RESEARCH PAPER

Data Analysis Prediction of pKa Values in Neural Network using Topological Indices

M.S.Magdum*[1], D.S.Revankar[2], J. R. Gurjar[1]

*[1]Department of Mathematics, KLS Gogte Institute of Technology, Udyambag Belagavi, Karnataka, 590008, India;
[1]msmagdum@git.edu, https://orcid.org/0009-0008-4482-9159[1], jeetendra.g8@gmail.com
https://orcid.org/0000-0002-6574-353X,

[2]Department of Mathematics , KLE DR. M. S. Sheshgeri College of Engineering & Technology, Belgaum – 590008,India.
revankards@gmail.com, https://orcid.org/0000-0002-3441-9964

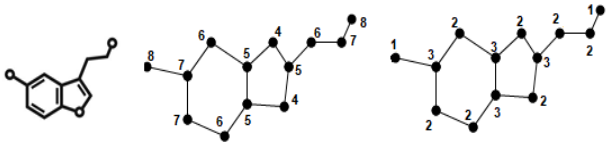
Abstract
This study aims to predict pKa value of neurotransmitters using eccentricity index and degree index. Materials and Methods: pKa values of neurotransmitters prescribed for specific data analysis tools like MATLAB. A study determines that the pKa prediction based on two methods, the eccentricity index and the degree index of chemical structures of eight neurotransmitters
.

Keywords: pKa value, molecular graph, neurotransmitter, eccentricity index, degree index.
Subject Classification: 05C07, 05C10
How to cite this article: Magdum MS, Revankar DS, Gurjar JR. Data Analysis Prediction of pKa Values in Neural Network using Topological Indices. Int J Drug Deliv Technol. 2026;16(16s): 217-223. DOI: 10.25258/ijddt.16.16s.23.

Introduction
Neurotransmitters are chemical messenger that send and receive messages from nerves to muscles. The nervous system is in charge of these chemical changes. These chemical changes are identified by their pKa and pH values. pKa and pH values are more essential for understanding the behaviour of chemical substances in everyday life. Every neurotransmitter has unique properties like its Molar mass, Boiling point, pKa, pH values, etc.
In this paper, we obtain the pKa values of neurotransmitters using graph eccentricity index and degree index. Our pKa prediction is based on two methods eccentricity index and the degree index of chemical structures of neurotransmitters. . Comparing to other approaches that use a far higher level of theory and practical computational cost, both sets were predicted with precision (rms errors of 0.7 – 1.0 log units).
Topological indices are mathematical descriptors of molecular graphs that predict chemical compounds' physiochemical, biological, toxicological and structural properties [7].
Let a graph G have vertex set V(G) and edge set E(G). The size of G is the number of edges, and the order is the shortest path between vertices u and v. The eccentricity $eccG(u)$ of u in G is the distance between u and the vertex that is farthest from u in G. The greatest eccentricity among the vertices' eccentricities in G is its diameter.

Eccentricity of graph G:
The eccentric index of G is defined as,
$\mathrm{Ecc(G)} = \sum_{v\epsilon\, V(G)} ecc(G)$.

Degree of graph G:
The degree index of G is defined as,
$\mathrm{Deg(G)} = \sum_{v\epsilon\, E(G)} deg(G)$.



Serotonin Eccen. of Serotonin Degree of Serotonin

Figure 1

Serotonin Eccentricity of Serotonin Degree of Serotonin
From figure of Serotonin, we observed that,
$\mathrm{Ecc(G)} = \sum_{v\epsilon\, V(G)} ecc(G) = 72$ and
$\mathrm{Deg(G)} = \sum_{v\epsilon\, E(G)} \deg(G) = 28$.

Directed graphs are used to calculate the DNA gap penalty [13], Chemical equations are represented [14], constructive approach for calculating the Wiener index [15], the molecular topological index of a tree [16], a brief description of graph theory's impact on chemistry [17], Boiling point and Wiener index correlation [18], eccentricity-related indices and reverse eccentricity connectivity index is examined for chemical compounds [19]. Motivated from above research authors have studied the pKa values of Neural network using eccentricity index and degree index.

