مجلة الدراسات التربوية والعلمية - كلية التربية - الجامعة العراقية
العدد السابع والعشرون - المجلد التاسع - علوم الفيزياء - نيسان 2026 م
doi.org/10.52866/esj.2026.09.27.22

Building Mathematical Models to Determine the Energy Gap Based on Temperature Using Cluster Analysis, the Neville Model, and a Hybrid Model

Enas A. Ahmad [1,a)], Ghassan E. Arif [2, b)], Saleh K. Mezaal [3, c)]
[1]*Department of Mathematics, College of Education for Pure Science, Tikrit University, Iraq*
[2]*Department of Mathematics, College of Education for Pure Science, Tikrit University, Iraq*
[3] *Department of Physics, College of Education for Pure Science, Tikrit University, Iraq*
[a)] *EA230050pep@st.tu.edu.iq* [b)] *ghasanarif@tu.edu.iq* [c)] *kissof_land@hotmail.com*

Abstract :

This research examined three different ways to calculate energy gap values based on temperatures:**Cluster analysis**, the Neville method, and a hybrid model that combines the two methods. Cluster analysis relied on grouping data into similar groups and converting complex equations into linear formulas to facilitate calculations, with an average error of about (0.0315) Volts, with standard deviation (0.0266) Volt. The Neville method, which approximates numerical values using polynomial equations, showed higher accuracy, with errors ranging from (0.01–0.02) Volts, with standard deviation (0.008) Volt. In contrast, the hybrid model, which combined the two methods, was distinguished by using a correction coefficient (L = 0.5), with the highest level of accuracy, where the error ranged from(0.001–0.06) Volts, with standard deviation(0.01573) Volt.

Keywords: cluster analysis, hybrid model, correction coefficient, error range, Neville method

بناء نماذج رياضية لتحديد فجوة الطاقة بناءً على درجة الحرارة باستخدام تحليل المجموعات، ونموذج نيفيل، ونموذج هجين

إيناس أحمد [a،1] ، غسان عارف [b ،2] ، صالح خضير [c ،3]
1. قسم الرياضيات، كلية التربية للعلوم الصرفة، جامعة تكريت، العراق
2. قسم الرياضيات، كلية التربية للعلوم الصرفة، جامعة تكريت، العراق
3. قسم الفيزياء، كلية التربية للعلوم الصرفة، جامعة تكريت، العراق
[a)] EA230050pep@st.tu.edu.iq [b)] ghasanarif@tu.edu.iq [c)] kissof_land@hotmail.com

مستخلص:

تناول هذا البحث دراسة ثلاث طرق مختلفة لحساب قيم فجوة الطاقة بناءً على درجات الحرارة، وهي: التحليل العنقودي، وطريقة نيفيل، ونموذج هجين يجمع بين الطريقتين. اعتمد التحليل العنقودي على تجميع البيانات في مجموعات متشابهة وتحويل المعادلات المعقدة إلى صيغ خطية لتسهيل الحسابات، حيث بلغ متوسط الخطأ فيها حوالي 0.0315 eV، مع انحراف معياري 0.0266 eV. أما طريقة نيفيل، التي تعتمد على تقريب القيم العددية باستخدام معادلات متعددة الحدود، فقد أظهرت دقة أعلى، حيث تراوح الخطأ بين 0.01–0.02 eV، مع انحراف معياري 0.008 eV. في المقابل، تميز النموذج الهجين، الذي دمج بين الطريقتين باستخدام معامل تصحيح ($\lambda = 0.5$)، بأعلى مستوى من الدقة، حيث تراوح الخطأ بين 0.001–0.06 eV، مع انحراف معياري 0.01573 eV.

الكلمات المفتاحية: التحليل العنقودي، النموذج الهجين، معامل التصحيح، مدى الخطأ، طريقة نيفيل .

Introduction

Cluster analysis and the Neville approach are two examples of numerical approximation techniques that play an important part in the process of forecasting material characteristics such as the energy gap in semiconductors. Research has shown that cluster analysis is an excellent method for grouping data points in order to uncover previously unseen patterns. This method also reduces the amount of computer complexity while retaining accuracy [1,3]. In the meanwhile, the Neville technique offers accurate polynomial interpolation, which is especially helpful for estimating the values of the energy gap at different temperatures [6,7]. The results of recent research indicate that hybrid models that combine both strategies produce higher accuracy by balancing local and global factors, hence lowering mistakes in comparison to standalone approaches [9,11]. In the fields of computational physics and materials science, where precise energy gap predictions are necessary for the design of semiconductors and optoelectronic applications, these kinds of research are especially useful. Thermodynamics and quantum mechanics are two examples of scientific fields that need high-precision numerical modeling. These hybrid techniques can also be extended to other scientific disciplines. The purpose of this research is to investigate a unique hybrid model that integrates cluster analysis with the Neville technique in order to improve energy gap forecasts.

