Materials Today Communications 34 (2023) 105311

Contents lists available at ScienceDirect

Materials Today Communications

journal homepage: www.elsevier.com/locate/mtcomm

A new ANN-PSO framework to chalcopyrite's energy band gaps prediction

Inas Bouzateur [a,b], Hamza Bennacer [a,c,*], Mohammed Assam Ouali [a,b], Mohamed Issam Ziane [d], Moufdi Hadjab [a], Mohamed Ladjal [a,b]

[a] Department of Electronics, Faculty of Technology, University of M'sila, Algeria
[b] LASS, Laboratory of Analysis of Signals and Systems, University of M'sila, Algeria
[c] Elaboration and Physico-Mechanical and Metallurgical Characterization of Materials Laboratory, ECP3M, University of Mostaganem, Algeria
[d] Higher School in Electrical and Energetic Engineering (ESGEE), Oran 31000, Algeria

ARTICLE INFO

Keywords:
Chalcopyrite
Band Gap Energy
Prediction
ANN
PSO

ABSTRACT

The electronic band gap energy is an essential photo-electronic parameter in the energy applications of engineering materials, particularly in solar cells and photo-catalysis domains. A prediction model that can correctly predict this band gap energy is desirable. A new approach for predicting a band gap energy is suggested in this paper. The proposed structure is based on artificial neural networks (ANN) and the particle swarm optimization algorithm (PSO); this structure can solve the artificial neural network's local minima issue while preserving the fitting quality. Our technique will hasten the identification of novel chalcopyrite in photovoltaic solar cells with improved resolution. The suggested model combines two sub-systems in a parallel configuration. A conventional prediction system with a low resolution for the training data being considered makes up the first ANN sub-system. A second ANN sub-system, labelled the error model, is introduced to the primary system to address the resolution quality issue, representing uncertainty in the primary model. The particle swarm optimization algorithm is used to identify the parameters of the proposed neural system. The method's effectiveness is assessed in terms of several criteria, and the output of our system shows good performance compared to experimental and other calculated results. Several benchmark approaches were compared with the proposed system in detail. Numerous computer tests show that the suggested strategy can significantly enhance convergence and resolution.

1. Introduction

Materials' high performance and energy efficiency are required conditions in smart optoelectronic devices nowadays. Therefore, research on novel materials with improved performance and higher energy efficiency has been steadily growing. Decades of research proved that ternary compounds with ABC2 structure (III-V-V_2 and I-III-VI_2) are exciting materials for electrical, photovoltaic, optoelectronic, and photonic applications due to their electronic, linear, and nonlinear optical properties. These materials have recently been used in light-emitting diodes, solar cells, and microelectronics industries [1–5]. There has been a considerable increase in interest in the use of thin-film solar cell devices, including silicon solar cells and photovoltaic absorbers. Due to their excellent band gap, copper indium gallium selenide (CIGS) absorber composites are increasingly popular for solar power production. They are thin-film solar cells with an efficiency equivalent to solar cells made from crystalline silicon (c-Si) wafers [6]. Developing high-efficiency and low-cost solar cells within novel photovoltaic materials is a crucial objective for research in this field. Thence, the identification of the electrical properties of their materials becomes crucial. The band gap energy is an essential characteristic of photovoltaic semiconductors. Materials with band gap energy between 1.1 and 1.8 eV are often utilized as solar absorbers in fabricating solar cells [7]. In principle, the ab-initio method is usually used to determine the physical properties of materials. The notable underestimation error occurs in calculating semiconductor band gap energy using classical density functional theory (DFT) [8]. More precise theoretical approaches can overcome this issue, such as the atomic-orbital-based approximate self-interaction correction scheme [8], the Wannier–Fermi–Löwdin self-interaction correction approach [9], the modified Becke-Johnson scheme (mBJ) [10,11], the screened hybrid functional [12], and the many-body perturbation theory (GW) [13–15]. However, the latter techniques can be computationally costly, which accounts for their rare implementation of a large set of materials.

Abbreviations: ANN, Artificial Neural Networks; PSO, Particle Swarm Optimization.

* Corresponding author at: Department of Electronics, Faculty of Technology, University of M'sila, Algeria.
E-mail address: hamza.bennacer@univ-msila.dz (H. Bennacer).

https://doi.org/10.1016/j.mtcomm.2023.105311
Received 23 November 2022; Received in revised form 21 December 2022; Accepted 1 January 2023
Available online 2 January 2023

Table 1
Properties of elements forming the I-III-IV_2, II-IV-V_2 Compounds in the data set [16,19,29–31].

I-III-IV_2 Compounds									II-IV-V_2 Compounds								
Grp	Elm	EN	AN	MP	PR	VL	MN1	MN2	Grp	Elm	EN	AN	MP	PR	VL	MN1	MN2
I	Cu	1.08	29	1358	2.04	11	66	68	II	Zn	1.44	30	692.7	1.88	12	69	71
	Ag	1.07	47	1235	2.375	11	67	67		Cd	1.4	48	594.3	2.215	12	70	70
	Au	1.19	79	1338	2.66	11	68	66		Hg	1.49	80	234.3	2.41	12	71	69
										Be	1.45	4	1562	1.08	2	7	12
										Mg	1.31	12	922	2.03	2	8	11
										Ca	1.17	20	1112	3	2	9	10
										Sr	1.13	38	1042	3.21	2	10	9
										Ba	1.08	56	1002	3.402	2	11	8
										Ra	0.9	88	973	3.53	2	12	7
III	B	1.9	5	2365	0.795	3	72	76	IV	C	2.37	6	3800	0.64	4	77	81
	Al	1.64	13	933.5	1.675	3	73	75		Si	1.98	14	1687	1.42	4	78	80
	Ga	1.7	31	302.9	1.695	3	74	74		Ge	1.99	32	1211	1.56	4	79	79
	In	1.63	49	429.8	2.05	3	75	73		Sn	1.88	50	505.1	1.88	4	80	78
	Tl	1.69	81	577	2.235	3	76	72		Pb	1.92	82	600.7	2.09	4	81	77
VI	O	3.32	8	54.36	0.465	6	87	91	V	N	2.85	7	63.15	0.54	5	82	86
	S	2.65	16	388.4	1.1	6	88	90		P	2.32	15	317.2	1.24	5	83	85
	Se	2.54	34	494	1.285	6	89	89		As	2.27	33	1089	1.415	5	84	84
	Te	2.38	52	722.7	1.67	6	90	88		Sb	2.14	51	903.9	1.765	5	85	83
										Bi	2.14	83	544.6	1.997	5	86	82

Abbreviation: Grp = Group, Elm = Element, EN = Electronigativity, AN: Atomic number, MP = Melting Point, PR = pseudopotential Radius, VL = Valency, MN1 = Mendeleev Number 1, MN2 =Mendeleev Number 2.

To overcome this issue, theoretical research on such compounds introduced a novel approach to accelerating the study and prediction of the band gap energy for novel ternary semiconductors. Using machine-learning methods to develop effective computational tools to handle this specific challenge is a new but exciting study area. In this regard, Zeng et al. [16] used an artificial neural network to relate the band gap energy and lattice constant of ternary semiconductors (ABC_2 chalcopyrites) to their chemical stoichiometric and fundamental element properties. This demonstrated that the relationship could be described linearly, thus paving the way for future research to employ linear regression approaches. Suh and Rajan [17] used the partial least square (PLS) regression method to predict the band gap energy of 205 new chalcopyrites. Their model revealed that 77 chalcopyrites had a predictive negative band gap energy value; also, their model failed to adequately explain the physical phenomena and causes of outliers. A subsequent study by Dey et al. [18] used multiple predictive techniques (OLS, LASSO, SPLS) coupled with feature ranking to develop a more robust quantitative model. Recently, Khmaissia et al. [19] proposed a novel approach based on feature selection and regression techniques to predict the band gap energy of 156 chalcopyrites. They achieved the same mean squared error (*MSE*) as reported in Dey et al. [18] when the same descriptors were used; the *MSE* was decreased to 630 meV when the optimal features selected by the correlation technique were used. Additionally, the number of predicted negative values has been reduced to 28, although the models can still not predict them.

In this paper, we suggest a new framework to predict the band gap energy of ABC_2 compounds. First, we employ all the properties of the four atoms in the compounds under consideration to elucidate the difference between the energy gap energy of the ABC and ABC_2 structures. The used input data set comprises chemical parameters. This research proposes a new and effective prediction method based on artificial neural networks (*ANN*) and particle swarm optimization algorithm (*PSO*). Artificial Neural networks (ANN) are computational algorithm that aims to mimic the behavior of "neurons-based" biological systems. This is presented as a system of interconnected “neurons” which can compute values from inputs. The two most important steps are the structure and parameter identification of an artificial neural network prediction system. The first, which deals with structure identification, is critical because it addresses the issue of constructing the framework for the ANN prediction system using input-output data [20]. The second step aims to identify the free parameters of the designed ANN structure. In this investigation, a stochastic algorithm is used to adjust the parameters of the proposed ANN framework adaptively due to its generally good performance; this technique is often used in various domains in various settings [21,22]. It is particularly well adapted to handling complex optimization issues. Through parallel processing in the population, the best solution is discovered. The stochastic algorithm used is the particle swarm optimization algorithm (*PSO*). Kennedy and Eberhart [23] developed PSO based on swarm intelligence. It uses a simple mathematical model to describe how groups of birds or schools of fish interact with one another.

