



Research article

Data-driven approaches to study the spectral properties of chemical structures

Ibtisam Masmali ^a, Muhammad Faisal Nadeem ^{b,*}, Zeeshan Saleem Mufti ^c, Ali Ahmad ^d, Ali N.A. Koam ^a, Haleemah Ghazwani ^a^a Department of Mathematics, College of Science, Jazan University, Jazan, 45142, Saudi Arabia^b Department of Mathematics, COMSATS University Islamabad Lahore Campus, Lahore, 54000, Pakistan^c Department of Mathematics and Statistics, The University of Lahore, Lahore, 54000, Pakistan^d Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan 45142, Kingdom of Saudi Arabia

ARTICLE INFO

Keywords:

Predictive modeling
Machine learning
Bismuth tri-iodide
Benzene ring
Energy
Data-driven methodologies
Eigenvalues

ABSTRACT

The molecular energy, which is the sum of all eigenvalues, is crucial in determining the total π -electron energy of conjugated hydrocarbon molecules. We used machine learning techniques to calculate the energy, inertia, nullity, signature, and Estrada index of molecular graphs for bismuth tri-iodide and benzene rings embedded in P-type surfaces within 2D networks. We applied MATLAB to extract the actual eigenvalues from the data and developed general equations for these molecular properties. We then used these equations to estimate the values and compared them to the actual values through graphical analysis. Our results demonstrate the potential of data-driven techniques in predicting molecular properties and enhancing our understanding of spectral theory.

1. Introduction

In the continuously developing discipline of chemistry, a comprehensive understanding of molecular energy and spectral characteristics is imperative for predicting the behavior and interactions of various molecules, such as conjugated hydrocarbon molecules. These molecules exhibit unique chemical properties, like increased stability and reactivity, due to the delocalization of π electrons throughout their structure [1–3]. The concept of molecular energy, which is the total sum of all eigenvalues, is central to studying these molecules. This concept is closely linked to spectral theory, a branch of mathematics that examines the relationship between eigenvalues and eigenvectors of linear operators. Spectral theory provides valuable insights into molecular properties and interactions by analyzing and classifying molecular structures based on their energy and spectral characteristics [4–6].

Researchers increasingly adopt interdisciplinary approaches to tackle complex scientific questions as chemistry advances [7]. Data science and machine learning have emerged as powerful tools to revolutionize the study of molecular properties [8]. These techniques provide a robust framework for analyzing extensive and complex datasets, identifying patterns, and building predictive models that significantly improve our understanding of molecular behavior [9,10]. Machine learning, a subset of artificial intelligence, can train algorithms to recognize patterns and relationships within data, leading to predictions or conclusions regarding molecular behavior. These algorithms can be applied to various tasks, such as predicting molecular energy, determining chemical reactivity, and

* Corresponding author.

E-mail addresses: iamasmali@jazanu.edu.sa (I. Masmali), mfaisalnadeem@gmail.com (M.F. Nadeem), zeeshansaleem009@gmail.com (Z.S. Mufti), ahmadms@gmail.com (A. Ahmad), akoum@jazanu.edu.sa (A.N.A. Koam), hqghazwani@jazanu.edu.sa (H. Ghazwani).

identifying potential drug candidates [11–13].

In spectral theory and molecular energy, data science and machine learning can be used to develop efficient computational methods for estimating molecular properties. By harnessing these tools, researchers can overcome the limitations of traditional approaches, which often depend on time-consuming and labor-intensive experiments. Additionally, these techniques enable the discovery of new relationships and patterns within data, potentially leading to groundbreaking findings and advancements in the field of chemistry [14–16].

The structure of a chemical compound can be represented by a molecular graph, which can be transformed into various matrices by utilizing several graph properties. The atoms of a structure are connected via a bond, which is directed to the adjacency and distance matrices. The polynomial obtained from the adjacency or distance matrix can be considered a structure's signature. The eigenvalue obtained from the polynomial considered the molecular descriptor and was used in the quantitative structure-property/activity relationship.

Graph of a molecule is a mathematical entity defined as $G = (V, E)$, where V is the set of vertices, also called atoms and E is the set of all the edges of the molecule, also called bonds. Usually, in molecular structure, hydrogen atoms are not contemplated. Adjacency matrices $A = [a_{ij}]$ of molecules are square matrices of order n , and eigenvalues $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_{q-1}, \lambda_q$ of A are called the eigenvalues of the molecular structure. The absolute sum of all the eigenvalues is known as the energy of a molecular structure G . Mathematically, it is denoted as $E = \sum_{i=1}^q |\lambda_i|$. The set consisting of eigenvalues is also named as the spectrum of the graph G . The spectral characteristics of the graph have been expansively investigated. There are many applications of graph theory in the field of chemical graph theory [17–19]. Among the many applications of spectral theory in chemistry, one of them is based on the adjacent equivalence between the eigenvalues of the structure and the molecular orbital energy level of electrons in conjugated hydrocarbons [20,21].

Rank and nullity play a vital role in graph theory and are associated with the area of linear algebra. Rank represents the sum of molecular structures' positive and negative inertia index. The nullity of a structure $\eta(G)$ is the number of roots having zero value in the characteristic polynomial of $A(G)$ and represents the stability of the molecular structure. If the molecule is stable, closed-shell means its nullity is zero, whereas if it is unstable, it is highly reactive, and open-shell means its nullity is more significant than zero. Every molecule structure can be expressed as square matrices with only 1's and 0's entries $p(G)$, the positive eigenvalues correspond to the positive inertia index, while $n(G)$, the negative eigenvalues correspond to the negative inertia index.

In recent decades, another important concept known as the Estrada index was introduced by Ernesto Estrada and defined as $EE(G) = \sum_{i=1}^n e^{\lambda_i}$. Initially, it was applied to quantify the degrees of folding of long-chain molecular structures, particularly proteins. In a continuation of Estrada index's applications, many studies have been done [22–27].

By the motivation of the above mathematical concepts, the structure of the molecule and its optimal properties are measured through these concepts in the current study. For the investigation, two famous molecules, Bismuth tri-iodide, and benzene, are considered due to their huge applications in chemistry, chemical engineering, and other fields of science. We measure the energy and Estrada index of these structures. In addition, we have also calculated the inertia, nullity, and signature of the molecules. This study aims to minimize the error between molecular graphs' exact and estimated values through polynomial curve fitting. We focus on two specific structures: Bismuth Tri-iodide (BiI₃) and a benzene ring embedded in a p-type surface. By comparing the exact and estimated energy and Estrada index for each structure, we aim to demonstrate the effectiveness of our methodology. To achieve this, we employ a multi-step computational process using various software tools.