Methodology

1) Cluster analysis (cluster analysis)

This technique uses the method of collecting data in the form of groups so that the points within the same group are more similar to each other to be compared with the points with other groups, and this method is used in many fields, including economics, life sciences, marketing, and social sciences, and can be used to recognize images to discover hidden patterns and relationships in the data[1].

2) Methodology, method of work and mathematical model construction

The mathematical model was formulated and the approximate values were calculated by creating a code in Python and calculating the error rate, and the model was chosen after converting the cluster analysis equation into a linear equation as shown in Figure (1), which shows the method of converting to the linear equation necessary to calculate approximation[2].

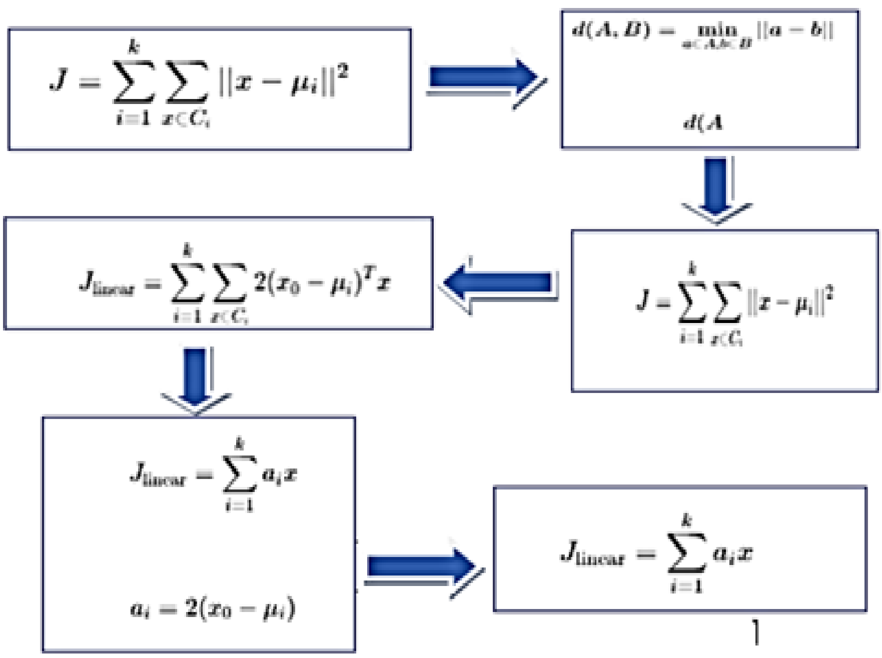

Figure (1) Converting the Cluster Analysis Equation to a Linear Equation

The final equation of cluster analysis as a linear function of the energy and temperature gap becomes as follows:

$$x = AE_g + BT \text{ ----------(1)}$$

Whereas (A,B) coefficients determine the effect of new points on cluster points and are calculated from the following relationship [3]:*A,B*

$$A = \alpha \sum_{i-1}^{k} ai,\ B = \beta \sum_{i=1}^{k} ai - - - (2)$$

After formulating the mathematical code, which provides three important functions as follows:

1) The first: dividing the domain of the function and choosing the new values to perform approximation through mathematical operations.

2) Second: Converting complex functions to simple linearity.

3) Third: Through this function, the relative error and the actual and ap-

proximate value of the integrals are calculated, and the calculation process is done by subtracting the approximate value from the experimental divided by the actual multiplied by (100).

3) Results

Figure (2) represents the method of cluster analysis of numerical approximation.