This study uses artificial neural networks and particle swarm optimization algorithm to predict Chalcopyrite’s band gap energy. A primary ANN prediction system is initially devised using training data. After that, the discrepancies between the experiment results and the output of the primary prediction system were utilized to develop an ANN-error model. This second model, which depicts the primary ANN prediction system’s uncertainties, is easily modifiable by subtracting the ANN error model output from the primary ANN prediction system output. Both of the models are connected in a parallel manner. After making the necessary adjustments to the model output, there is a noticeable increase in resolution. The parameters of the primary ANN prediction system and the ANN error model are tuned using a particle swarm optimization algorithm.

In overview, the most significant contribution made by this paper is the expansion of the concept of prediction by adding a new prediction module that is referred to as an ANN-error model. This error model supplements the primary ANN prediction system to enhance the latter output, resulting in a more accurate response. Moreover, the proposed structure deals with a known local minima issue due to the use of the descent gradient algorithm to tune the ANN’s free parameters. The convergence of the proposed method is demonstrated experimentally. Compared to more recently published benchmark approaches [16–19], the presented strategy has high model accuracy and a fast convergence rate.

2. Preliminaries

2.1. Data description

Solar energy is abundant and clean, reaching the earth’s surface at a rate of 120PW [24]. While it is used and exploited in a variety of industries by catching solar energy and transforming it into photovoltaic components like solar cells, these compounds families have a range of

band gaps that can be adjusted in multijunction cells to absorb different energy bands, which maximizes the absorption of the solar spectrum [25]. Several researchers have proposed systems to correlate the band gap energy with chemical properties [16–19,26–28]. The ABC_2 compounds contain two groups, namely I-III-VI_2 and II-IV-V_2 compounds. Each is based on three elements, bringing the total to 302 compounds. Each element has various chemical properties compiled from previous studies such as: Y. Zeng et al. [16], Khmaissia et al. [19], and P. Villars et al. [29–31]. The seven predictor variables relied upon in our work are Electronegativity (EN)($eV^{½}$), Atomic numbers (AN), the Melting point (MP)(k), pseudo potential radii (PR) (Atomic unit au), the number of the electron in the valence band (VL) and the Mendeleev numbers (MN), 1–2, are listed in Table 1. The Mendeleev number MN1 represents the left-top, top-down sequence in the periodic Mendeleev table where hydrogen (H) is placed above fluorine (F), which begins counting from the left-top corner and moves down to the bottom-left corner, going from the first group to the last. Whereas MN2 represents the left-bottom, down-top sequence, which begins counting from the left-down corner and moves up to the top-left corner, going from the first group to the last [30]. The band gap energy (Eg) of the compound ABC_2 may be correlated as a function of seven variables:

$$Eg = f(EN_X, AN_X, MP_X, VL_X, PR_X, MN1_X, MN2_X); \quad (1)$$

Where X denotes the four atoms (A, B and 2 C). Thus are the totals of twenty eight descriptors.

2.2. Artificial Neural Network (ANN)

Artificial Neural Network is one of the most widely used methodologies for modelling complex non-linear systems [32]. The artificial neural network is a biologically inspired computational technique that imitates the mechanism of the brain [21], whereas the human brain use association of neurons, and the neural network use connection weights of neurons, which are organized in layers (input layer, output layer and one or more hidden layers). The information of the neural network is stored in the form of weight and bias; every neuron in the hidden or output layer will first operate as a summing junction, combining and modifying the previous layer's inputs to each neuron using the following equation [21]:

$$y_i = \sum_{j=1}^{m} X_i W_{ij} + b_j; \quad (2)$$

where y_i is the net input to the node j in the hidden or output layer,X_i are the inputs to the node j or output of previous layers,W_{ij} are the weights representing the strength of the connection between the i^{th} node and j^{th} node; i is the number of nodes and b_j is the bias associated with node j. Each neuron has a transfer function that expresses the level of internal activation. A neuron's output is calculated by converting its input with the appropriate transfer function. For time series prediction, the sigmoid function, hyperbolic tangent, and linear function are the most commonly used transfer functions. After the creation of the neural network's structure, the networks will be trained using input-output data (learning phase) until we get the smallest value of the cost function and the networks were found the optimum value of the weights and bias [22]. ANN-based models have several advantages, including excellent generalization ability, which allows them to reliably anticipate output for a fresh input data set as well as the capacity to deal with noisy data and uncertainty; it was employed in different applications such as medicine, economics, engineering, and others [21,33–35].

2.3. Particle swarm optimization (PSO)

The Particle swarm optimization (PSO) algorithm suggested by Kennedy and Eberhart [23] is a computational technique used to find optimal solutions and solve hard optimization problems. PSO is a metaheuristic algorithm inspired by the social behavior of creatures that congregate, such as flocks of birds, schools of fish, and colonies of insects [36]. Particles are the population elements in PSO, and each particle, such as a bird or fish, is a candidate for the solution. Within the N-dimensional search space, every particle is considered a moving point with a specific velocity. The particle's velocity is constantly modified in response to its own experience and the experiences of its companions seeking a better solution area. In PSO, each particle k is presented with a state of two characteristics, position (x_k) and velocity(v_k), which are initialized in a population generation by a random process. These two characteristics are tuned during each iteration in order to reach the best solution [37,38] according to the following formulas:

$$v_{kd}^{t+1} = w.v_{kd}^{t} + c_1.r_1.\left(pbest_{kd} - x_{kd}^{t}\right) + c_2.r_2.\left(gbest_d - x_{kd}^{t}\right); \quad (3)$$

$$x_{kd}^{t+1} = x_{kd}^{t} + v_{kd}^{t+1}; \quad (4)$$

where x_{kd}^{t} and v_{kd}^{t} are position and velocity of k^{th} particle in the dimension d ($d = 1, 2.....m$) at t iteration, m is the swarm size, the velocity of each particle is limited to the maximum value within the interval $[LB, UB]$ defined in consideration of the limit of the decision variable,$pbest_k$ is the best position of the particle k, the best global position is $gbest$, c_1 and c_2 are the constants of acceleration, r_1 and r_2 represent a random parameters their values are in the range of $[0, 1]$, w is the inertial factor of particle swarm optimization algorithm, it gives momentum to particles moving in the design space and is also used for balancing global and local search options during the optimization process [39]. After updating position and velocity, a new cost function of the particles is calculated, and the same process will continue to run until a stop criterion is given [38].

The PSO is used in many applications to solve different problems, as reported in [39–42]; among these studies, we find that used PSO for predicting the crystal lattice parameters of pseudo-cubic and cubic perovskites materials to optimize the parameters of support vector regression (SVR) algorithms [43], therefore, in this paper, we use the PSO algorithm to optimize the parameters and improve the performance of ANN model.

3. Proposed ANN-PSO prediction method

This paper presents a new approach based on artificial neural networks and particle swarm optimization algorithm. The main aim is to give a new prediction system to accelerate the discovery of various materials properties with higher accuracy.

The method created to examine several descriptors of the materials' electronic and crystal structure properties under consideration is based on ab-initio calculations and experimentation. The generated input data is preprocessed to remove outliers and make the neural network more efficient, and all features have been subjected to a phase of preprocessing and normalization so that their ranges of variation are comparable to one another. Various features are applied to predict the band gap energy and prove the accuracy of our system. To get started, we are going to use our collected data which contains 28 features [16,19]; after that, we will use the set of a feature that Khmaissia et al. [19] applied in their work to provide a reliable comparative investigation, which then has been used to evaluate the accuracy of our system through of calculating prediction errors.

The suggested approach is divided into four phases, which are as follows:

***Phase* 1**: Data pre-processing;.

***Phase* 2**: Primary prediction system identification;.

***Phase* 3**: Error Process identification;.

***Phase* 4**: Final prediction system design.

The free parameters of the artificial neural networks systems to be adjusted using the particle swarm optimization algorithm are:

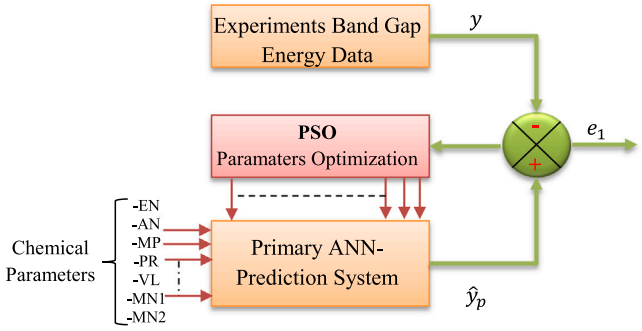

Fig. 1. Primary ANN-prediction system.

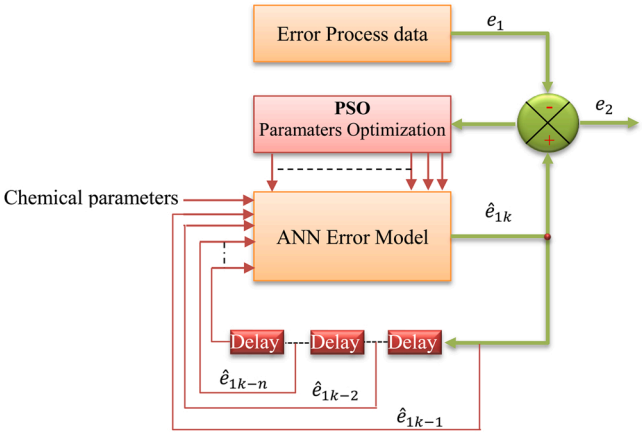

Fig. 2. ANN-error model.

