ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF RESEARCH IN PHARMACEUTICAL SCIENCES

Published by JK Welfare & Pharmascope Foundation Journal Home Page: www.ijrps.com

A Statistical Comparison between Zagreb indices for correlation with toxicity predictions of natural products

Siva Parvathi M.*[1], Sujatha D.[2], Sukeerthi T.[3]

[1]Department of Applied Mathematics, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India
[2]Institute of Pharmaceutical Technology, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India
[3]Department of Statistics, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India

Article History:

Received on: 24 Nov 2021
Revised on: 27 Dec 2021
Accepted on: 30 Dec 2021

Keywords:

Natural Compounds, First Zagreb Index, Second Zagreb Index, Hyper Zagreb Index, Toxicity

ABSTRACT

Graph theory had wide applications in developing *in silico* tools and it is widely used to calculate topological indices to establish structural activity relations of chemicals or compounds. However, usage of Zagreb indices with respect to natural compounds activity or toxicity prediction needs more attention. Many available online tools are using atom bond connectivity index (ABC Index), first and second Zagreb indices. The usage of the Hyper Zagreb index is very rare and using natural compounds is neglected. In this context, three types of Zagreb indices (first Zagreb index, second Zagreb index and hyper Zagreb index) were calculated to the selected chemical compounds of natural products and the relation between these indices and cytotoxicity of natural compounds were established. We have selected IC50 Values of the selected natural compounds in *Hela* cell lines as an index for cytotoxicity from the literature. The correlation of Zagreb indices and activity was performed using the R program, and we reached the conclusion that all indices correlate with the cytotoxicity of the studied compounds. This study acts as evidence to prove that, hyper Zagreb index correlates more with the cytotoxicity or activity of the studied natural compounds. Further studies using other Machine Learning tools to verify these findings will establish the importance of the hyper Zagreb index as one method to predict the toxicity of natural compounds.

*Corresponding Author

Name: Siva Parvathi M.
Phone: 9491315894
Email: parvathimani2008@gmail.com

ISSN: 0975-7538

DOI: https://doi.org/10.26452/ijrps.v13i1.32

Production and Hosted by

IJRPS | www.ijrps.com

© 2022 | All rights reserved.

INTRODUCTION

Toxicity predictions is one among the many variables impacting the success of drug discovery. To reduce the expenses and uncertainties of *in vitro*/*in vivo* experiments, it is crucial to perform high-throughput computer toxicity predictions. One dominant and most developed toxicity prediction method is Quantitative Structure-Activity Relationships (QSAR) based on chemical structural parameters [1]. In the era of Big Data and artificial intelligence, toxicity prediction can also be benefited from machine learning, which has been widely used in many fields. Characterization of chemical molecule descriptors and developing a suitable machine learning algorithm make it feasible to

develop computer-based toxicity, prediction models. Literature suggests a correlation of topological indices and toxicities using aryl hydrocarbons, organophosphorus compounds, small molecules etc.

In ADMET prediction models, molecular descriptors are utilized to correlate the structure-property relationship in order to predict the ADMET properties of molecules based on their descriptor values [2]. The molecular descriptors can be classified into three types, one dimensional, two dimensional and three-dimensional descriptors in ADMET models depending on the chemical representation level required for descriptor calculation. The simplest sort of molecular descriptor is one dimensional, which represents the information calculated from the molecule's molecular formula, such as the type and number of atoms in the molecule, as well as the molecular weight. The two-dimensional descriptors are much complicated than the one dimensional, and they usually represent the information about the molecule's size, structure, and electronic dispersion.

Numerous investigations have found a strong intrinsic association between chemical compounds, chemical properties (such as boiling and melting points) and their molecular structures. Researchers can use well-defined topological indices on these chemical molecular structures to better understand their physical properties, chemical reactivity and biological activity. Thus, the study of topological indices on the chemical structure of chemical materials and pharmaceuticals can compensate for the lack of chemical experiments and give a theoretical foundation for drug and chemical material synthesis. In 1972, Zagreb indices are developed and established as one of the graph-theoretical topological indices, which plays a crucial role to explain the structural properties of the chemical compounds in the study of QSPR and QSAR [3]. Das and Gutman [4, 5] introduced the Zagreb indices and some properties of chemical compounds at the molecular level were mathematically explained through them. Since then, the Zagreb indices have been extensively explored and concentrated due to their wide range of applications instead of the existing chemical methods, which required additional time and expanded expenses. For a variety of causes, many novel types of Zagreb indices [4–6] are provided. The first and second Zagreb indices are associated with different programs used for the typical count of topological indices Viz., POLLY, DRAGON, CERIUS, TAM and DISSIM [7].