*Author for Correspondence: msmagdum@git.edu

pKa and pH definition: pKa value is defined as,

$$pK_a = -log_{10}K_a.$$

An appropriate emphasis is placed on the most commonly used classical methods as like potentiometry and ultraviolet–visible spectroscopy, and to a separation technique, capillary electrophoresis, which nowadays is strongly recommended for pKa evaluation and shows the advantage that it does not require pure samples for measurements. Regression Statistics explains (i) Multiple R: (ii) R Square (R^2). (iii) Standard Error.

There are very good methods adopted for pKa determination using laboratory like Potentiometry, Spectrophotometry, and Capillary Electrophoresis.

DATA SET: A data set was collected from Neurotransmitter with its original given pKa and determined the eccentricity index of the graph (chemical structure) of each Neurotransmitter.

For any given set of data X and Y, a linear regression model is an equation of the form Polynomial Curve Fit. i.e. $f(x) = p1 * x^3 + p2 * x^2 + p3 * x + p4$.

Table 1. The Data of Eccentricity index(G) and Original pKa of Neurotransmitter

Sr. No	Neurotransmitter	Original pK_a value	Ecc. Index value
1	Adrenaline	9.2	88
2	Noradrenaline	8.58	65
3	Dopamine	8.93	59
4	Serotonin	9.5	72
5	Gamma Aminobutyric Acid	11.3	29
6	Glutamate	9.3	50
7	Edorphins	8.49	503
8	Acetylcholine	9	51

Table-2. SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.580890385
R Square	0.337433639
Adjusted R Square	0.194576497
Standard Error	8.113678548
Observations	8

Table -3. ANNOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	234.6889	234.6889	3.564979	0.107934
Residual	7	460.8225	65.83178		
Total	8	695.5114			

Table -4. t-stat and p-value

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Ecc(G)	0.0289727	0.015345	1.88811	0.10095	-0.00731	0.06525	-0.00731	0.065257