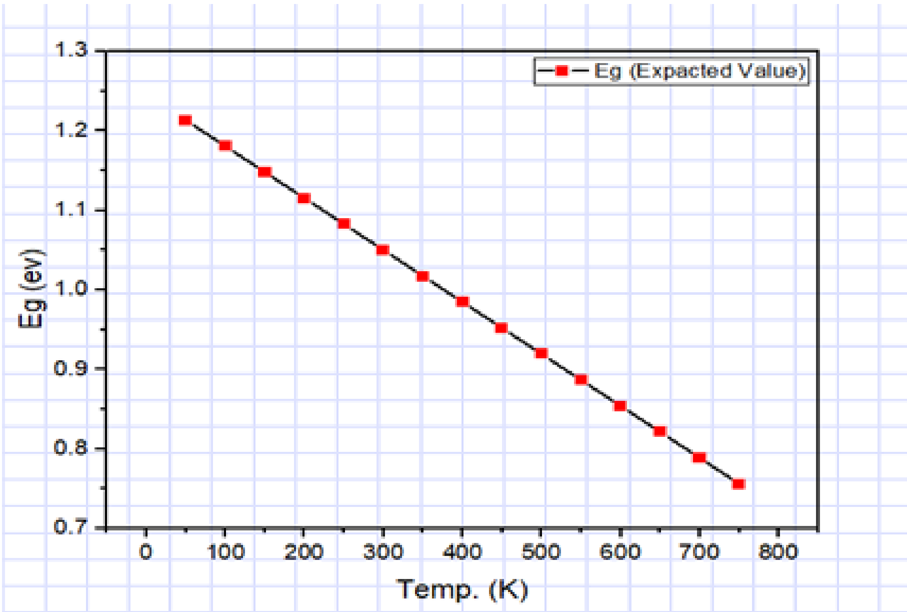

Figure (2) Approximate values calculated by cluster analysis method

The figure shows how to convert complex functions to linear functions and find values that are almost perfect for finding the values (A, B) in the linear equation of this approximation.

Through Table (1), the lowest error rate was (0.08595) and the highest error rate was (0.023952), the model can be considered acceptable because the relative error was relatively little, ranging between (0.016-0.12) and the average value of the error relativity is about (0.0315) and when calculating the standard deviation of this model, it was found that when it becomes within (0.0266 ev) and this is very acceptable from a scientific and mathematical point of view, which reflects the accuracy of the model more[4.5].

Figure (3) represents the relationship between the approximate value and the experimental value, noting the convergence between the approximate and experimental values.

Table (1) Approximate and experimental values of the energy gap calculated by cluster analysis method and error ratio as well.

	Temperature (K)	Experimental Energy Gap (eV) Values	Approximate values (eV)	Absolute error (eV)
1	50	1.25	1.213417	0.036583
2	100	1.22	1.180738	0.039262
3	150	1.18	1.14806	0.03194
4	200	1.12	1.115381	0.004619
5	250	1.06	1.082702	0.022702
6	300	1.02	1.050024	0.030024
7	350	0.98	1.017345	0.037345
8	400	0.95	0.984667	0.034667
9	450	0.93	0.951988	0.021988
10	500	0.9	0.91931	0.01931
11	550	0.87	0.886631	0.016631
12	600	0.83	0.853952	0.023952
13	650	0.8	0.821274	0.021274
14	700	0.78	0.788595	0.008595
15	750	0.88	0.755917	0.124083

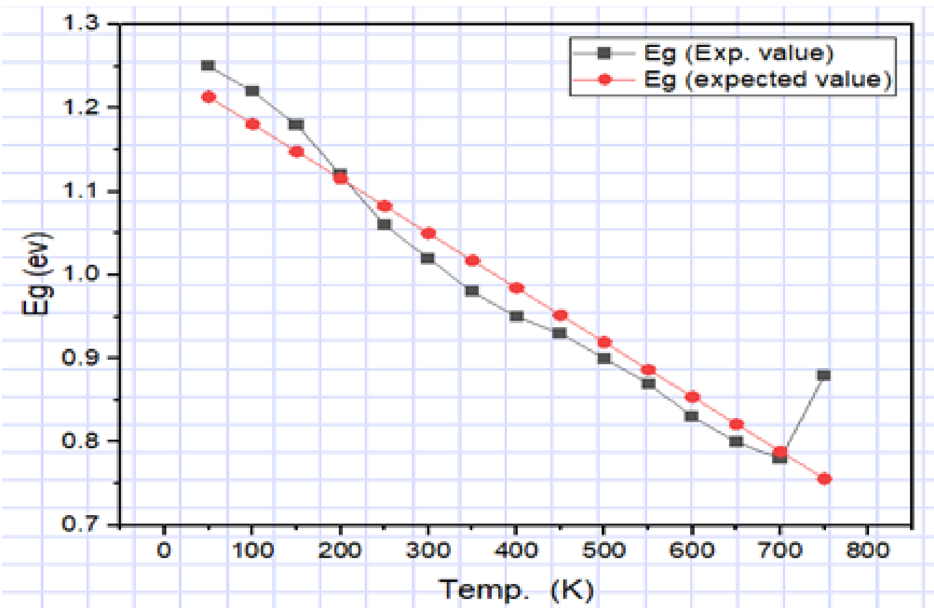

Figure (3) Experimental and approximate values measured by cluster analysis method

The observed convergence between approximate and experimental values can be observed when using this model, meaning that the percentage error in the selected values is small and can be acceptable.

4) Neville model Neville method

It is one of the methods through which an approximation is performed numerically and a new approximate value is calculated, as well as in this method the values of functions are calculated at certain points[6,7], and a first-order equation is formed and those values are used and entered into the Neville function.

5) Methodology, method of work and mathematical model construction

A mathematical programming model is formulated using the Python language through which approximate values are calculated, then the error rate is calculated, through which the effectiveness of the model is examined, as a set of points is selected and the function values are calculated at them, and this method is summarized as follows:

1) Choose approximate points of the actual values and be adjacent to them and then calculate the values of the function at them.