With this, the researchers try to find out the relationship between graph theory-based topological indices and the toxicity of the chemical compounds and explain internal code and mathematics. We have a notion that natural compounds are overall safe when compared to synthetic compounds. Studies using natural compounds to understand the association of different topological indices and activity or toxicity are sparse. Hence, this study was planned to calculate Zagreb indices to a set of natural molecules and understand the correlation of those indices with the reported biological activity or toxicity. We have also tried to compare different types of Zagreb indices for better correlation with biological activity or toxicity.

METHODS

Data Collection and preparation

As *in vivo* and *in vitro* data for cytotoxicity in a single cell line and animal model is required for development and finding an association of the calculated Zagreb indices using the R program, we have selected natural compounds available from the NPASS database {http://bidd.group/NPASS/} [8]. In the NPASS database, a search was carried using a target of Hela cell lines (Target ID – NPT165) which revealed data points for 1856 natural compounds.

In the next step, the data was cured by eliminating the compounds without a name, available as fatty acids, esters, salts, isomers etc. The 2D structures of the compounds were collected from PubChem. At this stage, compounds having a molecular weight above 1000 daltons were eliminated as plotting a molecular graph for large weight compounds will be more complex.

The cytotoxicity of finalized 800 compounds on Hela cell lines with Target ID- NPT165 were uniformly converted into nM data and then into their negative logarithmic values and designated as $pIC_{50.}$ All the cured data was prepared in CSV format.

Data Description

To predict *in silico* toxicology of compounds, many molecular descriptors within the scope of structure–toxicity interactions have been proposed. Topological descriptors are one among them and depend on graph theory concepts like connectivity, degree of a node and adjacency between nodes; refer to the overall topology of the compound and is determined by how atoms are related to one another. Topological characteristics of drug molecules are calculated using many mathematical indices. As Zagreb indices are based on vertices and their degrees, they are considered a method for estimating the topological characters of a chemical. Hence, we have calculated three types of Zagreb Indices for the above-prepared

© International Journal of Research in Pharmaceutical Sciences

dataset in CSV format.

Definition

Let G be a graph and d_i, d_j be the degrees of the vertices v_i, v_j respectively. Then the First Zagreb index of a graph G is Characterized as,

$$M_1(G) = \sum_{v_i v_j \in E(G)} (d_i + d_j)$$

and the Second Zagreb index of a graph G is,

$$M_2(G) = \sum_{v_i v_j \in E(G)} (d_i \times d_j)$$

In the year 2013, [6] presented another distance-based Zagreb index of a graph G named as Hyper-Zagreb Index.

Definition

Let G be a graph and d_i, d_j be the degrees of the vertices v_i, v_j respectively. Then the hyper Zagreb index of a graph G is Characterized as,

$$M_H(G) = \sum_{v_i v_j \in E(G)} (d_i + d_j)^2$$

R program

R is open-source programming and software for statistical computing. It can be available through the internet under the general public license (GPL). R can deal with a huge variate of mathematical and statistical tasks, and it can handle complex statistical approaches as easily as more simple ones. R language is widely used among statisticians and data minors for developing statistical software, data analytics, calculation and graphic tools [9–11]. It is an implementation of programming language combined with lexical, scoping, semantics, R and its library implements a wide variety of linear and non-linear regression models, time series, parametric and non-parametric tests, clustering and smoothing. R will give minimal output and store the result in an object.

Statistical Tool

The strength of the relationship between any two variables is measured in terms of an index is called the correlation coefficient. The study of these coefficients shows the variables which are closely associated with each other. There are different methods to perform correlation coefficients. i.e., Pearson correlation, Spearman correlation and Kendall. Here we perform the pearson correlation coefficient test. Pearson correlation(r) measures a linear dependence between two variables, and it is also known as parametric correlation.

$$r = \frac{\Sigma(x - mx)(y - my)}{\sqrt{\Sigma}(x - mx)2\Sigma(y - my)^2}$$

Where x and y are two vectors of length, m_x and m_y corresponds to the means of x and y, respectively. The following were the steps to calculate the Correlation coefficient using R programming:

1. Enter the data in a Microsoft Excel sheet and save it as data in CSV (comma delimited) format.

2. To import data from Excel into the R console.

Syntax:

> > CuredNP<-read.csv(file.choose ( ), header=T)