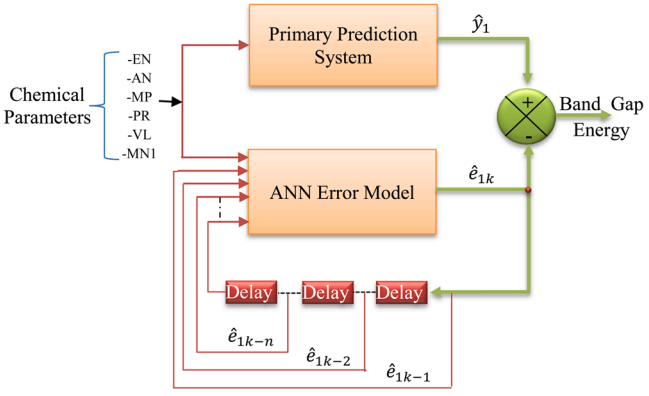

Fig. 3. Final prediction system.

- ✓ **Weights**, which are the parameters in a neural network's hidden layers, change the input data.
- ✓ **Biases**, which are the constants added to the product of features and weights. They are used to offset the result.

Through this study, the mean squared error (*MSE*) will be applied as the objective function, which can be described as follows:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2; \tag{5}$$

Where y_n is the actual output,$\hat{y}_n$ is the predicted output and N is the length of data.

Table 2
Compounds used in training phase and their experimental band gap energy (Eg) (eV) [19].

N°	Compounds	Eg (eV)		Compounds	Eg (eV)
1	$CuAlS_2$	3.49	15	$CuGaS_2$	2.43
2	$AgAlS_2$	3.13	16	$AgGaS_2$	2.64
3	$CuAlSe_2$	2.67	17	$CuGaSe_2$	1.68
4	$AgAlSe_2$	2.55	18	$AgGaSe_2$	1.8
5	$CuAlTe_2$	2.06	19	ZnSiP2	2.07
6	$AgAlTe_2$	2.27	20	$ZnSiAs_2$	1.74
7	$CuGaTe_2$	1.12	21	$ZnGeP_2$	2.05
8	$AgGaTe_2$	1.32	22	$ZnGeAs_2$	1.15
9	$CuInS_2$	1.53	23	$CdSiP_2$	2.33
10	$AgInS_2$	1.87	24	$CdSiAs_2$	1.55
11	$CuInSe_2$	1.04	25	$CdGeP_2$	1.72
12	$AgInSe_2$	1.24	26	$CdGeAs_2$	0.57
13	$CuInTe_2$	1.06	27	$CdSnP_2$	1.17
14	$AgInTe_2$	0.95	28	$CdSnAs_2$	0.26

Table 3
The Compounds used in the validation phase and their experimental band gap energy Eg (eV) [19].

N	Compounds	Eg (eV)
1	$GaInP_2$	2
2	$AgGaO_2$	4.1
3	$CuAlO_2$	3.5
4	$CuBO_2$	2.2
5	$CuBS_2$	3.61
6	$CuBSe_2$	3.13
7	$CuGaO_2$	3.37
8	$AgBO_2$	2.21
9	$AuBS_2$	2.55
10	$AuBSe_2$	1.53
11	$AuBTe_2$	1.33
12	$HgGeAs_2$	0.2
13	$AgAlO_2$	3.6
14	$CuInO_2$	3.9
15	$MgGeAs_2$	1.6
16	$MgSiAs_2$	2
17	$MgSiP_2$	2.3
18	$ZnGeN_2$	2.67

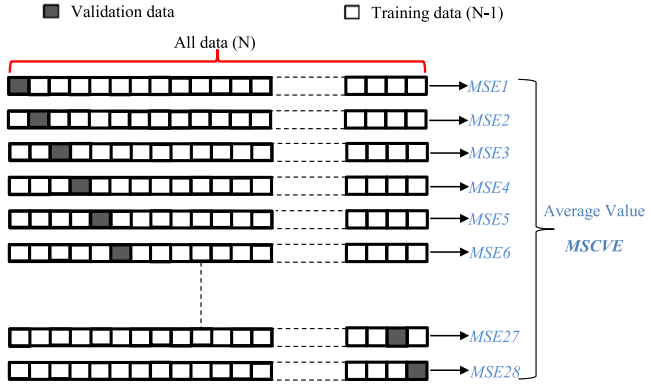

Fig. 4. Leave one out cross validation diagram.

3.1. Data pre-processing

The preprocessing phase is used to make the training of any machine learning more effective and successful; in the current work before training the networks, all data were normalized (their values will be in the same range between 0 and 1) using the following equation:

$$Normalized\ data = \frac{Original\ data - \mathrm{Min}\ of\ the\ original\ data}{standard\ deviation}. \tag{6}$$

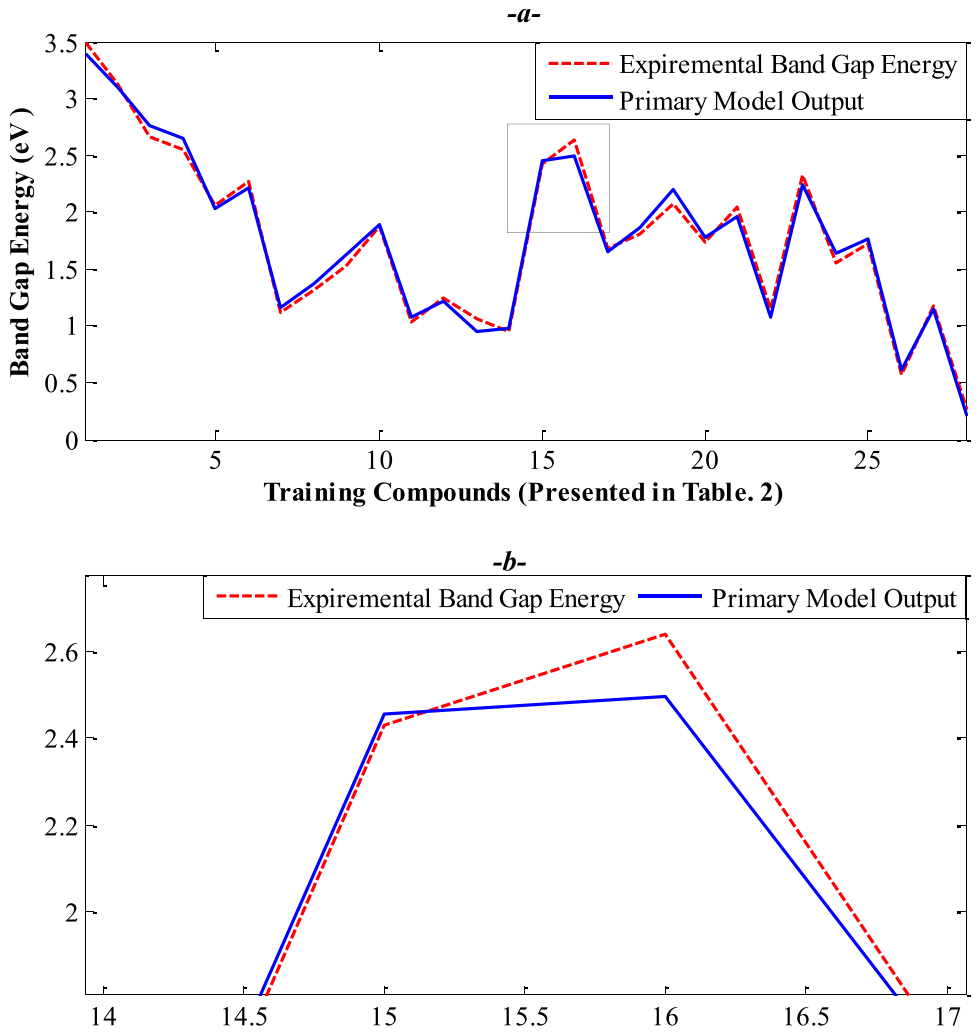

Fig. 5. PSO-based primary ANN-prediction system: (*a*) Primary ANN-prediction system output, **(*b*)** Zoomed segment of primary ANN- prediction system output.

3.2. Primary prediction system identification

In This phase, the input-output data set is used in the training phase to identify the primary ANN-prediction system (Fig. 1). Feedforward neural network is the kind of neural network employed in this research. The structure of the primary prediction system is mostly an online adaption of the feedforward neural network employed. The particle swarm optimization algorithm adjusts the primary ANN-prediction system's parameters (the parameters optimization block in Fig. 1) until the error between experiments band gaps values (y) and the output of the primary ANN-prediction system ($\hat{y}_p$) reaches its smallest value.

$$e_1 = y - \hat{y}_p; \tag{7}$$

3.3. Error process identification

This paper's major contribution is introducing the concept of prediction by including a new prediction module known as the error model. Let us define the error process e_1 (Eq.(1)) that represents the uncertainties in the primary ANN-prediction system. Using this error, a second ANN-system (ANN-error model) can be designed for the modelling of e_1.