2. Brief description of Bismuth Tri-iodide (BiI₃)

Bismuth tri-iodide (BiI₃) is an inorganic structure that is produced by the chemical reactions of iodine and bismuth; this motivated the interest of qualitative studies [28]. BiI₃ is extremely helpful in subjective inorganic investigations. It was experimentally shown that Bi-doped glass optical strands are among the most capable energetic laser media. Various types of Bi-doped fiber strands are formed and depleted to make Bi-doped fiber lasers and optical loudspeakers [29]. BiI₃ is a structure consisting of three layers, such that a bismuth atom is packed in between iodide particles to form a repeated $I - Bi - I$ plane [30]. Each monolayer unit of BiI₃ is stacked with each other via Vander Walls forces [31]. This structure provides ideal 2D material for photovoltaic cells, optoelectronics, and ambient temperature X-rays/gamma rays detectors [32]. These stacking patterns and interlayer distance affect the electronic structure and stability [33]. These electronic properties are modified by intercalation, chemical doping, and mechanical strength [34,35]. Particularly, the optical properties are significantly affected by the interlayer distance. The BiI₃ forms the material for photodetectors with excellent durability and stability with different bending strains, making them suitable for flexible devices such as for optoelectronics with advanced technologies like optical fiber communication, flexible imaging technologies, complex environmental monitoring, and wearable light sensors [36]. In addition, BiI₃ also gains attention in gamma-ray detectors or radiographic imaging owing to its strong photon inhibition power due to its high density (5.78 g/cm³), large band gap, and greater effective atomic number. These important characteristics are essential for huge resolutions of room-temperature gamma-ray spectroscopies [37,38]. The tremendous properties of BiI₃ also lead to the tactile applications of smart sensors, photovoltaic cells, human-machine interfacing and photonics. Through the modification of the morphologies the BiI₃ can convert into single and twin plates [39].

2.1. Computational methodology for molecular graph analysis

To measure the energy and Estrada index, different types of computation work are done by different software as shown in Fig. 1. Following the procedure outlined in Fig. 1, we first used *HyperChem* to draw the molecular structure of BiI₃, as depicted in Fig. 2.

Secondly, an adjacency matrix of the molecular graph is constructed by using TopoCluj. Third, the matrix's eigenvalues are calculated with Matlab's help. Finally, a polynomial curve of degree two is built through the eigenvalues attained from the adjacency matrix of the molecular structure using cf Toolbox in Matlab.

2.2. Energy and Estrada index of BiI_3

The 2 s-order polynomials which display the energy and Estrada index of the Bismuth Tri-iodide molecule are given by Eqs. (1) and (2) respectively.

$$E(B) = 7 \times 10^{-9}m^2n^2 - 9 \times 10^{-5}n^2m + 0.0003n^2 - 0.0003nm^2 + 13.794nm + 9.5506n - 0.0014m^2 + 5.1553m - 0.3239 \quad (1)$$

$$EE(B) = 3.167m^2n^2 - 15.839n^2m + 19.009n^2 - 12.67nm^2 + 110.86nm - 44.392n + 9.49m^2 - 32.462m + 57.857 \quad (2)$$

Where m is horizontal and n is vertical unit cells of BiI_3 .

Eqs. (1) and (2) can further be written in the form of coefficients as shown in Eqs. (3) and (4):

$$E(B) = \begin{cases} n^2 : & 7 \times 10^{-9}m^2 - 9 \times 10^{-5}m + 0.0003 \\ n : & -0.0003m^2 + 13.794m + 9.5506 \\ 1 : & -0.0014m^2 + 5.1553m - 0.3239 \end{cases} \quad (3)$$

$$EE(B) = \begin{cases} n^2 : & 3.167m^2 - 15.839m + 19.009 \\ n : & -12.67m^2 + 110.86m - 44.392 \\ 1 : & 9.49m^2 - 32.462m + 57.857 \end{cases} \quad (4)$$

The numerical results for the energy of and Estrada index of BiI_3 through Eqs (1) and (2) are calculated at different values of m in Table 1.

Further, estimated values of energy and Estrada are computed at different numbers of unit cells by using Eqs. (1) and (2) and compared with exact values of energy and Estrada index obtained through constructing the adjacency matrix, as shown in Figs. 3 and 4.

Here, exact values are denoted by a blue dotted line and estimated values by an orange line. These figures show a good agreement between exact and estimated values. For the close view, we also calculate the Mean Absolute percentage error between these values in Tables 2 and 3. The relative error is important because it gives us a valuation of the accuracy of calculations or projections. This allows us to analyze the method we use to identify areas for potential improvements.

Another way to support this study is to use another statistical method in which we first find the mean absolute error and the standard deviation of the errors across all data points to understand the overall accuracy of the estimation method. The average absolute error for the above data is 0.0763, and the standard deviation is 0.02423519. The normal distribution curve of the data given in Table 2 is shown in Fig. 5.

The average absolute error for the above data is 0.0855 and the standard deviation is 0.09604087671403254. The normal distribution curve of the data given in Table 3 is shown in Fig. 6.

In Tables 2 and it is noticed that the exact/actual values of the energy of BiI_3 are smaller than the value of energy of BiI_3 obtained from the quadratic equations, i.e. $E_{ext}(B) < E_{est}(B)$, where, we denote E_{ext} exact and E_{est} the estimated values of energy. Moreover, errors are positive among these values of energy, i.e. Error > 0 . Similarly, in Table 3, we have noted that actual values of Estrada index of BiI_3 are always more than the estimated values, i.e. $EE_{ext}(B) > EE_{est}(B)$, where, EE_{ext} and EE_{est} represents the exact and estimated value of Estrada index, respectively. We find that the mean absolute percentage error of energy of BiI_3 is 0.035 and mean absolute percentage error of Estrada index of BiI_3 is 0.00515. This error analysis shows that the relative errors are generally very small, indicating that the estimated values $EE_{est}(B)$ are quite close to the exact values $EE_{ext}(B)$. This suggests that the estimation method is accurate for this dataset, with errors typically less than 0.1 % relative to the exact values.

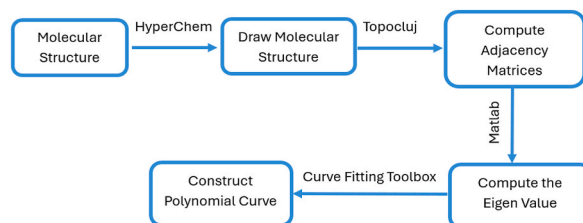


Fig. 1. Procedure to calculate the energy and Estrada index of molecular structures.