2) Use Python to formulate code and mathematical model, which provides the following functions:

First: Calculating the values of the function at the new points after the function and rounding points are chosen.

Second: Substituting the new values into the Neville function.

Third: Calculate approximate values.

Fourth: Calculating the percentage of error between the actual and calculated value by subtracting the approximate from the actual divided by the actual value, as well as calculating the standard deviation of the error rate value, which shows the effectiveness of the mathematical model[8].

Model construction

1) Points adjacent to the known values of the energy gap are selected in the exponential function.

2) According to these selected values.

3) Use the Neville equation to relate the point at the chosen points, which are:

$$f(x) = f(x_o) + (x + x_o)\frac{[f(x_1) - f(x_o)]}{(x + x_o)} - - - - - - - - - - - - - -(3)$$

4) Calculating the values of the required function after entering them into the Neville equation, to formulate the mathematical model of the multiple rounding method, we follow the following steps:

1) Choosing certain points that can be estimated, so that they are adjacent to the known points in the exponential function, which can be as follows:

$$f(x) = ax^2 + bx + c - - - (4)$$

so that the selected points () and () are a point at the beginning of the function $(x_o, x_1, x_2)x_o$

2) We calculate the value of each of $f(x_o), f(x_1), f(x_2)$

3) We enter the Neville equation as follows:

$$f(x) = f(x_o) + (x + x_o)\frac{[f(x_1) - f(x_o)]}{(x + x_o)} - - - - - - - - (5)$$

4) In the case of an energy and temperature gap it can be formulated in the following way:

$$Eg(T) = Eg(T_o) + (T + T_o)\frac{[Eg(T_1) - Eg(T_o)]}{(T + T_o)} - - - - - - (6)$$

6) Results

Figure (4) represents the approximate values of the energy gap

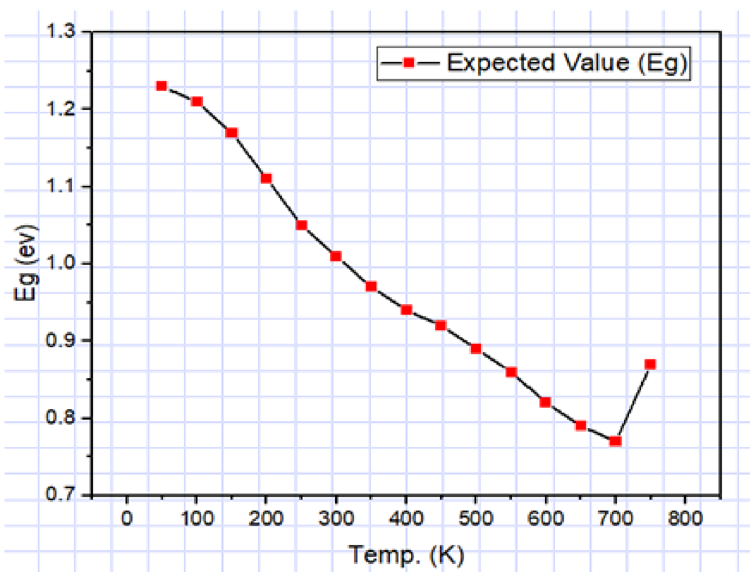

Figure (4) Approximate values calculated by the Neville method

Table (2) Approximate and experimental values of the energy gap calculated by cluster analysis method and error ratio as well.

	Temperature (K)	Experimental Energy Gap (eV) Values	Approximate values (eV)	Absolute error (eV)
1	50	1.25	1.23	0.02
2	100	1.22	1.21	0.01
3	150	1.18	1.17	0.01
4	200	1.12	1.11	0.01
5	250	1.06	1.05	0.01
6	300	1.02	1.01	0.01
7	350	0.98	0.97	0.01
8	400	0.95	0.94	0.01
9	450	0.93	0.92	0.01
10	500	0.9	0.89	0.01
11	550	0.87	0.86	0.01
12	600	0.83	0.82	0.01
13	650	0.8	0.79	0.01
14	700	0.78	0.77	0.01
15	750	0.88	0.87	0.01

Through Table (2), the lowest error rate was (0.01) and the highest error rate was about (0.2) The model can be considered acceptable because the relative error was relatively small ranging between (0.1-0.2) and when calculating the standard deviation of this model, it was found that when it becomes within (ev0.008) and this is very acceptable from a scientific and mathematical point of view, which reflects the accuracy of the model more.

Figure (5) represents the relationship between the approximate value and the experimental value, noting the convergence between the approximate and experimental values.

