Nat. Volatiles & Essent. Oils, 2021; 8(4): 9588-9597

Quantitative Structure Activity Relationship Study For The Prediction Of Complexity For 2-Acetamido-2-Deoxy-Beta-D- Glucopyranose Structurally Similar Compounds Using Regression

DR. D. PONMARY PUSHPA LATHA , DR. S. THANGA HELINA , DR. K. SIVARANJANI , D. JOSEPH PUSHPARAJ

Associate professor, Department of Digital Sciences, Karunya Institute Of Technology And Sciences, Coimbatore, ponmarymca@gmail.com, Assistant Professor, Department of Commerce , Karunya Institute Of Technology And Sciences,Coimbatore, thangahelina@karunya.edu, Assistant Professor,Department Of Mathematics, Karunya Institute Of Technology And Sciences,Coimbatore, sivaranjani@karunya.edu, Assistant Professor,Department Of Information Technolgy,Francisxavier Engineering College,Tirunelveli
er.joseph.raj@gmail.com

ABSTRACT- Coronavirus disease, which is called as COVID-19 and that is one of the infectious diseases which is infected by newly found coronavirus. Maching Learning has the major role in predicting the drugs of the particular disease. Lalmuanawma et al. 2020 has given the application of machine learning and artificial intelligence in COVID'19. It is used to develop the model design, Regression is one of the supervised Machine Learning Techniques. It is used to predict the values based on the data given. In this research work, Quantitative structure activity relationship (QSAR) study has been developed for structurally similar to 2-acetamido-2-deoxy-beta-D-glucopyranose as inhibitors for COVID-19 causing targets using regression. QSAR models for complexity was created with 40 training compounds, 20 test compounds, and 21 different descriptors. The structurally 95% similar compound of 2-acetamido-2-deoxy-beta-D-glucopyranose has been collected from pubchem[13] and molinspiration.com. Using 40 compounds, the linear regression model has been developed. The predictive capability of the QSAR models was evaluated by Correlation coefficient, mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error.

I. INTRODUCTION

Coronavirus disease, which is called as COVID-19 and that is one of the infectious diseases which is infected by newly found coronavirus. The person who is affected by Corona virus will have the problem of moderate respiratory illness. Aged people and people those who are affected by diabetes, cancer, High blood pressure have serious illness because of COVID-19.

History:

Covid-19 is caused by a virus called SARS-COV-2 and the abbreviation of this is Severe acute respirator syndrome coronavirus 2, World Health Organization (WHO) learned the novel virus during 31 december 2019. The disease was called by COVID-19 (coronavirus disease).

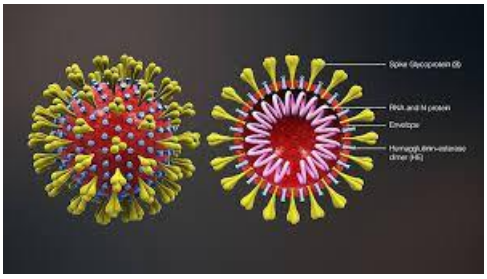

Fig. 1 Coronavirus disease

Prevention:

In order to stop the multiply of COVID-19: The following measures to be followed.

1. People has to clean their hands often by using soap and water, or n alcolhol-based hand rub.
2. Social distancing has to be maintained by the people.
3. Whenever the people find anybody is coughing or sneezing, Safe distance has to be maintained by the people.
4. People has to wear a mask.
5. People has to avoid touching their eyes, nose or mouth.
6. If the people have cough or sneeze, they have to cover their nose with a tissue
7. If the people feel that they are unwell, they have to stay in their homes
8. If the citizens have the problem of fever, cough and breathing, they have to contact the medical officer immediately.

How to make your environment safer:

Public has to avoid the places where it is closed, crowded and involved in close contact. Public has to avoid the places like restaurants, nightclubs, offices and places of worship where the public has close contact with each other. Outdoor gathering is good comparatively with indoor gatherings. Increase the natural ventilation and avoid the closed settings.

Medicine for COVID 19:

For moderate and mild covid attack, Hydroxychloroquine is the approved medicine. If the candidate is attacked in a moderate COVID-19, if they are lack in oxygen, Remdesivir is the medicine, which is used for the people those who lack in oxygen. Convalescent plasma is the medicine which are used for the moderate COVID-19, this is applicable when the demand of oxygen is increased. Tocilizumab is the medicine which is considered for patients who are affected by moderate COVID-19. This medicine is applicable for the people those who lack in oxygen and also applicable for the mechanical ventilation, this is used in spite of steroids. For clinical management of patients with moderate attack, Corticosteroid drugs such as Methylprednisolone and Dexamethasone have been approved for the clinical management of patients with moderate and severe COVID-19. Corticosteroid drugs such as Methylprednisolone and Dexamethasone have been approved for the clinical management of patients with moderate and severe COVID-19. In order to avoid blood clots and thrombogenic response, Low molecular weight Heparin has been given. It may be considered in patients with moderate COVID-19 who are not improving (progressively increasing oxygen requirements and in mechanical ventilation) despite the use of steroids [12].