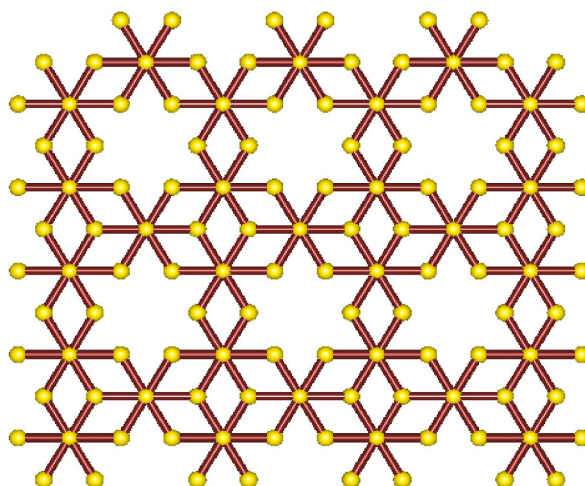


Fig. 2. Bismuth tri-iodide.

Table 1
The quadratic equations of the energy and Estrada index of BiI_3 .

(m, n)	Energy($E(B)$)	Estrada Index($EE(B)$)
$(1, n)$	$2.10 \times 10^{-4}n^2 + 23.34n + 4.83$	$6.342n^2 - 176.20n + 34.88$
$(2, n)$	$1.28 \times 10^{-4}n^2 + 19.86n + 9.98$	$0.009n^2 - 103.35n + 30.91$
$(3, n)$	$3.00 \times 10^{-5}n^2 + 50.92n + 15.12$	$-99.98n^2 + 244.16n - 154.05$
$(4, n)$	$-5.00 \times 10^{-5}n^2 + 64.72n + 20.27$	$-293.64n^2 + 866.33n - 520.03$
$(5, n)$	$-1.49 \times 10^{-4}n^2 + 78.51n + 25.44$	$-580.96n^2 + 1763.16n - 1067.02$
$(6, n)$	$-2.39 \times 10^{-4}n^2 + 92.30n + 30.55$	$-961.94n^2 + 2934.65n - 1795.01$
$(7, n)$	$-3.29 \times 10^{-4}n^2 + 106.12n + 35.69$	$-1436.58n^2 + 4380.80n - 2704.01$
$(8, n)$	$-4.19 \times 10^{-4}n^2 + 119.92n + 40.82$	$-2004.89n^2 + 6101.61n - 3794.02$
$(9, n)$	$-5.09 \times 10^{-4}n^2 + 133.72n + 45.96$	$-2666.85n^2 + 8097.08n - 5065.03$
$(10, n)$	$-5.99 \times 10^{-4}n^2 + 147.52n + 51.08$	$-3422.48n^2 + 10367.21n - 6517.05$

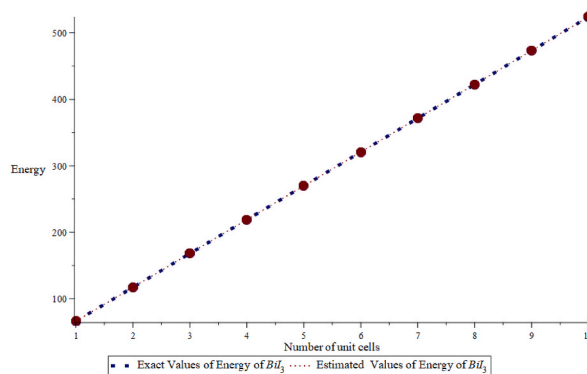


Fig. 3. Comparison of exact and estimated values of energy of BiI_3 .

2.3. The inertia, nullity and signature of BiI_3

This section analyzes the molecular structure of BiI_3 and its stability through optimal properties. For this, we calculate the numerical results of inertia, nullity and signature of BiI_3 in Table 4. In Table 4, $p(B)$ shows the positive inertia index whereas $n(B)$ represents the negative inertia index. When the vertical unit cells n are increased at constant m horizontal unit cells, it is found the balance between positive and negative inertia indexes. The difference between positive and negative eigenvalues is called the signature $s(B)$ and found no difference due to balanced behavior of inertia indexes. The results for nullity $\eta(B)$ are obtained by accounting the eigenvalues having zero value in the characteristic polynomial. The nullity of BiI_3 is increased with increasing of vertical unit cells of structure as shown in Table 4.

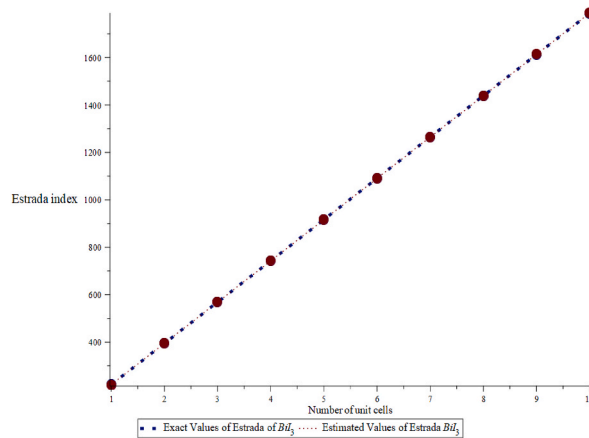


Fig. 4. Exact and estimated comparison of Estrada index of BiI₃.

Table 2

The exact values $E_{ext}(B)$ and estimated values $E_{est}(B)$ of the energy of BiI₃.

(m, n)	$E_{ext}(B)$	$E_{est}(B)$	Error	Absolute Error	Absolute Percentage Error	Relative Error (%)
(3,1)	66.950	66.966	0.016	0.016	0.023	$\frac{0.016}{66.966} \times 100\% = 0.02\%$
(3,2)	116.924	116.989	0.065	0.065	0.056	$\frac{0.065}{116.989} \times 100\% = 0.05\%$
(3,3)	167.849	167.919	0.070	0.070	0.041	$\frac{0.070}{167.919} \times 100\% = 0.04\%$
(3,4)	218.775	218.849	0.074	0.074	0.033	$\frac{0.074}{218.849} \times 100\% = 0.03\%$
(3,5)	269.701	269.779	0.078	0.078	0.029	$\frac{0.078}{269.779} \times 100\% = 0.03\%$
(3,6)	320.626	320.709	0.082	0.082	0.026	$\frac{0.082}{320.709} \times 100\% = 0.03\%$
(3,7)	371.552	371.640	0.088	0.088	0.024	$\frac{0.088}{371.640} \times 100\% = 0.02\%$
(3,8)	422.478	422.570	0.091	0.091	0.022	$\frac{0.091}{422.570} \times 100\% = 0.02\%$
(3,9)	473.403	473.500	0.097	0.097	0.020	$\frac{0.097}{473.500} \times 100\% = 0.02\%$
(3,10)	524.328	524.431	0.102	0.102	0.019	$\frac{0.102}{524.431} \times 100\% = 0.02\%$

Mean Absolute Percentage Error (MAPE) 0.029.