Because the error e_1 is a time series, it is appropriate to design its model using an ANN-Auto regressive model that attempts to predict the new output based on the previous outputs. Thus, we propose to identify e1 by the following ANN-AR model:

$$\hat{e}_{1k} = b_0 + b_1\hat{e}_{1k-1} + b_2\hat{e}_{1k-2} + \ldots + b_n\hat{e}_{1k-n} + \varepsilon_k; \tag{8}$$

Where $\hat{e}_{1k}$ is the estimated value of e_{1k},($\hat{e}_{1k-1}$,$\hat{e}_{1k-2,\ldots}\hat{e}_{1k-n}$) are the previous estimated values of the error process, ε_t is the residual and n is the number of regressors.

The structure of this phase is given in Fig. 2; where the boxes showing delay refer to the previous time series value of the error ($\hat{e}_1$). The task now is to adjust online the parameters of the ANN-Error model using the particle swarm optimization algorithm until the error between the error e_1 and the output of the ANN-Error model $\hat{e}_1$ reaches its smallest value.

$$e_2 = e_1 - \hat{e}_1. \tag{9}$$

3.4. Final prediction system

To obtain a final prediction model (Fig. 3), we should interconnect the primary ANN-prediction system and the ANN-Error model in a parallel configuration. This system will allow us to reduce the error obtained in the primary ANN prediction system and get a novel system with higher accuracy.

4. Results and discussion

To evaluate the efficiency of the proposed system, we use first data consisting of 7 predictor variables for every atom (A, B, 2C); after that,

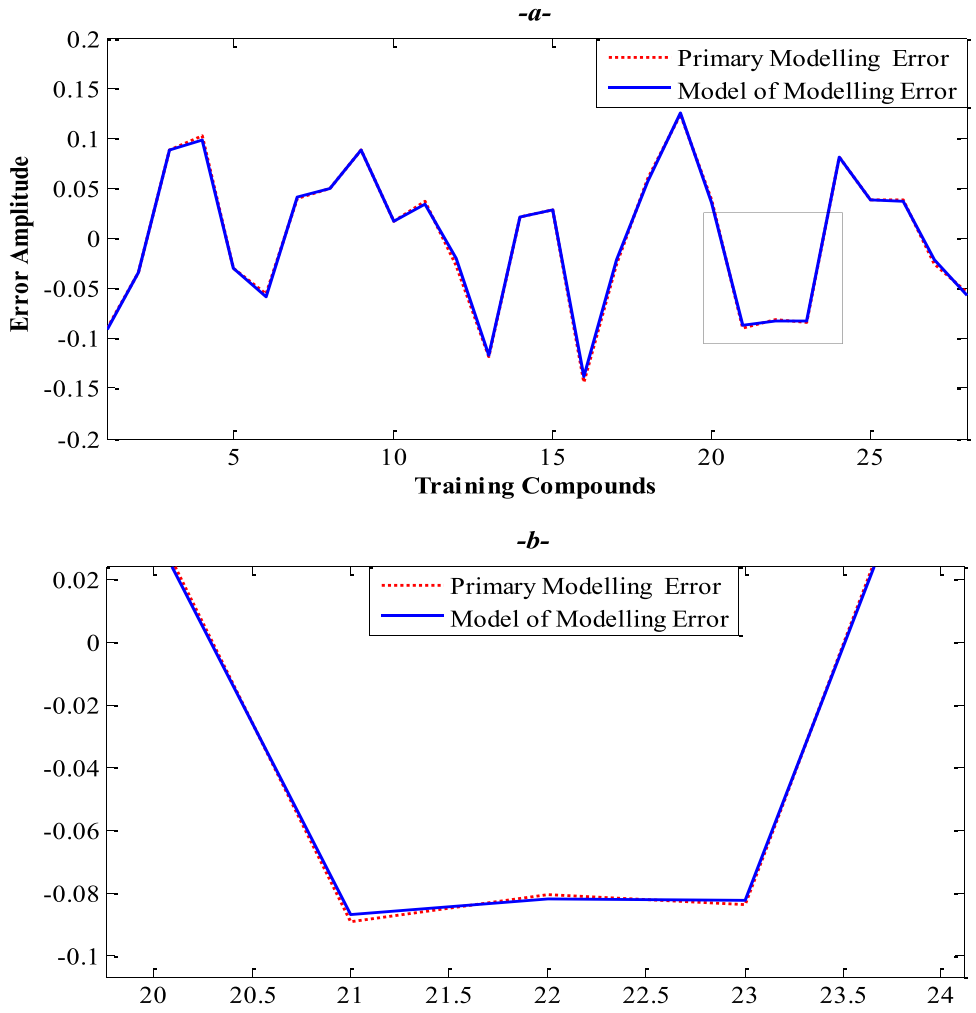

Fig. 6. PSO-based error process model: **(*a*)** Model of modelling error, **(*b*)** Zoomed segment of the model of modelling error.

we apply the optimal features used in the previous work reported in [16–19] to validate the accuracy of our system in terms of prediction errors, then we compare our results to the other benchmark techniques [16–19].

Before testing our approach, we divide our data into two parts. The first one is labeled data, including 46 compounds. We use 60% of them for training and cross-validation data (Trd) and 40% for validation (Tsd); these data are presented in Tables 2 and 3, respectively. The experimental band gap energy of each compound in this part is known. As a result, we use it to calculate the accuracy and investigate the relevance of the features. The second part is the unlabeled data is the largest data, which contains 266 compounds under consideration of unknown band gap energy. We cannot calculate their prediction error; thus, we present the anticipated band gap energy values. The descriptors of all data sets were preprocessed using the normalization technique to remove the outlier's problem and make the neural network more efficient.

To verify the effectiveness of our approach and the applied features, we use the cross-validation technique, a popular method for evaluating the predictive performance of a regression model. We use the training data TRd (28 compounds) set for this step. Given our small label data, we use the leave one out cross-validation technique (LOOCV); this cross-validation approach involves a consideration of each observation as the validation set, while the remaining observations ($N - 1$) are considered as the training set (is the total number of training data TRd=28 compounds). In LOOCV, model fitting is performed and predicted using an observation validation set. We also repeat these times for each observation as a validation set. This model is adjusted and used to predict an observation value. This is a special case of K-fold cross-validation in which the number of folds is the same as the number of observations ($K = N$). Finally, we calculate the overall errors MSCVE to be the overage of the N test. Fig. 4 illustrates the mechanism of LOOCV.

To evaluate the proposed approach's effectiveness, we first create a graphical comparison between the outputs of our suggested method and the experimental results. Thus, Figs. 5, 6 and 7 show the outputs of the various phases of the suggested system. Fig. 5 presents a superposition between the output of the Primary-ANN prediction system and the experiment results. By visually inspecting Fig. 5-***a*** and its zoomed one (Fig. 5-***b***), we can see that the output of the primary ANN-prediction system and the experimental bad gaps results are not well superposed. However, when we add the newly developed ANN-error model (Fig. 6), we notice an improvement in the congruence of the two curves as presented in Fig. 7. Thence, by visually inspecting Figs.5 and 7, we see that the final prediction system output is better than the primary prediction system output. Fig. 8 confirms this fact, showing that the primary error is smaller than the final error.

In the following, we assess the effectiveness of our system quantitatively in terms of the MSCVE criterion (Fig. 3). Table 4 lists the average MSCVE for our method and other benchmark methods to assess the importance of a subset of the features under consideration and the performance of our model. As can be seen, the result of the proposed method based on the 28 features performs better than the three other benchmark methods.