II.MACHINE LEARNING TO OVERCOME THE COVID'19

Identifying Risk:

People will be affected by various types of risk like: infection, severity and outcome. Using the machine learning, the people can identify the risk of a person or a group which are affected by COVID-19. Severity risk of a person can be identified through machine learning. Severity risk is the one the person who is having the severe COVID-19 symptoms. Hospitalization or intensive care admission can be done after identification. Through the machine learning techniques, outcome risk can be identified. In outcome risk, the medical people can check whether the specific treatement given to the individual or group is effective or not.

Predicting the risk of infection:

The following are the factors which are involved in the infection of COVID'19. Age, Pre-existing conditions, General hygiene habits, Social habits, Number of human interactions, Frequency of interactions, Location and climate, Socio-economic status. We can able to predict, how much the above factors are affecting the people.

Severity Risk:

There are some people are affected severely and some are affected mildly. Mildly affected people even do not have any symptom at all. There are some people are affected with severe lung disease or acute respirator syndrome (ARDS). When a person come to the doctor, the person can be predicted by whether he is affected mildly or severely.

Predicting Outcomes:

The people can be predicted that, how much is the medicine works. For example, if the person gaves medical care, how much percentage he got cure after the treatment can be predicted.

Drug development:

Machine Learning has the major role in predicting the drugs of the particular disease. Lalmuanawma et al. 2020 has given the application of machine learning and artificial intelligence in COVID'19. It is used to develop the model design,

Compound Details:

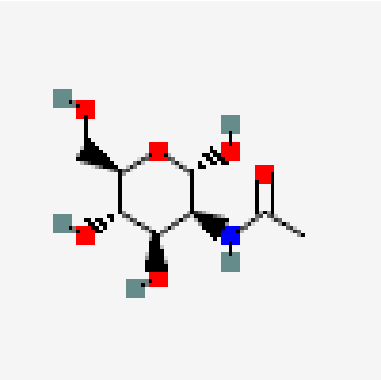

Fig.2 2-acetamido-2-deoxy-beta-D-glucopyranose

The compound 2-acetamido-2-deoxy-beta-D-glucopyranose has been taken from the PDB file 7BZ5 [11]. This is the ligand of COVID. 95% structurally similar to this compound has been taken. 40 structurally compound has been taken as a training set. The cross validation has used as 10. From the PubChem, the molecular properties has been taken as well as the smiles notation has been given as an input to molinspiration online tool and the properties have been calculated. The following attributes has been collected from pubchem Molinspiration property engine and pubchem. 1) miLogP 2) TPSA 3) natoms 4) MW 5) nON, 6) nOHNH 7) nviolations 8) nrotb 9) volume 10) XLogP3, 11) Hydrogen Bond Donor Count 12) Hydrogen Bond Acceptor Count 13) Rotatable Bond Count, 14) Exact Mass 15) Monoisotopic Mass 16) Topological Polar Surface Area 17) Heavy Atom Count 18) Formal Charge 19) Complexity 20) Isotope Atom Count 21) Defined Atom Stereocenter Count 22) Undefined Atom Stereocenter Count 23) Covalently-Bonded Unit Count 24) Compound Is Canonicalized

Preprocessing:

The following properties have been removed, since it has the similar values for all the rows: Formal Charge, Covalently-Bonded Unit Count and Compound Is Canonicalized

III.METHODOLOGY

Machine Learning Techniques:

Machine learning algorithm is useful in the business process and any applications. It will be useful in drug design. In the business, mainly it focuses on prediction of sales. It can able to predict the new drugs. There are some techniques are there in the machine learning: Regression, Classification, Clustering, Association Analysis and Dimensionality reduction etc., There are two types of machine learning like supervised Machine Learning techniques and Unsupervised Machine Learning Techniques. In, supervised Machine learning, the piece of data is predicted. In the unsupervised machine learning, target variable is not used, It relates and group the data points [1-5].

Regression:

Regression is one of the supervised Machine Learning Techniques. It is used to predict the values based on the data given. In order to build a model dataset, the following mathematical equation are used: y=m*x+b. where x and y represent the points and m is the slope. y-intercept(b) is a line which approximates the data observations [6-10] .