Table 3

Exact values $EE_{ext}(B)$ and estimated values $EE_{est}(B)$ of the Estrada index of BiI₃.

(m, n)	$EE_{ext}(B)$	$EE_{est}(B)$	Error	Absolute Error	Absolute Percentage Error	Relative Error (%)
(3,1)	215.519	215.515	-0.004	0.004	0.0018	$\frac{-0.004}{215.519} \times 100\% = -0.002\%$
(3,2)	427.787	427.780	-0.007	0.007	0.0016	$\frac{-0.007}{427.787} \times 100\% = -0.002\%$
(3,3)	640.066	640.054	-0.012	0.012	0.0018	$\frac{-0.012}{640.066} \times 100\% = -0.002\%$
(3,4)	852.338	852.341	-0.002	0.002	0.0002	$\frac{-0.002}{852.338} \times 100\% = -0.0002\%$
(3,5)	1064.611	1064.634	-0.023	0.023	0.0021	$\frac{-0.023}{1064.611} \times 100\% = -0.002\%$
(3,6)	1276.885	1276.940	-0.055	0.055	0.0043	$\frac{-0.055}{1276.885} \times 100\% = -0.004\%$
(3,7)	1489.158	1489.256	0.098	0.098	0.0066	$\frac{0.098}{1489.158} \times 100\% = 0.007\%$
(3,8)	1701.432	1701.583	0.151	0.151	0.0089	$\frac{0.151}{1701.432} \times 100\% = 0.009\%$
(3,9)	1919.705	1913.920	0.215	0.215	0.0112	$\frac{0.215}{1919.705} \times 100\% = 0.01\%$
(3,10)	2125.979	2126.267	0.288	0.288	0.013	$\frac{0.288}{2125.979} \times 100\% = 0.01\%$

Mean Absolute Percentage Error (MAPE) 0.0343.

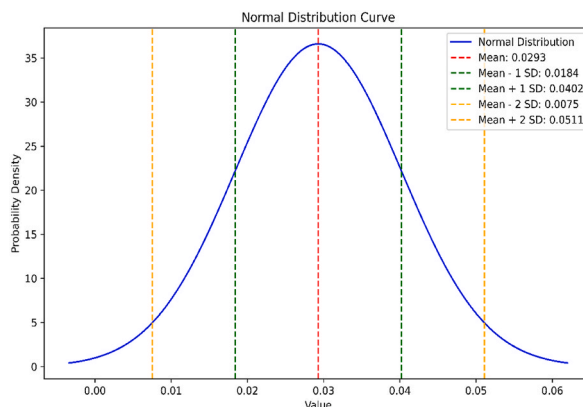


Fig. 5. Normal distribution curve for the absolute errors in energy estimation of BiI₃.

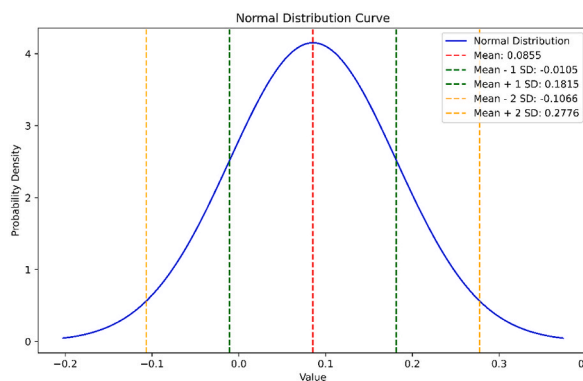


Fig. 6. Normal distribution curve for the absolute errors in Estrada index estimation of BiI₃.

Table 4

The inertia, nullity and signature of BiI₃.

(m, n)	$p(B)$	$n(B)$	$\eta(B)$	$s(B)$
(3,1)	16	16	40	0
(3,2)	26	26	60	0
(3,3)	36	36	80	0
(3,4)	46	46	100	0
(3,5)	56	56	120	0
(3,6)	66	66	140	0
(3,7)	76	76	160	0
(3,8)	86	86	180	0
(3,9)	96	96	200	0
(3,10)	106	106	220	0
(3,11)	116	116	240	0
(3,12)	126	126	260	0

3. Benzene Ring Embedded (BRE) in P-type-surface

P-type networks are embeddings of sp^2 carbons in triply periodic surfaces with the same regularity of single-node simple cubic Bravais tilings. It linked among the 230 symmetry classes of Euclidean space. In these embeddings, the edges of the structure are without crossings, and it splits the space into two disjoint regions. A molecular structure consists of entirely sp^2 atoms, is embeddable in a triply periodic surface, with nonpositive Gaussian arc. These types of carbon structures are called Schwarzites. Schwarzites have exceptional electronic, magnetic, and optical characteristics. The Shwarzites, which are embedded in P-type surfaces, decorate the Bravais lattice in three-dimensional Euclidean space. P-type surfaces can be filled with various coverings of polygons having more sides than hexagons, which are required to create the negative Gaussian curvature.

3.1. Energy and Estrada index of BRE

The same procedure is followed to measure the energy and Estrada index of BRE as for BiI₃. The structure of BRE is constructed through HyperChem, as shown in Fig. 7.

The results of energy and Estrada index in the form of 2 s-order polynomials are shown in Eqs. (5) and (6) respectively.

$$E(BRE) = 2 \times 10^{-7} m^2 n^2 - 2 \times 10^{-5} n^2 m + 2 \times 10^{-5} n^2 - 0.01 n m^2 + 35.2 n m + 0.98 n - 0.0085 m^2 + 0.9773 m - 0.1176 \quad (5)$$

$$EE(BRE) = 0.0542 m^2 n^2 - 0.1387 n^2 m - 0.0665 n^2 - 0.0145 n m^2 + 70.062 n m + 2.193 n - 0.1985 m^2 + 2.3475 m - 1.994 \quad (6)$$

Eqs. (5) and (6) can further be written in the form of coefficients as shown in Eqs. (7) and (8):

$$E(BRE) = \begin{cases} n^2 : & 2 \times 10^{-7} m^2 - 2 \times 10^{-5} m + 2 \times 10^{-5} \\ n : & -0.01 m^2 + 35.2 m + 0.98 \\ 1 : & -0.0085 m^2 + 0.9773 m - 0.1176 \end{cases} \quad (7)$$

$$EE(BRE) = \begin{cases} n^2 : & 0.0542 m^2 - 0.1387 m - 0.0665 \\ n : & -0.0145 m^2 + 70.062 m + 2.193 \\ 1 : & -0.1985 m^2 + 2.3475 m - 1.994 \end{cases} \quad (8)$$

The numerical set of results for the energy and Estrada index at different values of m by using Eqs. (5) and (6) are shown in following Table 5.