*Author for Correspondence: msmagdum@git.edu

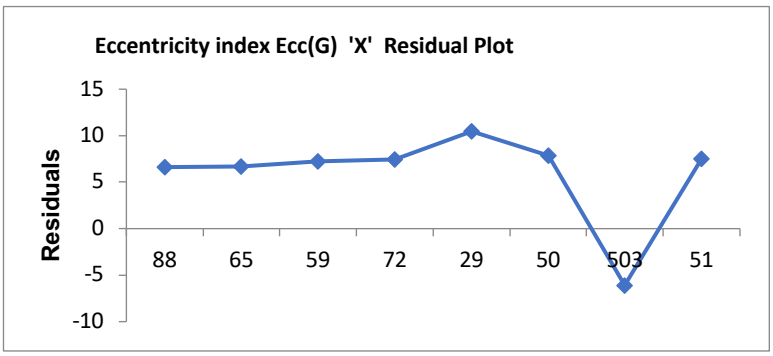


Figure-2 Eccentricity index Ecc(G) 'X' Residual Plot

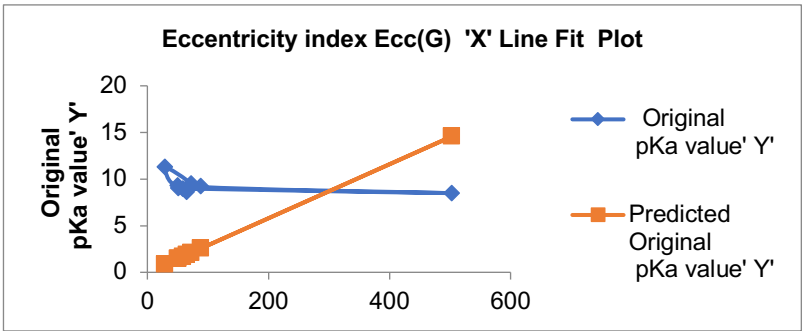


Figure-3 Eccentricity index Ecc(G) 'X' Line Fit Plot

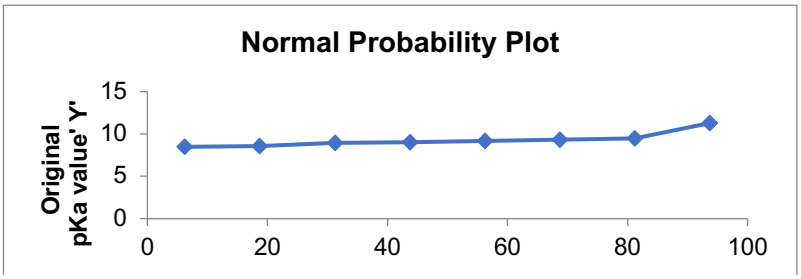


Figure-4 Normal Probability Plot

Table 5: Eccentricity index of Neurotransmitters

Fit Name: Eccentricity index			
Polynomial Curve Fit (poly3)			
$f(x) = p1 * x^3 + p2 * x^2 + p3 * x + p4.$			
x is normalized by mean 114.6 and std 157.9			
Coefficients and 95% Confidence Bounds			
	Value	Lower	Upper
p1	-13.279	-23.117	-3.4415
p2	24.712	7.5693	41.854
p3	18.256	1.7024	34.81
p4	11.724	8.6081	14.84
Goodness of Fit			
Value			
SSE	0.5951		
R-square	0.8908		
DFE	4		
Adj R-sq	0.8089		
RMSE	0.3857		
Goodness of Validation			
Value			

SSE	0.5951
RMSE	0.2727

From table-1 to table-5, comparing pKa values and eccentricity index of G values for different neurotransmitters and comparing to descriptors of neurotransmitters. Combined analysis of pKa and eccentricity index values reveals a complementary relationship between chemical reactivity and structural complexity. Regression Statistics R value is 0.5809 indicates positive correlation between eccentricity index and pka values of neurotransmitter.

In table-1 to table-5, we observe all coefficients have confidence intervals that do not include 0, which means they are statistically significant at the 95% level. In goodness of fit, sum of squared errors – lower is better, **89.08%** of variance in the data is explained by the model, adjusted for number of predictors – still very strong, Root Mean Square Error – average error magnitude (~0.39 units). Goodness of Validation, Sum of Squared Errors – lower is better, Lower than fit RMSE suggests good predictive performance.

Table-6. Data of degree index(G) and Original pKa of Neurotransmitter

Sr.no	**Neurotransmitter**	**Original pK_a value**	**degree Index deg(G)**
1	Adrenaline	9.2	25
2	Noradrenaline	8.58	23
3	Dopamine	8.93	22
4	Serotonin	9.5	28
5	Gamma Aminobutyric Acid	11.3	11
6	Glutamate	9.3	18
7	Endorphins	8.49	68
8	Acetylcholine	9	17

Table-7. SUMMARY OUTPUT

Regression Statistics	
Multiple R	**0.821873686**
R Square	0.675476355
Adjusted R Square	0.532619212
Standard Error	5.678403882
Observations	8

Table-8. ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	469.801505	469.8015054	14.57008	0.008790638
Residual	7	225.709895	32.24427065		
Total	8	695.5114			

Table-9. t -Stat and p-value

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
deg(G) 'X'	0.245735219	0.06437785	3.81707702	0.006567	0.093505789	0.397964648

Table-10 RESIDUAL OUTPUT AND PROBABILITY OUTPUT

Sr.no	Predicted Original pK_a value' Y'	Residuals	Standard Residuals	Percentile	Original pK_a value' Y'

1	6.143380463	3.05661954	0.575454618	6.25	8.49
2	5.651910026	2.92808997	0.551256993	18.75	8.58
3	5.406174807	3.52382519	0.663413112	31.25	8.93
4	6.880586118	2.61941388	0.493144075	43.75	9
5	2.703087404	8.5969126	1.618498145	56.25	9.2
6	4.423233933	4.87676607	0.918124588	68.75	9.3
7	16.70999486	-8.21999486	-1.54753771	81.25	9.5
8	4.177498715	4.82250129	0.907908426	93.75	11.3