> CuredNP

3. To calculate correlation test between each variable in the given data.

Syntax:

>cor.test(CuredNP$FZI,CuredNP$pIC50,method="pearson")

>cor.test(CuredNP$SZI,CuredNP$pIC50,method="pearson")

>cor.test(CuredNP$HZI,CuredNP$pIC50,method="pearson")

4. By plotting the correlation test, we display the plot between the variables.

> plot (FZI, pIC50)

Similarly,

> plot (SZI, pIC50)

> plot (HZI, pIC50)

RESULTS

In this section, the relations between the First Zagreb Index (FZI), Second Zagreb Index (SZI), Hyper Zagreb Index (HZI) and cytotoxicity values of natural compounds were established.

```
> cor.test(CuredNP$FZI,CuredNP$pIC50,method="pearson")

        Pearson's product-moment correlation

data:  CuredNP$FZI and CuredNP$pIC50
t = 1.9293, df = 413, p-value = 0.05438
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.001768398  0.189050787
sample estimates:
      cor
0.09450919
```

Figure 1: Output in R – Correlation between FZI and pIC50

The correlation between FZI and pIC50 were assessed using R program, and the t-test statistic value was found to be t = 1.9293, which checks whether there is any significant differences between the FZI and pIC50 means, df is the degrees of freedom (n-2) is the number of data points minus 2 (df = 413) value. A p-value represents the probability of the correlation between the FZI and pIC50 in the sample data.

In our study, p = 0.05438, which means that a statistically significance correlation exists between FZI

```
> cor.test(CuredNP$SZI,CuredNP$pIC50,method="pearson")

        Pearson's product-moment correlation

data:  CuredNP$SZI and CuredNP$pIC50
t = 1.9778, df = 413, p-value = 0.04862
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.0006065625 0.1913398355
sample estimates:
      cor
0.09686241
```

Figure 2: Output in R – Correlation between SZI and pIC50

```
> cor.test(CuredNP$HZI,CuredNP$pIC50,method="pearson")

        Pearson's product-moment correlation

data:  CuredNP$HZI and CuredNP$pIC50
t = 1.9815, df = 413, p-value = 0.0482
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.0007901129 0.1915166598
sample estimates:
      cor
0.09704423
```

Figure 3: Output in R– Correlation between HZI and pIC50

and pIC50 values. The confidence interval means the distance from the lower confidence limit to the upper confidence limit of the correlation coefficient values at 95% (conf.int = [-0.001768398, 0.189050787]). The sample estimated value was assumed by the actual calculation between the sample size and width of the data, and the accuracy correlation coefficient (cor. coeff) value was cor. coeff=0.0945.

The p-value of the correlation test between these two variables is 0.05. At the 5% significance level, we reject the null hypothesis. There was a reasonably strong positive correlation between the two variables of FZI and pIC50, and an increase in the FZI values was associated with increases in pIC50 variables. Therefore, we conclude that there was a linear relationship between the First Zagreb Index (FZI) and pIC 50 values (Figure 1).

In the correlation studies between SZI and pIC50, t value was t = 1.9778, which indicates a significant difference between the SZI and pIC50 means, df was 413 with p-value = 0.04862 and the confidence interval at 95% was conf.int=[-0.000606,0.191339]. The sample estimated value was assumed by the actual calculation between the sample size and width of the data, and the accuracy correlation coefficient was found to be cor. coeff = 0.094.

The p-value of the correlation test between these two variables was 0.04. At the 5% significance level, we reject the null hypothesis. There was a reasonably strong positive correlation between the two variables of SZI and pIC50, increasing in SZI values was associated with increasing in pIC50 variables.

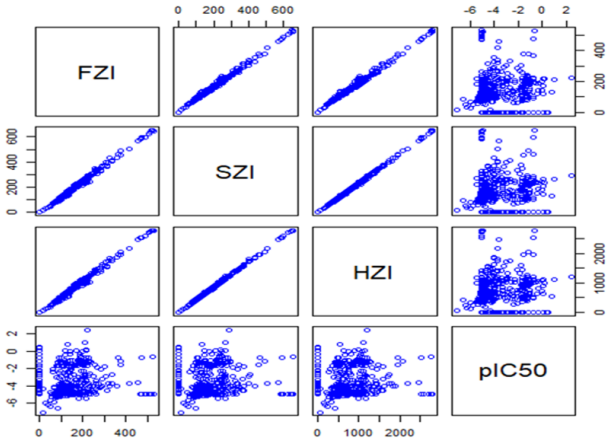

Figure 4: First Zagreb Index (FZI), Second Zagreb Index (SZI), Hyper Zagreb Index (HZI) correlation with pIC 50

Therefore, we conclude that there is a linear relationship between the Second Zagreb Index (SZI) and pIC 50 values (Figure 2).