A second experiment is performed using all the data TRd indicates

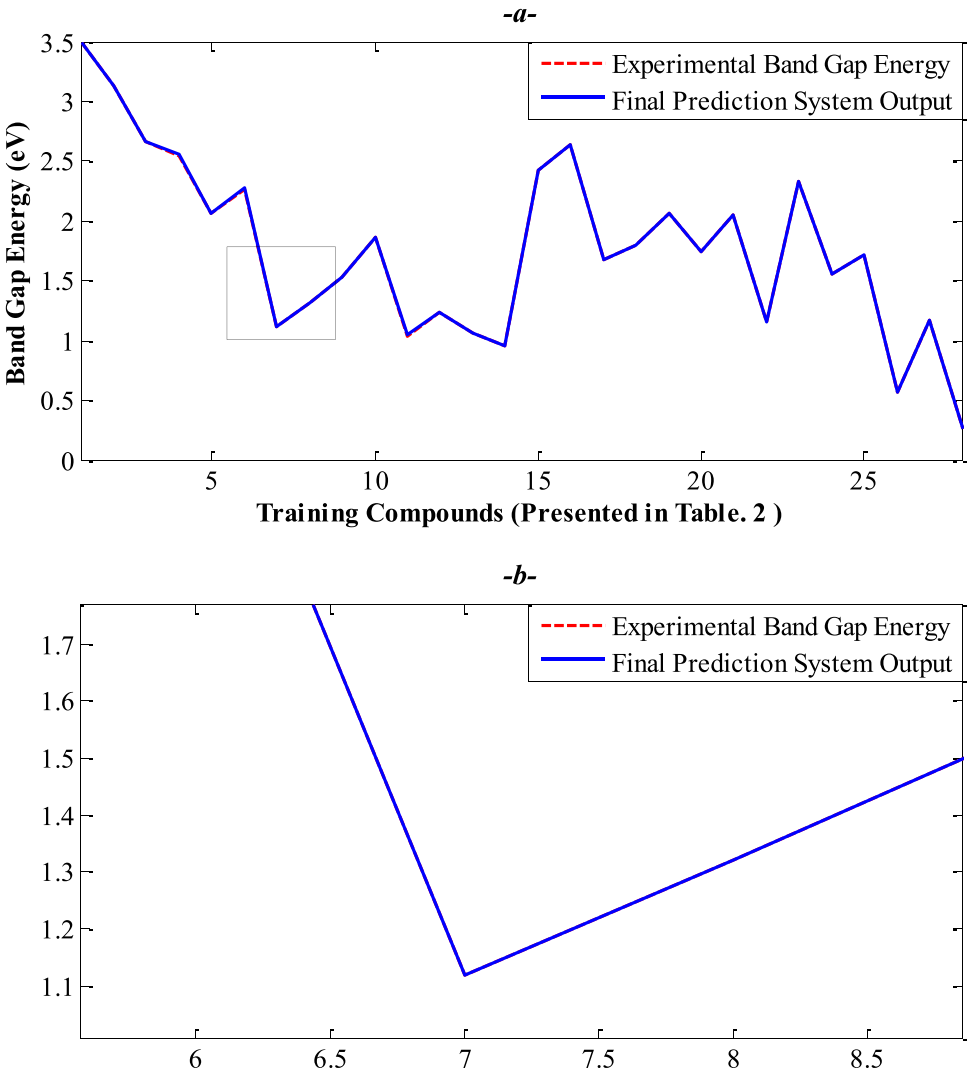

Fig. 7. PSO-based Final ANN-prediction system: **(*a*)** Final ANN-prediction system output, **(*b*)** Zoomed segment of the final ANN-prediction system output.

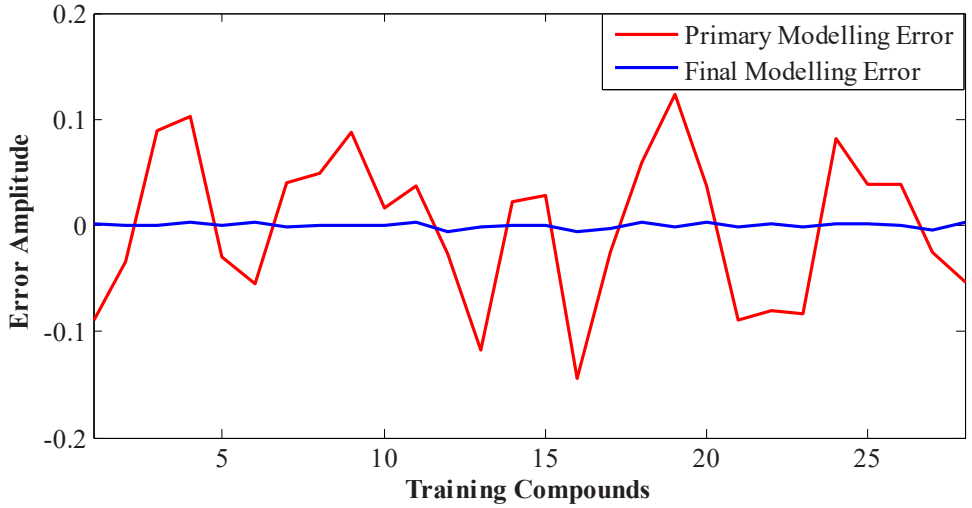

Fig. 8. Primary modelling error VS Final modelling error.

the 28 compounds from training to learn the regression model and the selected feature subset's generalization. Then, we use data containing 18 compounds TSd that are not present in training set for validation. The prediction results for all training data are shown in Fig. 9, where we present the regression curves without (Fig. 9-*a*-) and with (Fig. 9-*b*-) the ANN-error model, to highlight the importance of the introduced ANN-

Table 4
Performance comparison between the proposed method and other benchmark methods in terms of MSCVE(eV^2) criterion.

Models	MSCVE (eV^2)		
	PLS		
Suh C and Rajan[17]	0.1849		
	OLS	PLS	LASSO
Dey et al.[18]	0.501	0.0619	0.0527
Khmaissia et al.[19]	0.0731[a]	0.0750[a]	0.0658[a]
	0.0727[b]	0.0646[b]	0.0593[b]
Proposed method	**0.0454**		

a: MSCVE for the 15 features using by *Khmaissia* et al. [19].
b: MSCVE for the 7 optimal features using by Khmaissia et al. [19].

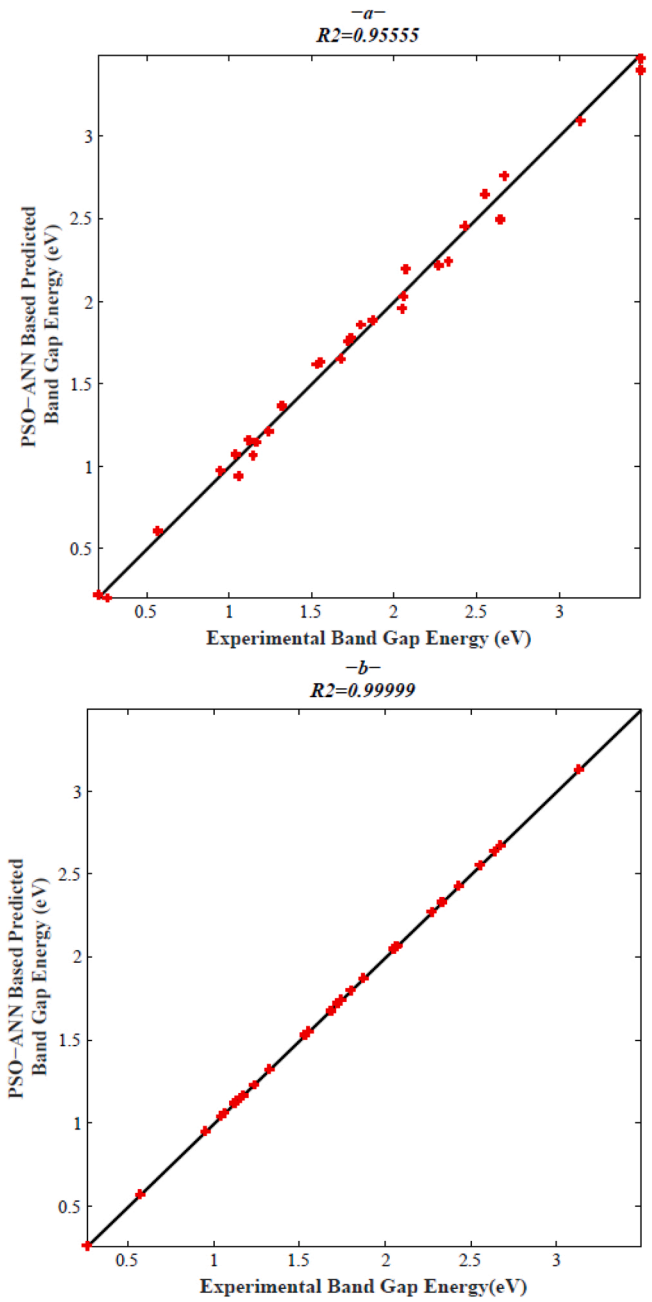

Fig. 9. The PSO-ANN base predicted Band Gaps Energy **Vs** Experimental band gap energy: **(*a*)** Without the ANN-error Model, **(*b*)** With the ANN-error Model.

Table 5
Absolute errors of predicted band gap energy values (eV) and MSE (eV^2) for each of the 18 TS_d compounds, bolded values display errors within approximately a 0.5 eV interval for each of them.

Compounds	Experimental band gaps value (eV)	$\|$**predicted** − **experimental**$\|$(**eV**)
$GaInP_2$	2	**0.24**
$AgGaO_2$	4.1	1.25
$CuAlO_2$	3.5	**0.31**
$CuBO_2$	2.2	0.98
$CuBS_2$	3.61	**0.30**
$CuBSe_2$	3.13	**0.14**
$CuGaO_2$	3.37	**0.20**
$AgBO_2$	2.21	1.02
$AuBS_2$	2.55	**0.27**
$AuBSe_2$	1.53	1.04
$AuBTe_2$	1.33	**0.04**
$HgGeAs_2$	0.2	0.61
$AgAlO_2$	3.6	**0.31**
$CuInO_2$	3.9	0.88
$MgGeAs_2$	1.6	0.63
$MgSiAs_2$	2	**0.55**
$MgSiP_2$	2.3	**0.25**
$ZnGeN_2$	2.67	0.93
MSE(eV^2)	-	**0.50**

error model. To evaluate the performance of the trained model in terms of accuracy score (the percentage of correct predictions made by a model) we have used the coefficient of determination (R^2), which is a number between 0 and 1 that measures how well a model predicts an outcome (the closer is to 1 the better it is), the interpretation of the coefficient of determination can be as following:

- ✓ $R^2 = 0$ means the model does not predict the outcome (Accuracy with low score);
- ✓ $0 < R^2 < 1$ means the model partially predicts the outcome (Accuracy with Medium score);
- ✓ $R^2 = 1$ the model perfectly predicts the outcome (Accuracy with High score).

As a result, the proposed prediction model's accuracy score was high, with a coefficient of determination for the training data $R^2 = 0.9999$ (Fig. 9-*b*-). The presented method's accuracy indicates that the training data used in the model was adequate for describing the band gap energy.