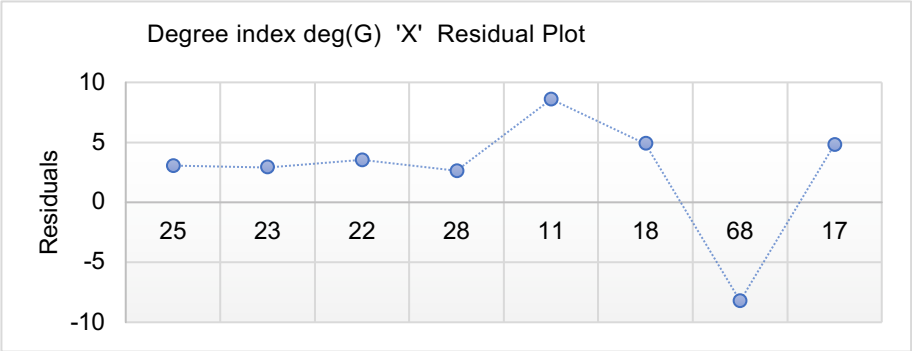


Figure-5 Degree index deg(G) 'X' Residual Plot

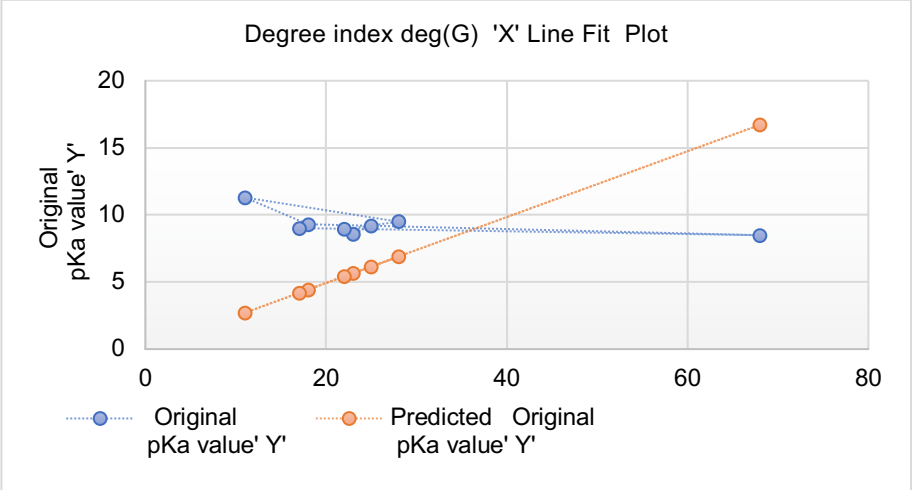


Figure-6 Degree index deg(G) 'X' Line Fit Plot

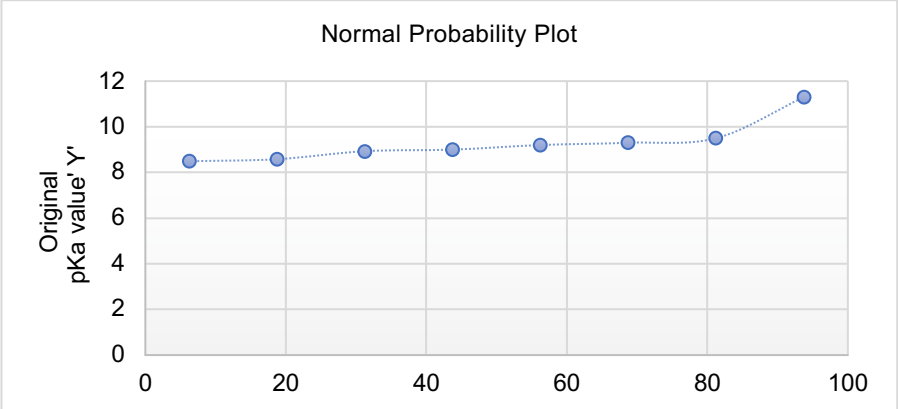


Figure-7 Normal Probability Plot

Table 11: Degree index of Neurotransmitters

Fit Name: Degree index			
Polynomial Curve Fit (poly3)			
$f(x) = p1 * x^3 + p2 * x^2 + p3 * x + p4.$			
x is normalized by mean 26.5 and std 17.57			
Coefficients and 95% Confidence Bounds			
	Value	Lower	Upper
p1	-2.2028	-3.1908	-1.2147
p2	3.8706	2.3654	5.3759
p3	2.8186	0.7979	4.8393
p4	9.256	8.8377	9.6744
Goodness of Fit			
Value			
SSE	0.2257		
R-square	0.9586		
DFE	4		
Adj R-sq	0.9275		
RMSE	0.2375		
Goodness of Validation			
Value			
SSE	0.2257		
RMSE	0.168		

Table-6 to Table-11, it's a topological descriptor from graph theory that represents the sum of vertex degrees in the molecular graph of the compound.