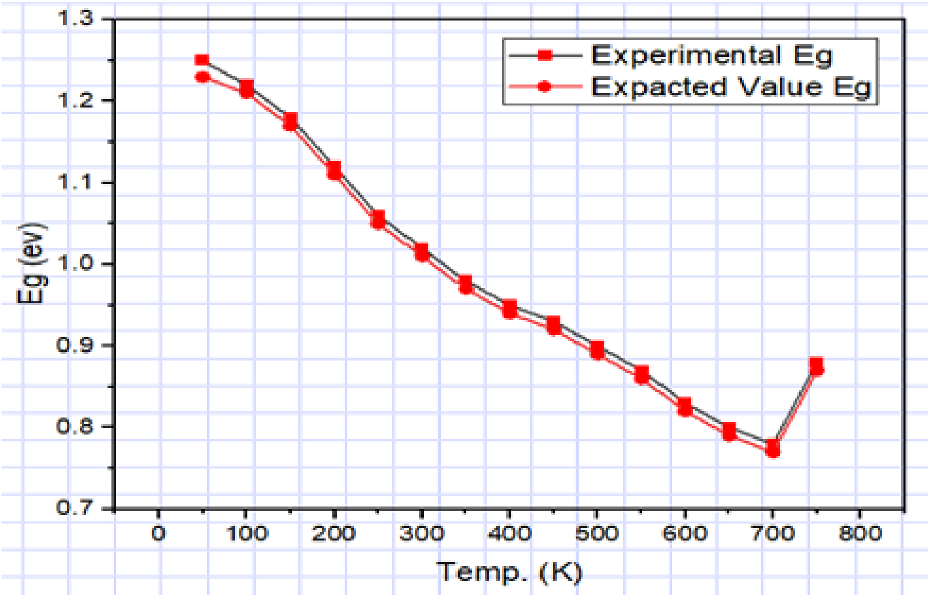

Figure (5) Relationship between approximate value and experimental value

The remarkable convergence between approximate and experimental values can be observed when using this model, meaning that the error rate in the selected values is small and can be very scientifically acceptable.

7) Hybrid model formulation

As passed in the previous hybrid model there must be a coefficient **Merge** which controls the balance between numerical interpolation and aggregation, where The proposed formula ensures that interpolation is applied within appropriate clusters taking into account global and local influences. It can be formulated in the following way:

$$P_\lambda(x) = \lambda \sum_{k=1}^{k} \beta_k P_k(x) + (1-\lambda) P_{i,j}(x) - - - - - - (7)$$

Whereas:

($P_\lambda(x)$:is the final approximate value that takes into account the analysis of aggregation.

(k): number of clusters

(β_k): The weight assigned to the cluster can be calculated in the following way:

$$\beta_k = \frac{e^{-d(x,C_k)}}{\sum_{m=1}^{K} e^{-d(x,C_m)}} - - - - - - - - - (8)$$

f($d(x,C_k)$):is the distance between the point (x) and the center of the cluster (C_k).

($P_k(x)$): Interpolation factor for the Neville method within the cluster.

($P_{i,j}(x)$): Data-level computed interpolation

(T,λ): hypothetical correction factor depends on a certain value ranging between (0-1), if the value is (0), then the hybrid model tends to the polynomial method and if it is (1), the model tends to be hybrid to the Cray method, so an average value will be chosen between (0-1) and be (0.5) [9].

8) Methodology, modus operandi and formulation of the hybrid mathematical model

1) The basic equation of the hybrid model was formulated based on the basic principles of mathematics that are concerned with merging two equations and adding a specific correction factor.

2) A mathematical model was formulated based on the Python programming language to calculate the numerical approximation and error rate during the selected periods, then calculate the value of the error rate, and this will be done by choosing a set of different points and calculating the values of the function at each of them, and the method is summarized as equation (9):

$$E_{g\lambda}(T) = \lambda \sum_{k=1}^{k} \beta_k E_{gk}(T) + (1-\lambda) E_{igNivelle}(T) --(9)$$

Whereas:

($E_{g\lambda}(T)$):):is the final approximate value of the energy gap at point (T)

(k): number of clusters

(β_k): The weight of each cluster can be calculated in the following way:

$$\beta_k = \frac{e^{-d(T,C_k)}}{\sum_{m=1}^{K} e^{-d(T,C_m)}} -----------(10)$$

and ($d(T,C_k)$)): is the distance between point (T) and the center of cluster (C_k)

($P_k(x)$):Interpolation factor for the Neville method within the cluster.

($P_{i,j}(x)$): Interpolation calculated at the entire data level without aggregation.

(T,λ): A hypothetical correction factor depends on a certain value ranging between (0-1), if the value is (0), then the hybrid model tends to the polynomial method and if it is (1), the model tends to be hybrid to the Cray method, so an average value will be chosen between (0-1) and be (0.5) [9.10].