To check the accuracy of the result, we have compared the estimated values of energy and Estrada index which are obtained from Eqs. (5) and (6) with the exact values of energy and Estrada index in Figs. 8 and 9 and found a good agreement between the results.

Here, exact values are denoted by a blue dotted line and estimated values by an orange line. In addition, we calculate the mean absolute percentage error between these values in Tables 6 and 7.

The average absolute error for the above data is 0.3761 and the standard deviation is 0.42581697. The normal distribution curve of the data given above is shown in Fig. 10.

The average absolute error for the above data is 0.0855 and the standard deviation is 0.09604087671403254. The normal distribution curve of the data is shown in Fig. 11.

We have observed that the exact value of the energy of BRE is less than first two terms, the gradual increase from the estimated values of energy of BRE is seen. Similarly, exact value of Estrada index of BRE is less than the estimated values of Estrada index of BRE, for few terms, which later have a sudden change in behavior for the remaining five values. We find that the mean absolute percentage error of energy of BRE is 0.04283 and mean absolute percentage error of Estrada index of BRE is 0.00523.

3.2. The inertia, nullity and signature of BRE

The molecular structure of BRE and its stability are analysis through the numerical results of inertia, nullity and signature in Table 8. In Table 8, $p(BRE)$ and $n(BRE)$ shows the positive and the negative inertia indexes and found a balance between results. The signature of molecular structure is displayed by $s(BRE)$ and found the zero values because of the balance behavior of inertia indexes. In the results of nullity denoted by $\eta(BRE)$, found a constant behavior at all unit cells of the structure as shown in Table 8.

4. Conclusion

We studied the energy, Estrada index, inertia, nullity, signature for BiI₃ and BRE in current investigation. The following inequalities

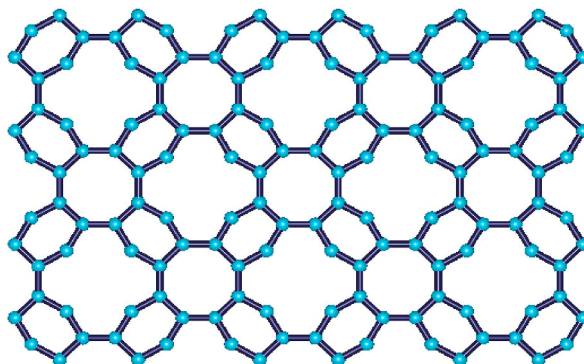


Fig. 7. Benzene ring embedded in P-type surface.

Table 5
The quadratic curves for the Energy and Estrada index of BRE.

(m, n)	Energy($E(BRE)$)	Estrada index($EE(BRE)$)
(1, n)	$\frac{1}{5000000} n^2 + 36.17 n + 0.8682$	$-0.1510 n^2 + 72.2405 n + 0.1550$
(2, n)	$-\frac{3}{156250} n^2 + 71.34 n + 1.8710$	$-0.1271 n^2 + 142.2590 n + 1.9070$
(3, n)	$-\frac{191}{5000000} n^2 + 106.49 n + 2.8908$	$0.0052 n^2 + 212.2485 n + 3.2620$
(4, n)	$-\frac{71}{1250000} n^2 + 141.62 n + 3.9276$	$0.2459 n^2 + 282.2090 n + 4.2200$
(5, n)	$-\frac{3}{40000} n^2 + 176.73 n + 4.9814$	$0.5950 n^2 + 352.1405 n + 4.7810$
(6, n)	$-\frac{29}{312500} n^2 + 211.82 n + 6.0522$	$1.0525 n^2 + 422.0430 n + 4.9450$
(7, n)	$-\frac{551}{5000000} n^2 + 246.89 n + 7.1400$	$1.6184 n^2 + 491.9165 n + 4.7120$
(8, n)	$-\frac{159}{1250000} n^2 + 281.94 n + 8.2448$	$2.2927 n^2 + 561.7610 n + 4.0820$
(9, n)	$-\frac{719}{5000000} n^2 + 316.97 n + 9.3666$	$3.0754 n^2 + 631.5765 n + 3.0550$
(10, n)	$-\frac{1}{6250} n^2 + 351.98 n + 10.5054$	$3.9665 n^2 + 701.3630 n + 1.6310$

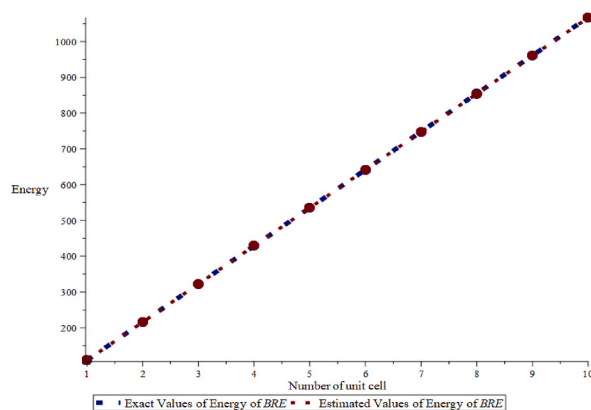


Fig. 8. Exact and estimated comparison of energy of BRE.

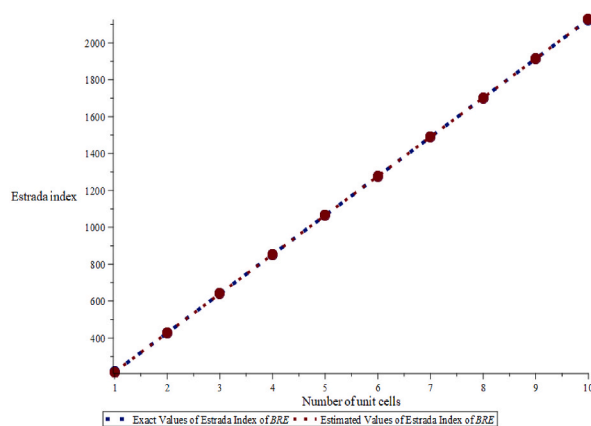


Fig. 9. Exact and estimated comparison of Estrada index of BRE.