Table 5 lists the absolute errors between the experimental and predicted band gap energy using the proposed technique for each compound in the data test. For solar cells, the range prediction error $\|predicted - experimental\|$ should be smaller than 0.5 eV [18]. As illustrated in Table 5, 55% of the validation data have a predicted error of less than 0.5 eV. This indicates that the training algorithm does not fit the training data completely due to the limited size of the training data. Nevertheless, its performance is better than previous work; the final raw value shows the average squared error MSE of the prediction for all 18 test compounds. It was lesser than other works reported in [18,19].

In the following, we compare the performance of the previous model with the performance of the current work in terms of the *MSE* criterion, using the important properties determined by the various filter selection techniques provided in the previous work [19] listed in Table 6. The proposed framework performs better than the Khmaissia et al. [19] regressions methods. While the mean squared predicted errors of the validation data are reduced for every selected feature, as shown for all six-filter selection strategies. The lowest squared error average is attained when a subset of relevant characteristics is chosen using the correlation criterion technique. However, it did not improve the prediction accuracy significantly; maybe because they did not add the features of the second atom in the third element in ABC_2 compounds; one possible explanation is that all features of each atom in the compounds under consideration must be defined to increase the strength of

Table 6
Performance comparison between the developed system and Khmaissia *et al.* system [19] using the same data in their work.

Filter selection technique	Selected features	MSE (eV^2)			
		Regression methods			
		OLS [19]	**PLS** [19]	**LASSO** [19]	**Proposed method**
15 Features	AN(B), PR(B), AN(C), EN(C), PR(C), MP(B), MP(C), MP(A), PR(A), AN(A),VL(B),VL(C), EN(B), EN(A), VL(A),	19.63	1.05	0.79	**0.73**
Correlation	AN(B), PR(B), AN(C), EN(C), PR(C), MP(C), MP(A)	0.66	0.66	0.63	**0.61**
Lasso regularization	AN(B), PR(B), AN(C), EN(C), PR(A), MP(B), MP(C), MP(A), PR(C), AN(A),VL(B),VL(C),	1.32	2.06	1.32	**1.21**
SFS-OLS	AN(B), PR(B), AN(C), EN(C), PR(C), MP(C), VL(B),VL(C), VL(A)	1.23	-	-	**0.75**
SFS-PLS	AN(B), MP(B), EN(C), PR(C), MP(C), EN(A), AN(C),	-	0.99	-	**0.71**
SFS-Lasso	AN(B), PR(B), AN(C), EN(C), PR(C), MP(C), VL(B),VL(C), VL(A)	-	-	1.32	**1.11**

Table 7
The predicted band gap energy values (eV) for the 266 compounds.

Compounds	Eg (eV)	Compounds	Eg (eV)	Compounds	Eg (eV)	Compounds	Eg (eV)	Compounds	Eg (eV)
$AgBS_2$	3.40	$BeCP_2$	0.82	$MgPbBi_2$	2.51	$ZnPbSb_2$	0.95	$SrSnN_2$	4.52
$AgBSe_2$	2.65	$BeCSb_2$	-0.69	$MgPbN_2$	1.93	$ZnSiBi_2$	1.34	$SrSnP_2$	2.98
$AgBTe_2$	1.19	$BeGeAs_2$	0.35	MgPbP2	2.14	$ZnSiN_2$	2.19	$SrSnAs_2$	1.61
$AgTlS_2$	1.31	$BeGeBi_2$	1.00	$MgPbSb_2$	2.57	$ZnSnBi_2$	0.99	$SrSnSb_2$	0.95
$AgTlSe_2$	1.30	$BeGeN_2$	3.53	$BePbP_2$	1.54	$ZnSnN_2$	3.47	$SrSnBi_2$	0.77
$AgTlTe_2$	1.50	$BeGeSb_2$	1.02	$BePbSb_2$	2.83	$ZnSnSb_2$	0.67	$SrPbN_2$	3.53
$AuAlS_2$	2.95	$BePbAs_2$	1.62	$BeSiBi_2$	1.19	$ZnSnAs_2$	0.15	$SrPbP_2$	2.67
$AuAlSe_2$	2.88	$BePbBi_2$	2.68	$BeSiN_2$	3.50	$ZnSnP_2$	1.43	$SrPbAs_2$	2.43
$AuAlTe_2$	1.35	$BePbN_2$	1.36	$BeSiSb_2$	0.71	$AgBPo_2$	1.08	$SrPbSb_2$	2.11
$AuBO_2$	2.55	$CdSnN_2$	3.63	$BeSnAs_2$	1.84	$CaCN_2$	1.38	$SrPbBi_2$	1.36
$CuTlS_2$	1.49	$CdSnSb_2$	0.23	$BeSnBi_2$	1.35	$CaCP_2$	0.97	$BaCN_2$	0.89
$CuTlSe_2$	1.28	$CuAlPo_2$	1.69	$BeSnN_2$	3.24	$CaCAs_2$	1.78	$BaCP_2$	1.28
$CuTlTe_2$	1.06	$CuBPo_2$	1.31	$BeSnP_2$	1.39	$CaCSb_2$	1.37	$BaCAs_2$	1.60
$AuInS_2$	1.59	$CuInPo_2$	0.93	$BeSnSb_2$	1.65	$CaCBi_2$	1.16	$BaCSb_2$	0.90
$AuInSe_2$	1.18	$CuTlPo_2$	1.02	$CdCAs_2$	0.72	$CaSiN_2$	3.39	$BaCBi_2$	0.91
$AuInTe_2$	1.34	$HgCAs_2$	0.72	$CdCBi_2$	0.71	$CaSiP_2$	2.84	$BaSiN_2$	3.32
$CuBTe_2$	1.44	$HgCBi_2$	-0.26	$CdCN_2$	1.62	$CaSiAs_2$	1.95	$BaSiP_2$	2.27
$CuTlO_2$	3.28	$HgCN_2$	1.09	$CdCP_2$	1.09	$CaSiSb_2$	1.53	$BaSiAs_2$	1.97
$AuTlTe_2$	1.00	$HgCP_2$	0.66	$CdCSb_2$	0.25	$CaSiBi_2$	1.27	$BaSiSb_2$	1.03
$AgGaPo_2$	0.99	$HgCSb_2$	-0.29	$CdGeBi_2$	0.32	$CaGeN_2$	3.65	$BaSiBi_2$	0.99
$AgInO_2$	4.40	$HgGeBi_2$	0.41	$CdGeN_2$	2.86	$CaGeP_2$	3.22	$BaGeN_2$	3.98
$AgInPo_2$	1.36	$HgGeN_2$	3.06	$CdGeSb_2$	0.69	$CaGeAs_2$	1.55	$BaGeP_2$	2.49
$AgTlO_2$	3.35	$HgGeP_2$	1.23	$CdPbAs_2$	0.36	$CaGeSb_2$	0.75	$BaGeAs_2$	1.17
$AgTlPo_2$	1.08	$HgGeSb_2$	-0.35	$CdPbBi_2$	1.49	$CaGeBi_2$	0.57	$BaGeSb_2$	0.52
$AgAlPo_2$	1.11	$HgPbAs_2$	0.66	$CdPbN_2$	2.56	$CaSnN_2$	3.94	$BaGeBi_2$	0.48
$AuAlO_2$	3.52	$HgPbBi_2$	1.36	$CdPbP_2$	0.63	$CaSnP_2$	2.72	$BaSnN_2$	3.97
$AuAlPo_2$	0.60	$HgPbN_2$	2.04	$CdPbSb_2$	0.77	$CaSnAs_2$	2.22	$BaSnP_2$	2.91
$AuBPo_2$	1.29	$HgPbP_2$	1.14	$CdSiBi_2$	1.47	$CaSnSb_2$	1.46	$BaSnAs_2$	1.60
$AuGaO_2$	3.43	$HgPbSb_2$	1.07	$CdSiN_2$	2.18	$CaSnBi_2$	0.50	$BaSnSb_2$	0.97
$AuGaPo_2$	0.78	$HgSiAs_2$	1.15	$CdSiSb_2$	1.49	$CaPbN_2$	3.65	$BaSnBi_2$	1.13
$AuGaS_2$	2.26	$HgSiBi_2$	0.69	$CdSnBi_2$	1.22	$CaPbP_2$	2.62	$BaPbN_2$	2.92
$AuGaSe_2$	1.72	$HgSiN_2$	2.39	$MgSiBi_2$	0.67	$CaPbAs_2$	2.32	$BaPbP_2$	2.78
$AuGaTe_2$	1.85	$HgSiP_2$	2.66	$MgSiN_2$	3.97	$CaPbSb_2$	3.08	$BaPbAs_2$	3.00
$AuInO_2$	4.34	$HgSiSb_2$	0.53	$MgSiSb_2$	1.24	$CaPbBi_2$	1.54	$BaPbSb_2$	2.37
$AuInPo_2$	0.91	$HgSnAs_2$	0.20	$MgSnAs_2$	2.29	$SrCN_2$	0.92	$BaPbBi_2$	0.81
$AuTlPo_2$	0.81	$HgSnBi_2$	0.46	$MgSnBi_2$	1.20	$SrCP_2$	0.94	$RaCN_2$	0.83
$Be_{0.5}C_{0.5}Sb$	-0.74	$HgSnN_2$	3.37	$MgSnN_2$	3.39	$SrCAs_2$	1.76	$RaCP_2$	0.98
$Be_{0.5}Ge_{0.5}As$	0.21	$HgSnP_2$	0.59	$MgSnP_2$	1.73	$SrCSb_2$	0.89	$RaCAs_2$	0.84
$Be_{0.5}Ge_{0.5}P$	1.65	$HgSnSb_2$	0.88	$MgSnSb_2$	1.99	$SrCBi_2$	1.18	$RaCSb_2$	0.74
$Be_{0.5}Ge_{0.5}Sb$	0.68	$MgCAs_2$	0.28	$ZnCAs_2$	0.90	$SrSiN_2$	3.34	$RaCBi_2$	0.74
$Be_{0.5}Si_{0.5}As$	1.08	$MgCBi_2$	0.57	$ZnCBi_2$	0.44	$SrSiP_2$	2.52	$RaSiN_2$	3.03
$Be_{0.5}Si_{0.5}P$	2.05	$MgCN_2$	0.89	$ZnCN_2$	1.91	$SrSiAs_2$	1.88	$RaSiP_2$	2.32
$Be_{0.5}Si_{0.5}Sb$	1.11	$MgCP_2$	0.58	$ZnCP_2$	1.58	$SrSiSb_2$	1.54	$RaSiAs_2$	1.96
$Be_{0.5}Sn_{0.5}As$	1.44	$MgCSb_2$	0.22	$ZnGeBi_2$	0.37	$SrSiBi_2$	1.12	$RaSiSb_2$	0.72
$Be_{0.5}Sn_{0.5}P$	1.53	$MgGeBi_2$	0.76	ZnGeSb2	0.59	$SrGeN_2$	3.83	$RaSiBi_2$	0.79
$Be_{0.5}Sn_{0.5}Sb$	1.72	$MgGeN_2$	3.35	$ZnPbAs_2$	0.16	$SrGeP_2$	2.50	$RaGeN_2$	3.51
$BeCAs_2$	-0.19	$MgGeP_2$	1.84	$ZnPbBi_2$	1.44	$SrGeAs_2$	1.14	$RaGeP_2$	1.85
$BeCBi_2$	-0.46	$MgGeSb_2$	1.23	$ZnPbN_2$	2.72	$SrGeSb_2$	0.58	$RaGeAs_2$	1.15
$BeCN_2$	2.04	$MgPbAs_2$	2.78	$ZnPbP_2$	1.00	$SrGeBi_2$	0.56	$RaGeSb_2$	0.65
$RaGeBi_2$	0.92	$RaSnSb_2$	1.45	$RaPbAs_2$	2.07	$BeSiAs_2$	1.02	$AuTlSe_2$	1.07
$RaSnN_2$	4.21	$RaSnBi_2$	1.36	$RaPbSb_2$	2.32	$BeGeP_2$	1.45	$CuGaPo_2$	1.17
$RaSnP_2$	2.53	$RaPbN_2$	2.66	$RaPbBi_2$	1.05	$AuTlO_2$	3.04	$ZnCSb_2$	0.31
$RaSnAs_2$	1.09	$RaPbP_2$	2.91	$BeSiP_2$	2.28	$AuTlS_2$	2.03	$ZnSiSb_2$	1.30
$CdGeBi_2$	0.61								