3) This hybrid model allows combining both cluster integration and the Neville method to obtain accurate values of the energy gap at different temperatures.

4) Dividing the domain of functions into equal intervals, forming an approximate function for each of them, calculating the value of the integrals of subfunctions, grouping the values to obtain their approximate value, and then calculating the error rate value.

5) Using the Python programming language to create a mathematical model of the Cray method includes three functions to perform this integration, which are:

1) The first function: dividing the domain of the function into intervals related to each of the approximate values.

2) The second function: calculating the approximate values of the function by subfunctions.

3) The third function: through which the relative error is calculated as we have shown earlier by subtracting the actual value from the approximate and dividing by the actual value and multiplying by 100 to get its percentage.

4) The data was plotted to compare the results and verify the compatibility of the hybrid model with experimental values.

4.3.2 Results

Figure (6) represents the approximate values calculated for the energy gap

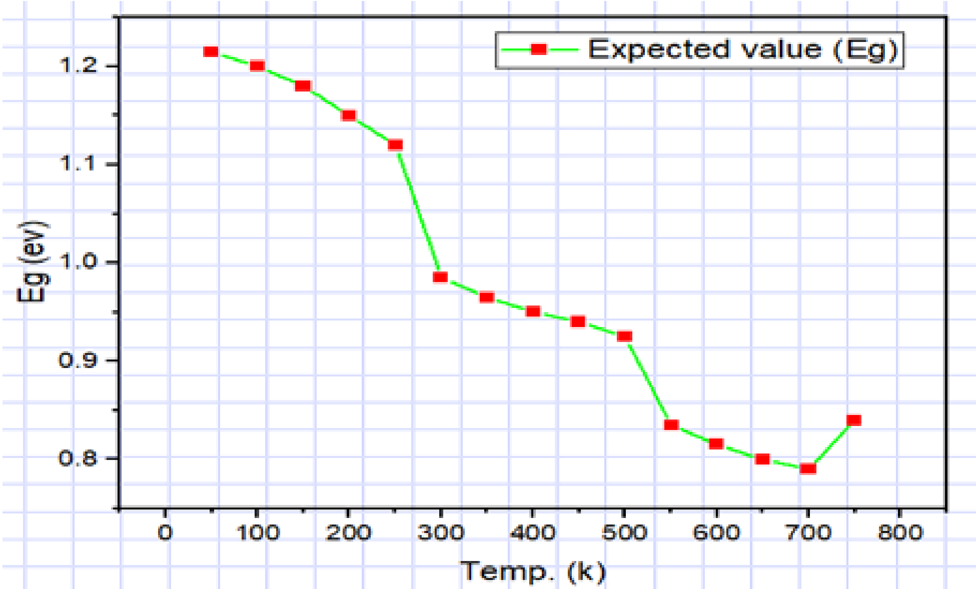

Figure (6) Approximate values calculated for the energy gap

Table (3) Approximate and experimental values of the energy gap calculated by cluster analysis method and error ratio as well.

	Temperature(K)	Experimental Energy Gap (eV) Values	Approximate values (eV)	error Absolute (eV)
1	50	1.25	1.215	0.035
2	100	1.22	1.2	0.02
3	150	1.18	1.18	0.002
4	200	1.12	1.15	0.03
5	250	1.06	1.12	0.06
6	300	1.02	0.985	0.035
7	350	0.98	0.965	0.015
8	400	0.95	0.947	0.01
9	450	0.93	0.94	0.01
10	500	0.9	0.925	0.025
11	550	0.87	0.835	0.035
12	600	0.83	0.815	0.015
13	650	0.8	0.8	0.001
14	700	0.78	0.79	0.01
15	750	0.88	0.84	0.04

Through Table (3), the lowest error rate was (0.001) and the highest error rate was (0.06), the model can be considered acceptable because the relative error was relatively little, ranging between (0.01-0.06) and when calculating the standard deviation of this model, it was found that when it becomes within (ev0.01573) and this is very acceptable from a scientific and mathematical point of view, which reflects the accuracy of the model more.

Figure (6) represents the relationship between the approximate value and the experimental value, noting the convergence between the approximate and experimental values.