When HZI and pIC50 were correlated, t value was t = 1.9815, df = 413 and p-value = 0.0482, which means that a statistically significant correlation existed between the inputs with conf.int=[0.0007901129,0.1915166598] at 95%. The sample estimated value was assumed by the actual calculation between the sample size and width of the data, which implicated cor. coeff = 0.097 level.

The p-value of the correlation test between these two variables was 0.04. At the 5% significance level, we reject the null hypothesis. There was a reasonably strong positive correlation between the two variables of HZI and pIC50, increasing in HZI values were associated with increasing in pIC50 variables. Therefore, we conclude that there was a linear relationship between the Hyper Zagreb Index (HZI) and pIC 50 values (Figure 3).

The pictorial representation of correlations between FZI, SZI and HZI with pIC50 was represented in Figure 4, and Hyper Zagreb indices were highly correlated positively to the pIC50 values of natural compounds. Further, using additional machine learning technologies to identify the rationale for correlation will have more uses in the field of drug discovery.

CONCLUSION

Toxicity studies are a very cumbersome, time-consuming and costly affair in the process of drug development. Moreover, the focus on the toxicity of natural compounds is very sparse. In this context, we attempted to use Zagreb indices to calculate

© International Journal of Research in Pharmaceutical Sciences

the topological characteristics of the collected natural compounds. These inputs were used to predict a correlation between the pIC50 values of cytotoxicity and Zagreb indices from the prepared dataset. R programming was used in the pilot evaluation and significant relation between hyper Zagreb indices and toxicities were found. Further studies to compare between FZI, SZI and HZI to understand the best indices among the study, for toxicity correlation will be useful. Moreover, applying other ML tools to understand the basis for correlation will have more applications in the field of drug discovery.

ACKNOWLEDGEMENT

The authors acknowledge Prof.S.Jyothi, Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India , for her scientific inputs.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding Support

This study was supported by seed funding sanctioned to Dr.M. Siva Parvathi from CURIE-AI grant sanctioned to Sri Padmavati Mahila Visvavidyalayam, Tirupati, India, funded by the Department of Science and Technology, India.

REFERENCES

[1] A Cherkasov, E N Muratov, D Fourches, A Varnek, I I Baskin, M Cronin, J Dearden, P Gramatica, Y C Martin, and R Todeschini. QSAR modelling: where have you been? Where are you going to? *Journal of medicinal chemistry*, 57(12):4977–5010, 2014.

[2] M T Khan and I Sylte. Predictive QSAR modelling for the successful predictions of the ADMET properties of candidate drug molecules. *Current drug discovery technologies*, 4(3):141–149, 2007.

[3] J Braun, A Kerber, M Meringer, and C Rucker. The similarity of molecular descriptors: the equivalence of Zagreb indices and walk counts. *MATCH Communications in Mathematical and in Computer Chemistry*, 54:163–176, 2005.

[4] I Gutman and K C Das. The first Zagreb index 30 years after. *MATCH Communications in Mathematical and in Computer Chemistry*, 50(1):83–92, 2004.

[5] K C Das and I Gutman. Some properties of the second Zagreb index. *MATCH Communications in Mathematical and in Computer Chemistry*, 52(1):103–112, 2004.

[6] G H Shirdel, H Rezapour, and A M sayadi. The hyper-Zagreb index of graph operations. *Iranian Journal of Mathematical Chemistry*, 4(2):213–220, 2013.

[7] S R Jammalamadaka. Essential Statistics with Python and R. *UC Santa Barbara*, pages 1–262, 2019.

[8] K Nordhausen and Peter Dalgaard. Introductory Statistics with R, Second edition. *International Statistical Review*, 77(1):155–156, 2009.

[9] K V S Sarma. Statistics made simple. Do it Yourself on PC. Second edition. page 303, New Delhi, 2010. Prentice hall of India Learning Private Limited. ISBN: 978-81-203-4017-6.

[10] M Sorokina and C Steinbeck. Review on natural products databases: where to find data in 2020. *Journal of cheminformatics*, 12(1):1–51, 2020.

[11] R Todeschini and V Consonni. Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references. page 1257. John Wiley & Sons, 2009. ISBN: 9783527628773.