The regression statistics include the correlation coefficient (R), which quantifies the agreement between observed and predicted pKa values. A value of 0.822 indicates a strong positive correlation, model explains 67.5% of the variance in pKa values based on degree index deg(G). This is a fairly strong fit, on average, the predicted pKa is off by ±5.68, which is a moderately high error. ANOVA (Analysis of Variance) table, we have $p < \mathbf{0.01}$, the model is statistically significant, the degree index deg(G) has a strong effect on pKa.

Regression coefficients table, which provides details about the linear relationship between the degree index deg(G) and the pKa values. Standard Error is 0.0644 smaller values indicate more reliable estimates, high absolute value of t -distribution indicates stronger evidence against the null hypothesis, p-value (0.00657) At the 1% significance level, the result is statistically significant. The 95% confidence interval (0.0935 to 0.3980), being entirely above zero, further supports the presence of a positive and meaningful relationship.

Conclusions:

We have described a Graph theory and Data analysis methodologies for the prediction of pKa for Neurotransmitters. We get the eccentricity index (G) data in relation to the neurotransmitter's original pKa. ANOVA explains significant differences in the pka values, a polynomial curve is fitted, t-statistics and p-values are provided in table 4, the goodness of fit and goodness of validation for the eccentricity index of neurotransmitters, etc. We obtain R=0.580890385, which is a positive correlation. Consequently, we have shown a number of charts, such as the eccentricity index vs. residual, line plot, and normal probability plot.

We get the degree index (G) data in relation to the neurotransmitter's original pKa. ANOVA explains significant variations in the pka values, a polynomial curve is fitted, t-statistics and p-values are provided in table 4, the degree index of neurotransmitters has a goodness of fit and a goodness of validation, and we gain a positive correlation (R=0.821873686).

Consequently, a number of charts have been shown, such as the degree index vs. residual, line plot, and normal probability plot. Using neurotransmitter data analysis, we compare the eccentricity index and degree index and find that the degree index has a higher pKa value**.**

Furthermore, a variety of physicochemical properties, such as vapor pressure, melting point, and boiling point, can be determined using the degree index and eccentricity index approaches

Conflicts of Interest:

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgment:

No support.

REFERENCES

1. Wiener H. *"Structural determination of paraffin boiling points"*. J. Am Chem. Soc.;69:p17-20,1947.
2. Divya T, & Yamuna M. "*Drug pKa Value Prediction – From Compressed Graph"*. Innovare Journal of Engineering and Technology, 7(5), p1–6. (2019).
3. Yamuna M. "*DNA Gap Penalty Using Directed Graph*." Pharm Lett;7:p392-8. 2015.
4. Yamuna M, Elakkiya A. "*Chemical Equation Representation As Directed Graph"*. Pharm Chem 2015;7:p49-55.
5. Yamuna M. *"Wiener index of chemical trees from its subtree"*. Pharm Chem;6:p235-42. 2014.
6. Yamuna M, Divya T. "*Molecular Topological Index of Tree with Equal Number of Vertices of a Given Tree"*. Vol. 263. IOP Conference Series: Materials Science and Engineering;. p. 42119.2017.
7. Yamuna M. "*A brief review on contribution of graph theory and wiener index to chemistry"*. Int J Pharm Technol;8:p11182-92. 2016.
8. Adams C, Franzosa R. "*Introduction to Topology*". London: Pearson Education;. p. 420-1. 2008.
9. Wu J, Farahani MR, Yu X, Gao W. *"Physical-chemical properties studying of molecular structures via topological index calculating"*. Open Phys;15:p261-9. 2017.