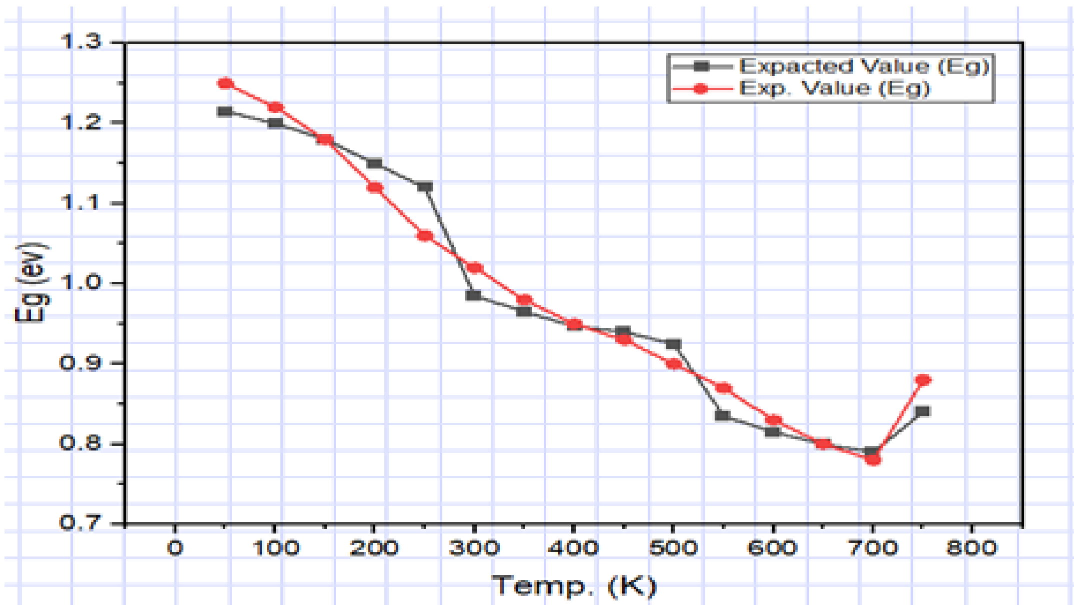

Figure (7) Relationship between approximate value and experimental value

The observed convergence between approximate and experimental values can be observed when using this model, meaning that the error rate in the selected values is small and can be scientifically acceptable.

The proposed hybrid model for calculating the energy gap, which combines the cluster analysis method with the Neville method, is promising due to its ability to integrate the physical and statistical advantages of both methods. This integration allows for improved accuracy of estimates and reduced potential errors.

Evaluation of model accuracy:

- **Absolute error and standard deviation: The** absolute error between the actual values and the expected values of the hybrid model was calculated, and the mean absolute error and standard deviation were found to have low values, indicating the model's accuracy in predicting the energy gap across the studied temperature range.
- **Mixing modulus (correction) (λ):** The use of a mixing coefficient λ of 0.5 ensures an even balance between the two combined methods, enhancing the accuracy of the model.

The use of hybrid models can lead to improved prediction accuracy in multiple domains, for example, in a study on the prediction of electrical load fluctuations, a hybrid approach was used to improve the accuracy of models in predicting the values of daily electrical energy consumption [11,12].

Figure (8) represents the relationship between experimental values and approximate value of the hybrid model, cluster analysis, and Neville method.

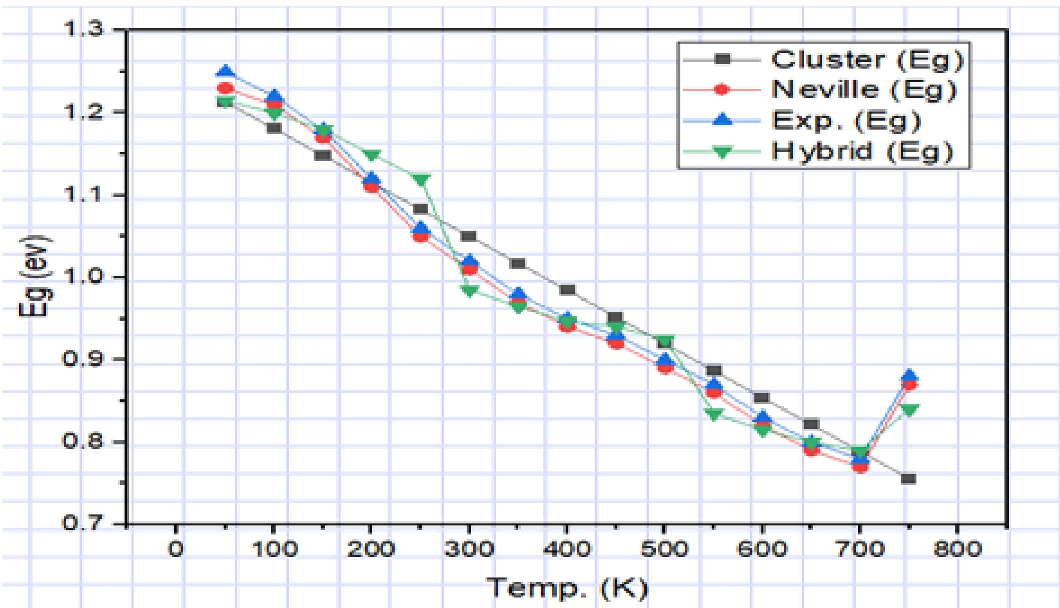

Figure (8): Relationship between experimental values and approximate value of hybrid model, cluster analysis and Neville method