$E_{ext}(B) > E_{est}(B)$, $EE_{ext}(B) > EE_{est}(B)$, $E_{ext}(BRE) > E_{est}(BRE)$ and $EE_{ext}(BRE) > EE_{est}(BRE)$ have been observed between exact and estimated values of energy of BiI_3 and BRE. In addition, since the nullity of BiI_3 and BRE is zero so the molecule of these structures is stable and closed shell.

The numerical values of the energy told us about the correlation between the bond energy of π -electrons and every orbital in

Table 6
The exact values $E_{ext}(BRE)$ and estimated values $E_{est}(BRE)$ of the energy for BRE.

(m, n)	$E_{ext}(B)$	$E_{est}(B)$	Error	Absolute Error	Absolute Percentage Error	Relative Error (%)
(3,1)	109.376	109.378	0.0014	0.0014	0.0013	$\frac{0.0014}{109.376} \times 100\% = 0.001\%$
(3,2)	215.887	215.890	0.0036	0.0036	0.0017	$\frac{0.0036}{215.887} \times 100\% = 0.002\%$
(3,3)	322.361	322.360	-0.0010	0.0010	0.0003	$\frac{-0.0010}{322.361} \times 100\% = -0.003\%$
(3,4)	428.834	428.825	-0.0087	0.0087	0.0020	$\frac{-0.0087}{428.834} \times 100\% = -0.002\%$
(3,5)	535.305	535.170	-0.1347	0.1347	0.0252	$\frac{-0.1347}{535.305} \times 100\% = -0.025\%$
(3,6)	641.779	641.651	-0.2677	0.2677	0.0417	$\frac{-0.2677}{641.779} \times 100\% = -0.04\%$
(3,7)	748.264	748.809	-0.4554	0.4554	0.0609	$\frac{-0.4554}{748.264} \times 100\% = 0.06\%$
(3,8)	854.742	854.063	-0.6790	0.6790	0.0794	$\frac{-0.6790}{854.742} \times 100\% = -0.08\%$
(3,9)	961.220	961.275	-0.9456	0.9456	0.0983	$\frac{-0.9456}{961.220} \times 100\% = -0.09\%$
(3,10)	1067.699	1067.444	1.2551	1.2551	0.1175	$\frac{1.2551}{1067.699} \times 100\% = 0.01\%$

Mean Absolute Percentage Error (MAPE) 0.0232.

Table 7
The exact values $EE_{ext}(BRE)$ and estimated values $EE_{est}(BRE)$ of the Estrada Index for BRE.

(m, n)	$E_{ext}(B)$	$E_{est}(B)$	Error	Absolute Error	Absolute Percentage Error	Relative Error (%)
(3,1)	215.519	215.515	-0.004	0.004	0.0019	$\frac{-0.004}{215.519} \times 100\% = -0.002\%$
(3,2)	427.787	427.780	-0.007	0.007	0.0016	$\frac{-0.007}{427.787} \times 100\% = -0.002\%$
(3,3)	640.066	640.054	-0.012	0.012	0.0019	$\frac{-0.012}{640.066} \times 100\% = -0.002\%$
(3,4)	852.338	852.341	-0.002	0.002	0.0002	$\frac{-0.002}{852.338} \times 100\% = -0.0002\%$
(3,5)	1064.611	1064.634	0.023	0.023	0.0022	$\frac{0.023}{1064.611} \times 100\% = 0.002\%$
(3,6)	1276.885	1276.940	-0.055	0.055	0.0043	$\frac{-0.055}{1276.885} \times 100\% = -0.004\%$
(3,7)	1489.158	1489.256	0.098	0.098	0.0066	$\frac{0.098}{1489.158} \times 100\% = 0.007\%$
(3,8)	1701.432	1701.583	0.151	0.151	0.0089	$\frac{0.151}{1701.432} \times 100\% = 0.009\%$
(3,9)	1913.705	1913.920	0.215	0.215	0.0112	$\frac{0.215}{1913.705} \times 100\% = 0.01\%$
(3,10)	2125.979	2125.267	0.288	0.288	0.0135	$\frac{0.288}{2125.979} \times 100\% = 0.01\%$

Mean Absolute Percentage Error (MAPE) 0.00723.

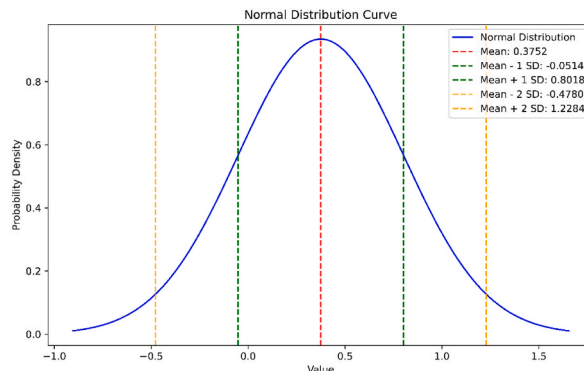


Fig. 10. Normal distribution curve for the absolute errors in energy estimation of BRE.

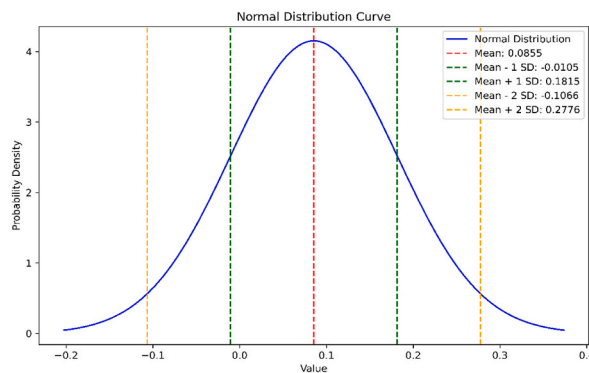


Fig. 11. Normal distribution curve for the absolute errors in Estrada index estimation of BRE.

Table 8

The inertia, nullity and signature of BRE.