the correlation between inputs and output and show the difference between the ABC and ABC_2 structure. They know that the most important features are B and C elements [18,19].

In all this study, 60% of the labelled data was trained with our algorithm and 40% for validation and evaluation. In the second field, we will use all the labelled data for 46 compounds to learn more about the training model and validate the trained predictive model with 266 unlabeled compounds. The drawback of these experiments is that they cannot verify the accuracy of the predicted band gap energy. It just lists the values predicted by the proposed system. We assume that the consistency of the predicted values according to the feature set and regression model could be used as an indicator of the validity of the predicted values. These predicted band gaps will be available to other researchers for further analysis and validation. The results are presented in Table 7.

Since the band gap energy cannot be negative, any negative numbers in the table should be considered invalid and show that the system did not predict well. As a result, the learned regression models cannot predict significant band gaps for all compounds. This indicates that our model needs more labelled training data to make it more robust and learned. On the other hand, in comparison to the results published in [16–19], the developed ANN-PSO model contains 3% of compounds with a negative band gap out of 266 new compounds. Thus, it is reasonable to conclude that the proposed approach outperformed the previous models and reduced the prediction error, but it was still given a negative value. At this point, we conclude that our model needs additional data to continue making accurate predictions.

5. Conclusion

In this study, we proposed a novel approach based on artificial neural networks and particle swarm optimization algorithms to predict the band gap energy for 266 ABC_2 compounds with high resolution. This method will deal with the problem of local minima of the neural networks while preserving the fitting quality by introducing a notion of prediction using a new module called the ANN- error model. Secondly, we showed that we had added the seven features to the second atom of the third element in the ABC2 compounds. Furthermore, features related to the final element's two atoms are necessary for band gap energy prediction. Adding these feature sets and the proposed method reduced the prediction error to 0.50 eV compared to the previously reported results. This result demonstrates that our underside training data set is sufficiently reliable to generate significant predictions. We also tested our approach using the previous works' data set. This step further demonstrates the accuracy of our approach. The prediction errors are improved for all five-feature selections, though they are larger than we acquired in the first study with the original 28 features. Hence, it is reasonable to assume that the features of the second atom for the third element (C) in the compounds under consideration strongly correlate with the band gap energy. In the final section, to maximize the size of the training set, we trained the model on all labelled data (46 compounds). We tested the model on the large unlabeled data (266 compounds) since the real band gap energy of these data is not labelled; the result of this part cannot be validated. We compiled all predicted values. The number of invalid predictions (negative values) decreased to 3% compared to previously reported results. Our future work aspires to augment the data size by considering the data augmentation technique, followed by the repetition of the learning system with other metaheuristics algorithms.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data generated or analyzed during this study are included in this published article. Extended data will be made available on reasonable request.

Acknowledgments

The authors are deeply grateful to all collaborators who participated in this work. One of the authors Dr. Hamza Bennacer gratefully acknowledges to Fadoua Khmaissia from University of Louisville for here valuable and useful discussions. Also, we are grateful to the General Directorate for Scientific Research and Technological Development (DGRSDT) Algeria. This work falls under the PRFU research project N°: A10N01UN280120220006.

References

[1] A. Gani, O. Cheref, M. Ghezali, M. Rabah, A.H. Reshak, Y. Djaballah, A. Belasri, Mechanical stability and optoelectronic behavior of BeXP2 (X= Si and Ge) chalcopyrite, Chin. J. Phys. 64 (2020) 174–182, https://doi.org/10.1016/j.cjph.2020.01.007.

[2] W.H. Liu, W. Zeng, F.S. Liu, B. Tang, Q.J. Liu, X.J. Ma, First-principles analysis of desired inherent photovoltaic functionalities of tetragonal CuAlX2 (X= O, S, Se and Te), J. Solid State Chem. 303 (2021), 122516, https://doi.org/10.1016/j.jssc.2021.122516.

[3] L. Wei, X. Lv, Y. Yang, J. Xu, H. Yu, H. Zhang, J. Zhou, Theoretical investigation on the microscopic mechanism of lattice thermal conductivity of ZnXP2 (X= Si, Ge, and Sn), Inorg. Chem. 58 (7) (2019) 4320–4327, https://doi.org/10.1021/acs.inorgchem.8b03421.

[4] H. BENNACER, S. Berrah, A. Boukortt, M. Ziane, Electronic and optical properties of GaInX2 (X=As, P) from first principles study, Indian J. Pure Appl. Phys. 53 (3) (2015) 181–189, https://doi.org/10.56042/ijpap.v53i3.5836.

[5] H. Bennacer, A. Boukortt, S. Meskine, M. Hadjab, M.I. Ziane, A. Zaoui, First principles investigation of optoelectronic properties of ZnXP2 (X= Si, Ge) lattice matched with silicon for tandem solar cells applications using the mBJ exchange potential, Optik 159 (2018) 229–244, https://doi.org/10.1016/j.ijleo.2018.01.079.

[6] J. Ramanujam, U.P. Singh, Copper indium gallium selenide based solar cells–a review, Energy Environ. Sci. 10 (6) (2017) 1306–1319, https://doi.org/10.1039/C7EE00826K.

[7] J. Keller, K.V. Sopiha, O. Stolt, L. Stolt, C. Persson, J.J. Scragg, M. Edoff, Wide-gap (Ag, Cu)(In, Ga) Se2 solar cells with different buffer materials—a path to a better heterojunction, Prog. Photovolt.: Res. Appl. 28 (4) (2020) 237–250, https://doi.org/10.1002/pip.3232.

[8] C.D. Pemmaraju, T. Archer, D. Sánchez-Portal, S. Sanvito, Atomic-orbital-based approximate self-interaction correction scheme for molecules and solids, Phys. Rev. B 75 (4) (2007), 045101, https://doi.org/10.1103/PhysRevB.75.045101.

[9] R. Shinde, S.S. Yamijala, B.M. Wong, Improved band gaps and structural properties from Wannier–Fermi–Löwdin self-interaction corrections for periodic systems, J. Phys.: Condens. Matter 33 (11) (2020), 115501, https://doi.org/10.1088/1361-648X/abc407.