Table (4) represents the approximate value of the energy gap for each of the three models plus the experimental value

	Temperature(K)	Cluster energy gap (eV)	Neville Energy Gap (eV)	Experimental power gap (eV)	Energy gap of hybrid model (eV)
1	50	1.213417	1.23	1.25	1.215
2	100	1.180738	1.21	1.22	1.2
3	150	1.14806	1.17	1.18	1.18
4	200	1.115381	1.11	1.12	1.15
5	250	1.082702	1.05	1.06	1.12
6	300	1.050024	1.01	1.02	0.985
7	350	1.017345	0.97	0.98	0.965
8	400	0.984667	0.94	0.95	0.947
9	450	0.951988	0.92	0.93	0.94
10	500	0.91931	0.89	0.9	0.925
11	550	0.886631	0.86	0.87	0.835
12	600	0.853952	0.82	0.83	0.815
13	650	0.821274	0.79	0.8	0.8
14	700	0.788595	0.77	0.78	0.79
15	750	0.755917	0.87	0.88	0.84

Table (4) shows that the best approximate values for calculating the energy gap were for the hybrid model, as they were closer to the experimental values than other models because this model took into account the following:

1) The hybrid model is characterized by accuracy, flexibility, it balances the exact interpolation within the clusters with the general data estimation.

2) The selection of optimal parameters takes into account the correctness of the results, and the effect of temperature on the calculations.

3) Achieves the best balance between physical and statistical methods which makes it effective in advanced applications.

4) Reduces the impact of outliers It reduces noise because data is collected before interpolation.

5) It is **flexible enough** to be eas-

ily adjusted to suit **other areas**, such as forecasting material states or energy calculations in semiconductor physics.

6) Improves the overall estimation of irregular data by distributing the data into clusters and then applying Neville interpolation within each group.

7) Ensures improved forecast accuracy without the need for overly complex calculations, as it combines cluster analysis with the Neville method, reducing the number of calculations compared to applying the Neville method to all data at once.

Result and Conclusions

The results showed the superiority **of the hybrid model** in calculating energy gap values compared to individual methods, thanks to its combination of the efficiency of cluster analysis in collecting data and the accuracy of the Neville method in numerical approximation. This model also reduced the impact of outliers and improved the accuracy of the results, especially when dealing with irregular data. On the other hand, the Neville method proved efficient in providing accurate results, despite the complexity of its calculations when applied to unclassified data, while cluster analysis was more appropriate. To handle large, albeit in some cases, less accurate datasets, the proposed hybrid model can be an advanced solution that combines accuracy and efficiency, making it an ideal choice for complex scientific applications, especially in the fields of physical and engineering data analysis. The results confirm that the integration of statistical and mathematical methods can lead to improved forecast quality and reduced errors, opening up new avenues for research and development in this field.

References

1. Varshni, Y. P. (1967). Temperature dependence of the energy gap in semiconductors. *Physica, 34*(1), 149–154. https://doi.org/10.1016/0031-8914(67)90062-6
2. MacDonald, DJ, Shalizi, C. R. & Sherfish, M. (2011). *Estimation of mixing coefficients β. Electronic Journal of Statistics* , 5, 242–270. Retrieved from https://projecteuclid.org/euclid.ejs/1300198553.

3. Al-Rashidi, H.H.H.J. (2023). *Modeling and predicting electrical load fluctuations using the slope of support vectors*. University of Basra.
4. Jane, A. K. (2010). Data aggregation: 50 years after K-Means. *Pattern Recognition Letters* , 31(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011.
5. Janardan Decali, 2022, Cluster Analysis Method, https://byjus.com/maths/cluster-analysis/.
6. Rokotashi, L., Maimon and Khrun (2005). Feed larai. ref or auspicious w. Rokesh (detained) *v. Fixer Ian* Ep. 2 (2005).
7. Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2), 179-188.
8. Trion, R.C. (1939). *Mass analysis: Cookie and evaluative (factor) analysis to isolate units in the mind and personality*. Edwards Brothers.
9. Fix, E., & Hodges, J.L. (1951). Discriminatory analysis - non-parametric discrimination: characteristics of consistency. *Technical Report 4, USAF Aviation Medical College, Randolph Field, Texas.*
10. Esther, M., Craigel, H.P., Sander, J., Shaw, X. (1996). A density-based algorithm for detecting clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96),* 226-231.
11. Cohonen, T. (1998). Self-regulation map. *Neural Computing* , 21(1-3), 1-6.
12. Bezdik, J.C. (1981). *Identify patterns using fuzzy target function algorithms*. Springer.