(m, n)	$p(BRE)$	$n(BRE)$	$\eta(BRE)$	$s(BRE)$
(3,1)	39	39	2	0
(3,2)	76	76	2	0
(3,3)	113	113	2	0
(3,4)	150	150	2	0
(3,5)	187	187	2	0
(3,6)	224	224	2	0
(3,7)	261	261	2	0
(3,8)	298	298	2	0
(3,9)	335	335	2	0
(3,10)	372	372	2	0
(3,11)	409	409	2	0
(3,12)	446	446	2	0

π -electrons correspond to the each eigen value of the graph under consideration. If we calculate the energy for a single unit, such as (1,1), we obtain a positive integer. As we increase the order horizontally, for example, moving from (1,1) to (1,2), (1,3), (1,4) and so on, the energy of each subsequent unit will also be increased. This is because the number of vertices grows, and consequently, the order of the adjacency matrix also increases. Similarly, we proceeded vertically, for example, from (1,1) to (2,1), (3,1), and so on. As the order increased, handling the calculations became increasingly difficult. Therefore, we generalized our graph and applied statistical methods to analyze the general behavior of the data. To determine the energy of a particular unit, we generated a polynomial of a certain order to estimate the energy of the desired unit. The positive eigenvalues were linked with the antibonding level, negative eigenvalues were linked with bonding levels, and zero eigenvalues were associated with the nonbonding level.

By employing MATLAB to extract actual eigenvalues from the data and generate general equations, we aimed to bridge the gap between the actual and estimated values of these molecular properties. We used the Mean Absolute Percentage Error (MAPE) as the standard metric for error in Table 2, Table 3, Tables 6, and Table 7. When the MAPE between actual and expected values was close to zero, it generally indicated that the model's predictions were accurate on average. Additionally, we observed that the energy of BiI₃ and BRE was 0.035 and 0.04, respectively, from units (3,1) to (3,10). Similarly, we observed that the Estrada index of BiI₃ and BRE was 0.00515 and 0.0523, respectively, from units (3,1) to (3,10). We have used two other methods to support our results. In the first method, we have found the relative error percentage (%) and in the second method, we have found the average absolute error and the standard deviation along with their normal curve. In the first method, the relative error percentage (%) was less than 0.1 %, which suggested that this method was accurate for this dataset, with errors typically less than 0.1 % relative to the exact values. In the second method, a small value of standard deviation indicated that the data points were very close to the mean, reflecting low variability, high consistency, tight distribution, and predictability.

For future research, we propose to conduct comparative studies with other data-driven and other machine learning approaches, such as neural networks, SVM, and decision trees, to evaluate the relative performance and applicability of different methods. Apply the polynomial curve fitting method to a wider range of molecular structures and chemical systems to test its generalizability and robustness. Explore the integration of more advanced machine learning techniques to improve prediction accuracy and computational efficiency. Investigate the use of our methodology in real-world applications, such as drug discovery, materials science, and chemical engineering, to assess its practical utility and impact. Develop a more comprehensive framework that combines various data-driven approaches for a holistic understanding and prediction of molecular properties. The error analysis method we used can be applied to any model. Relative error percentage is scale-independent, making it useful for comparing errors across different datasets or models. It provides a direct interpretation of how large the errors are relative to the actual values, making it easier to understand the model's performance in practical terms. Standard deviation offers a well-rounded understanding of the model's performance, highlighting

different aspects of accuracy and consistency, which can be crucial for model evaluation and improvement.

Funding

This research was funded by Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia through the project number ISP-2024.

Data availability statement

No external data is used, all the data required to understand, evaluate, and reproduce the findings presented in this manuscript are included in the manuscript.

CRediT authorship contribution statement

Ibtisam Masmali: Writing – review & editing, Data curation, Conceptualization. **Muhammad Faisal Nadeem:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Zeeshan Saleem Mufti:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis. **Ali Ahmad:** Writing – review & editing, Validation, Software, Methodology, Data curation. **Ali N.A. Koam:** Writing – review & editing, Validation, Investigation. **Haleemah Ghazwani:** Writing – review & editing, Validation, Software.

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.

Acknowledgement

The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number ISP-2024.

References

- [1] M. Schüler, Y. Pavlyukh, Spectral properties from Matsubara Green's function approach: application to molecules, *Phys. Rev. B* 97 (11) (2018) 115164.
- [2] S. Manivel, B.S. Gangadharappa, N. Elangovan, R. Thomas, O.A.A. Ali, D.I. Saleh, Schiff base (Z)-4-((furan-2-ylmethylene) amino) benzenesulfonamide: synthesis, solvent interactions through hydrogen bond, structural and spectral properties, quantum chemical modeling and biological studies, *J. Mol. Liq.* 350 (2022) 118531.
- [3] J.T. Holaj-Krzak, N. Rezik, N.A. Alsaif, G. Lakshminarayana, Elucidating the infrared spectral properties of succinic molecular acid crystals: illustration of the structure and the hydrogen bond energies of the crystal and its deuterated analogs, *J. Phys. Chem.* 126 (33) (2022) 5604–5620.
- [4] P.W. Langhoff, R.J. Hinde, J.A. Boatz, J.A. Sheehy, Spectral theory of the chemical bond, *Chem. Phys. Lett.* 358 (3–4) (2002) 231–236.
- [5] L.A. Gribov, V.I. Baranov, I.V. Mikhailov, From the theory of spectra to the theory of chemical reactions, in: *Advances in Geochemistry, Analytical Chemistry, and Planetary Sciences: 75th Anniversary of the Vernadsky Institute of the Russian Academy of Sciences*, 2023, pp. 475–483. Cham: Springer International Publishing.
- [6] V. Choudhary, A. Bhatt, D. Dash, N. Sharma, DFT calculations on molecular structures, HOMO–LUMO study, reactivity descriptors and spectral analyses of newly synthesized diorganotin (IV) 2-chloridophenylacetohydroxamate complexes, *J. Comput. Chem.* 40 (27) (2019) 2354–2363.
- [7] J.A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.R. Müller, Tkatchenko A. Combining machine learning and computational chemistry for predictive insights into chemical systems, *Chem. Rev.* 121 (16) (2021) 9816–9872.
- [8] F. Strieth-Kalthoff, F. Sandfort, M.H. Segler, F. Glorius, Machine learning the ropes: principles, applications and directions in synthetic chemistry, *Chem. Soc. Rev.* 49 (17) (2020) 6154–6168.
- [9] N. Artrith, K.T. Butler, F.X. Coudert, S. Han, O. Isayev, A. Jain, A. Walsh, Best practices in machine learning for chemistry, *Nat. Chem.* 13 (6) (2021) 505–508.
- [10] F. Oviedo, J.L. Ferres, T. Buonassisi, K.T. Butler, Interpretable and explainable machine learning for materials science and chemistry, *Acc. Mater. Res.* 3 (6) (2022) 597–607.
- [11] S.M. Moosavi, K.M. Jablonka, B. Smit, The role of machine learning in the understanding and design of materials, *J. Am. Chem. Soc.* 142 (48) (2020) 20273–20287.
- [12] Z. Jiao, P. Hu, H. Xu, Q. Wang, Machine learning and deep learning in chemical health and safety: a systematic review of techniques and applications, *ACS Chemical Health & Safety* 27 (6) (2020) 316–334.
- [13] M. Haghightalari, J. Hachmann, Advances of machine learning in molecular modeling and simulation, *Current Opinion in Chemical Engineering* 23 (2019) 51–57.
- [14] U.W. Liebal, A.N. Phan, M. Sudhakar, K. Raman, L.M. Blank, Machine learning applications for mass spectrometry-based metabolomics, *Metabolites* 10 (6) (2020) 243.
- [15] K.K. Hyde, M.N. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden, D.R. Dixon, E. Linstead, Applications of supervised machine learning in autism spectrum disorder research: a review, *Review Journal of Autism and Developmental Disorders* 6 (2019) 128–146.
- [16] Y. Mao, N. Dong, L. Wang, X. Chen, H. Wang, Z. Wang, I.M. Kislyakov, J. Wang, Machine learning analysis of Raman spectra of MoS₂, *Nanomaterials* 10 (11) (2020) 2223.
- [17] J.R. Boes, O. Mamun, K. Winther, T. Bliigaard, Graph theory approach to high-throughput surface adsorption structure generation, *J. Phys. Chem.* 123 (11) (2019) 2281–2285.
- [18] A.T. Balaban, Applications of graph theory in chemistry, *J. Chem. Inf. Comput. Sci.* 25 (3) (1985) 334–343.
- [19] K.J. Burch, Mathematical physics in theoretical chemistry. Chemical Applications of Graph Theory, Elsevier, 2019, pp. 261–294.
- [20] G.G. Hall, On the eigenvalues of molecular graphs, *Mol. Phys.* 33 (2) (1977) 551–557.
- [21] I. Gutman, D. Kiani, M. Mirzakhah, On incidence energy of graphs, *MATCH Communications in Mathematical and in Computer Chemistry* 62 (3) (2009) 573–580.

