si), and Iron (Fe), neural network identify that the sample contain three cubic structure component of FCC, Diamond and SC with accuracy rate of 78%, 67%, and 51% respectively. Since the cubic structure type of Iron (Fe) is BCC, the result given by neural network is wrong. While SVM gives more accurate result of FCC, Diamond and BCC with the accuracy rate of 90%, 82%, and 73% respectively.

Neural network failure is probably due to the fact that the fingerprint structure of BCC and SC is almost the same and among the mixture sample component, Iron (Fe) is the crystalline material that has the smallest number of $R_{m,n}$ ratio values. And after being mixed with the two other components, it makes its ratio pattern difficult to track.

SVM results also outperform the neural network in the mixture of two component of Aluminium (Al) and Silicon (Si). Though both successfully detect the correct cubic structure types (FCC and diamond), but the accuracy of SVM is slightly higher, 91% and 92% respectively for FCC and diamond. As for single component crystalline material samples, both SVM and neural network gives equivalent results.

6. Summary

In this paper we propose the use of support vector machine to enhance the performance of cubic structures identification on multi component crystalline material. The complexity of multi component crystalline material $R_{m,n}$ ratio needs a sophisticated and powerful methods to accurately classify its cubic structure type. And support vector machine which exhibits more excellent performances such as no local optimum problem, no over-fit or under-fit problem, better convergence property, less training samples, higher correct identification rate, and higher reliability suits this job.

This work proves that the use of artificial intelligence techniques such as support vector machine and neural network towards crystalline materials cubic structure identification research to be very useful and can simplify and fasten up the work on crystallography research area. Our future research direction includes applying these artificial intelligence techniques to other problems in crystallography research area.

Acknowledgments

Author would like to acknowledge University of Sumatera Utara for providing space and the equipments used in this research. Also, author would like to thank Muhammad Fermi Pasha for the fruitful discussion and insightful comments on the support vector machine part.

References

[1] Suryanarayana, C. and Norton, M.G. X-Ray Diffraction - A Practical Approach. Plenum Press - NewYork, 1998.
[2] Syukur, M. and Zarlis, M. A Standard Calculation for Lattice Analysis on Cubic Structures' Solid Materials using X-Ray Diffraction Method and its Experimental Study. Journal of Solid State Science and Technology Letter, Vol. 6, No. 2, 1999.
[3] Joachims, T. Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (Eds.), MIT Press, 1999.
[4] Vapnik, V.N. Statistical Learning Theory. New York: John Wiley and Sons, 1998.
[5] Syukur, M., Pasha, M.F. and Budiarto, R. A Neural Network-Based Application to Identify Cubic Structures in Multi Component Crystalline Materials using X-Ray Diffraction Data. International Journal of Computer Science and Network Security (IJCSNS), 7(2): 49-54, 2007.
[6] Cullity, B.D. Elements of X-ray Diffraction. Addison-Wesley - London, 1978.
[7] The International Center for Diffraction Data, http://www.icdd.com
[8] Ewald, P.P. Fifty Years of X-Ray Diffraction. International Union of Crystallography, 1999.
[9] Gubbi, J., Shilton, A., Parker, M. and Palaniswami, M. Protein Topology Classification using Two-Stage Support Vector Machines. Genome Informatics Vol.17, No.2, pp. 259-269, 2006.
[10] Krause, L., McHardy, A.C., Nattkemper, T.W., Pühler, A., Stoye, J. and Meyer, F. GISMO - gene identification using a support vector machine for ORF classification. Nucleic Acids Research, 35(2): 540–549, 2006.
[11] Brockett, C. and Dolan, W.B. Support vector machines for paraphrase identification and corpus construction. Proceedings of The 3rd International Workshop on Paraphrasing (IWP2005), 2005.

Mohammad Syukur received the B.Sc. degree in Physics from University of Sumatera Utara, Indonesia and M.Sc. degrees in Materials Physics from Bandung Institute of Technology, Indonesia in 1974 and 1982, respectively. He is currently a senior lecturer and the head of Crystallography & Radiation Physics Research Laboratory at Faculty of Mathematics and Natural Sciences, University of Sumatera Utara, Indonesia. His research interests include Material Physics, X-Ray Diffraction methods, Artificial Intelligence Application and Crystallography. He has written two book chapters and numerous refereed research papers.