[10] F. Tran, P. Blaha, Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential, Phys. Rev. Lett. 102 (22) (2009), 226401, https://doi.org/10.1103/PhysRevLett.102.226401.

[11] D. Koller, F. Tran, P. Blaha, Merits and limits of the modified Becke-Johnson exchange potential, Phys. Rev. B 83 (19) (2011), 195134, https://doi.org/10.1103/PhysRevB.83.195134.

[12] B.G. Janesko, T.M. Henderson, G.E. Scuseria, Screened hybrid density functionals for solid-state chemistry and physics, Phys. Chem. Chem. Phys. 11 (3) (2009) 443–454, https://doi.org/10.1039/B812838C.

[13] M.J. Van Setten, F. Weigend, F. Evers, The GW-method for quantum chemistry applications: Theory and implementation, J. Chem. Theory Comput. 9 (1) (2013) 232–246, https://doi.org/10.1021/ct300648t.

[14] L. Reining, The GW approximation: content, successes and limitations, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 8 (3) (2018), e1344, https://doi.org/10.1002/wcms.1344.

[15] L. Weston, C. Stampfl, Machine learning the band gap properties of kesterite I 2− II− IV− V 4 quaternary compounds for photovoltaics applications, Phys. Rev. Mater. 2 (8) (2018), 085407, https://doi.org/10.1103/PhysRevMaterials.2.085407.

[16] Y. Zeng, S.J. Chua, P. Wu, On the prediction of ternary semiconductor properties by artificial intelligence methods, Chem. Mater. 14 (7) (2002) 2989–2998, https://doi.org/10.1021/cm0103996.

[17] C. Suh, A. Rajagopalan, X. Li, K. Rajan, Combinatorial materials design through database science, MRS Online Proc. Libr. (OPL) (2003) 804, https://doi.org/10.1557/PROC-804-JJ9.23.

[18] P. Dey, J. Bible, S. Datta, S. Broderick, J. Jasinski, M. Sunkara, K. Rajan, Informatics-aided bandgap engineering for solar materials, Comput. Mater. Sci. 83 (2014) 185–195, https://doi.org/10.1016/j.commatsci.2013.10.016.

[19] F. Khmaissia, H. Frigui, M. Sunkara, J. Jasinski, A.M. Garcia, T. Pace, M. Menon, Accelerating band gap prediction for solar materials using feature selection and regression techniques, Comput. Mater. Sci. 147 (2018) 304–315, https://doi.org/10.1016/j.commatsci.2018.02.012.
[20] S.A. Kalogirou, Artificial neural networks in renewable energy systems applications: a review, Renew. Sustain. Energy Rev. 5 (4) (2001) 373–401, https://doi.org/10.1016/S1364-0321(01)00006-5.
[21] S. Azizi, M.M. Awad, E. Ahmadloo, Prediction of water holdup in vertical and inclined oil–water two-phase flow using artificial neural network, Int. J. Multiph. Flow. 80 (2016) 181–187, https://doi.org/10.1016/j.ijmultiphaseflow.2015.12.010.
[22] E. Jorjani, S.C. Chelgani, S.H. Mesroghli, Application of artificial neural networks to predict chemical desulfurization of Tabas coal, Fuel 87 (12) (2008) 2727–2734, https://doi.org/10.1016/j.fuel.2008.01.029.
[23] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95-international conference on neural networks, Vol. 4, IEEE, 1995, pp. 1942–1948, https://doi.org/10.1109/ICNN.1995.488968.
[24] T. Lopes, P. Dias, L. Andrade, A. Mendes, An innovative photoelectrochemical lab device for solar water splitting, Sol. Energy Mater. Sol. Cells 128 (2014) 399–410, https://doi.org/10.1016/j.solmat.2014.05.051.
[25] P. Dey, J. Bible, S. Datta, S. Broderick, J. Jasinski, M. Sunkara, K. Rajan, Informatics-aided bandgap engineering for solar materials, Comput. Mater. Sci. 83 (2014) 185–195, https://doi.org/10.1016/j.commatsci.2013.10.016.
[26] Y. Zhang, Bandgap nature of chalcopyrite ZnXP2 (X= Si, Ge, Sn), Comput. Mater. Sci. 133 (2017) 152–158, https://doi.org/10.1016/j.commatsci.2017.03.016.
[27] Y.M. Basalaev, N.I. Gordienok, Energy band structure of be–(c, si, ge, sn)–n2 crystals, Russ. Phys. J. 60 (5) (2017) 900–907, https://doi.org/10.1007/s11182-017-1155-y.
[28] B. Kocak, Y.O. Ciftci, Determination of the basic physical properties of semiconductor chalcopyrite type MgSnT2 (T= P, As, Sb) from first-principles calculations, J. Mater. Res. 31 (10) (2016) 1518–1531, https://doi.org/10.1557/jmr.2016.133.
[29] P. Villars, K. Brandenburg, M. Berndt, S. LeClair, A. Jackson, Y.H. Paod, S. Iwata, Interplay of large materials databases, semi-empirical methods, neuro-computing and first principle calculations for ternary compound former/nonformer prediction, Eng. Appl. Artif. Intell. 13 (5) (2000) 497–505, https://doi.org/10.1016/S0952-1976(00)00028-2.
[30] P. Villars, K. Brandenburg, M. Berndt, S. LeClair, A. Jackson, Y.H. Pao, S. Iwata, Binary, ternary and quaternary compound former/nonformer prediction via Mendeleev number, J. Alloy. Compd. 317 (2001) 26–38, https://doi.org/10.1016/S0925-8388(00)01410-9.
[31] P. Villars, K. Cenzual, J. Daams, Y. Chen, S. Iwata, Data-driven atomic environment prediction for binaries using the Mendeleev number: Part 1. Composition AB, J. Alloy. Compd. 367 (1–2) (2004) 167–175, https://doi.org/10.1016/j.jallcom.2003.08.060.
[32] S.K. Lahiri, K.C. Ghanta, Artificial neural network model with the parameter tuning assisted by a differential evolution technique: The study of the hold up of the slurry flow in a pipeline, Chem. Ind. Chem. Eng. Q. /CICEQ 15 (2) (2009) 103–117, https://doi.org/10.2298/CICEQ0902103L.
[33] R. Kumar, R.K. Aggarwal, J.D. Sharma, Comparison of regression and artificial neural network models for estimation of global solar radiations, Renew. Sustain. Energy Rev. 52 (2015) 1294–1299, https://doi.org/10.1016/j.rser.2015.08.021.
[34] C.W. Kan, L.J. Song, An artificial neural network model for prediction of colour properties of knitted fabrics induced by laser engraving, Neural Process. Lett. 44 (3) (2016) 639–650, https://doi.org/10.1007/s11063-015-9485-7.
[35] S.O. Fadlallah, T.N. Anderson, R.J. Nates, Artificial neural network–particle swarm optimization (ANN-PSO) approach for behaviour prediction and structural optimization of lightweight sandwich composite heliostats, Arab. J. Sci. Eng. 46 (12) (2021) 12721–12742, https://doi.org/10.1007/s13369-021-06126-0.
[36] N.C. Long, P. Meesad, Meta-heuristic algorithms applied to the optimization of type-1 and type 2 TSK fuzzy logic systems for sea water level prediction. 2013 IEEE 6th International Workshop on Computational Intelligence and Applications (IWCIA), IEEE, 2013, pp. 69–74, https://doi.org/10.1109/IWCIA.2013.6624787.
[37] J. Du, Y. Liu, Y. Yu, W. Yan, A prediction of precipitation data based on support vector machine and particle swarm optimization (PSO-SVM) algorithms, Algorithms 10 (2) (2017) 57, https://doi.org/10.3390/a10020057.
[38] X. Zhang, P. Wang, D. Liang, C. Fan, C. Li, A soft self-repairing for FBG sensor network in SHM system based on PSO–SVR model reconstruction, Opt. Commun. 343 (2015) 38–46, https://doi.org/10.1016/j.optcom.2014.12.079.
[39] Z. Liu, J. Lu, P. Zhu, Lightweight design of automotive composite bumper system using modified particle swarm optimizer, Compos. Struct. 140 (2016) 630–643, https://doi.org/10.1016/j.compstruct.2015.12.031.
[40] W. Tao, Z. Liu, P. Zhu, C. Zhu, W. Chen, Multi-scale design of three dimensional woven composite automobile fender using modified particle swarm optimization algorithm, Compos. Struct. 181 (2017) 73–83, https://doi.org/10.1016/j.compstruct.2017.08.065.
[41] H.M. Hasanien, Particle swarm design optimization of transverse flux linear motor for weight reduction and improvement of thrust force, IEEE Trans. Ind. Electron. 58 (9) (2010) 4048–4056, https://doi.org/10.1109/TIE.2010.2100338.
[42] S.O. Fadlallah, T.N. Anderson, R.J. Nates, Artificial neural network–particle swarm optimization (ANN-PSO) approach for behaviour prediction and structural optimization of lightweight sandwich composite heliostats, Arab. J. Sci. Eng. 46 (12) (2021) 12721–12742, https://doi.org/10.1007/s13369-021-06126-0.
[43] T.O. Owolabi, Extreme learning machine and swarm-based support vector regression methods for predicting crystal lattice parameters of pseudo-cubic/cubic perovskites, J. Appl. Phys. 127 (24) (2020), 245107, https://doi.org/10.1063/5.0008809.