- [22] Y. Chen, J. Zhu, A graph theory-based method for regional integrated energy network planning: a case study of a China–US low-carbon demonstration city, *Energies* 12 (23) (2019) 4491.
- [23] S.B. Bozkurt, I. Gutman, Estimating the incidence energy, *MATCH Communications in Mathematical and in Computer Chemistry* 70 (2013) 143–156.
- [24] K.C. Das, I. Gutman, On incidence energy of graphs, *Lin. Algebra Appl.* 446 (2014) 329–344.
- [25] E. Estrada, Characterization of 3D molecular structure, *Chem. Phys. Lett.* 319 (5–6) (2000) 713–718.
- [26] M.A. Malik, R. Farooq, Computational results on the energy and Estrada index of TUC4C8 (R)[m, n] nanotubes, *Optoelectronics and Advanced Materials-Rapid Communications* 9 (1–2) (2015) 311–313.
- [27] G.R. Omid, On the nullity of bipartite graphs, *Graph. Combinator* 25 (1) (2009) 111–114.
- [28] M. Imran, M.A. Ali, S. Ahmad, M.K. Siddiqui, A.Q. Baig, Topological characterization of the symmetrical structure of bismuth tri-iodide, *Symmetry* 10 (6) (2018) 201.
- [29] R.A. Mackay, W. Henderson, *Introduction to Modern Inorganic Chemistry*, CRC Press, 2002 Nov 18.
- [30] L.E. Smart, E.A. Moore, *Solid state chemistry: an introduction*, CRC press (2012 May 29).
- [31] S. Rebouillat, F. Pla, Recent strategies for the development of biosourced-monomers, oligomers and polymers-based materials: a review with an innovation and a bigger data focus, *J. Biomaterials Nanobiotechnol.* 7 (4) (2016) 167–213.
- [32] R.E. Brandt, R.C. Kurchin, R.L. Hoye, J.R. Poindexter, M.W. Wilson, S. Sulekar, F. Lenahan, P.X. Yen, V. Stevanovic, J.C. Nino, M.G. Bawendi, Investigation of bismuth triiodide (BiI₃) for photovoltaic applications, *J. Phys. Chem. Lett.* 6 (21) (2015) 4297–4302.
- [33] S. Sankaran, K. Deshmukh, M.B. Ahamed, S.K. Pasha, *Electrospun polymeric nanofibers: fundamental aspects of electrospinning processes, optimization of electrospinning parameters, properties, and applications*. In *Polymer Nanocomposites in Biomedical Engineering*, Springer, Cham, 2019, pp. 375–409.
- [34] Lintereur AT, Qiu W, Nino JC, Baciak JE. Bismuth tri-iodide radiation detector development. In *Hard X-Ray, Gamma-Ray, and Neutron Detector Physics XI 2009 Sep 17 (Vol. vol. 7449, p. 74491M)*. International Society for Optics and Photonics.
- [35] S. Tongay, W. Fan, J. Kang, J. Park, U. Koldemir, J. Suh, D.S. Narang, K. Liu, J. Ji, J. Li, R. Sinclair, Tuning interlayer coupling in large-area heterostructures with CVD-grown MoS₂ and WS₂ monolayers, *Nano Lett.* 14 (6) (2014) 3185–3190.
- [36] C. Xie, F. Yan, Flexible photodetectors based on novel functional materials, *Small* 13 (43) (2017) 1701822.
- [37] M. Matsumoto, K. Hitomi, T. Shoji, Y. Hiratate, Bismuth tri-iodide crystal for nuclear radiation detectors, *IEEE Trans. Nucl. Sci.* 49 (5) (2002) 2517–2520.
- [38] S.S. Gokhale, H. Han, J.E. Baciak, J.C. Nino, K.A. Jordan, Growth, fabrication, and testing of bismuth tri-iodide semiconductor radiation detectors, *Radiat. Meas.* 74 (2015) 47–52.
- [39] R.E. Brandt, R.C. Kurchin, R.L. Hoye, J.R. Poindexter, M.W. Wilson, S. Sulekar, F. Lenahan, P.X. Yen, V. Stevanovic, J.C. Nino, M.G. Bawendi, Investigation of bismuth triiodide (BiI₃) for photovoltaic applications, *J. Phys. Chem. Lett.* 6 (21) (2015) 4297–4302.