Data Flow Diagram:

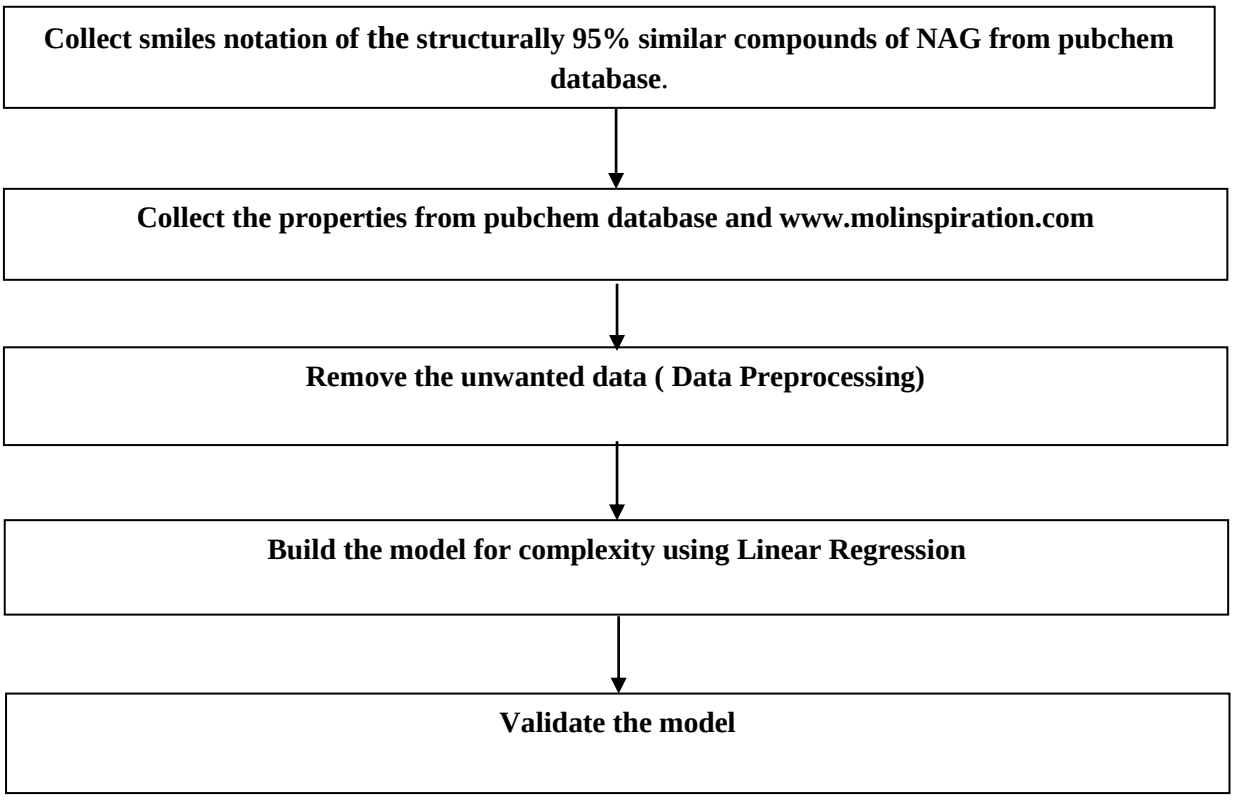

Fig.3 Build the model

Procedure:

1. Collect smiles notation of the structurally 95% similar compounds of NAG from pubchem database.
2. Collect the properties from pubchem database and www.molinspiration.com
3. Remove the unwanted data (Data Preprocessing)
4. Build the model for complexity using Linear Regression
5. Validate the model

IV.RESULT AND DISCUSSION

Following is the formula for the complexity in linear regression.

complexity = -8.0418 * miLogP + 17.4242 * natoms + -50.1303 * nON + -22.3247 * nOHNH + -6.5114 * nviolations + -9.3873 * nrotb + -22.4941 * HBDC + -9.3873 * RBC + 0.0515 * EM + 4.1555 * TPSA + 17.4242 * HAC + -204.7607

In this model, miLogP, nON, nOHNH, nviolations, nrotb, HBDC, RBC are decreasing. For example, miLogp decreases nON is also decreasing. Natoms, EM, TPSA, HAC are increasing. For example Natoms increases EM is also increases. miLog P and natoms are negatively correlated. miLogP and EM are negatively correlated. The compound id 129800705 and 14160777 are negatively correlated with miLogP and EM. The compound id 9549241 and 10944029 are positively correlated in TPSA and HAC. Table 1 represents the validation parameter. Since the correlation coefficient value is 0.6238, the model is acceptable. Based on the graphical representation (Fig. 4), It has been noted that except the 7th component all the other component observed and predicted values are similar. Thus, the model has the good accuracy.

Table 1: Validation

XCorrelation coefficient	0.6238
Mean absolute error	17.4951
Root mean squared error	78.9244

Relative absolute error	47.2166 %
Root relative squared error	166.9907 %

Table 2: Observed vs Predicted value

Observed Complexity	**Predicted Complexity**
321	319.2212
299	308.0846
341	350.0828
235	236.1343
277	269.1069
277	252.939
374	720.2037
248	247.0678
248	248.7566
287	285.8744
248	247.0678
261	258.4825
321	320.9904
235	236.1212
248	247.0678
318	318.14
235	236.0694
231	258.4825

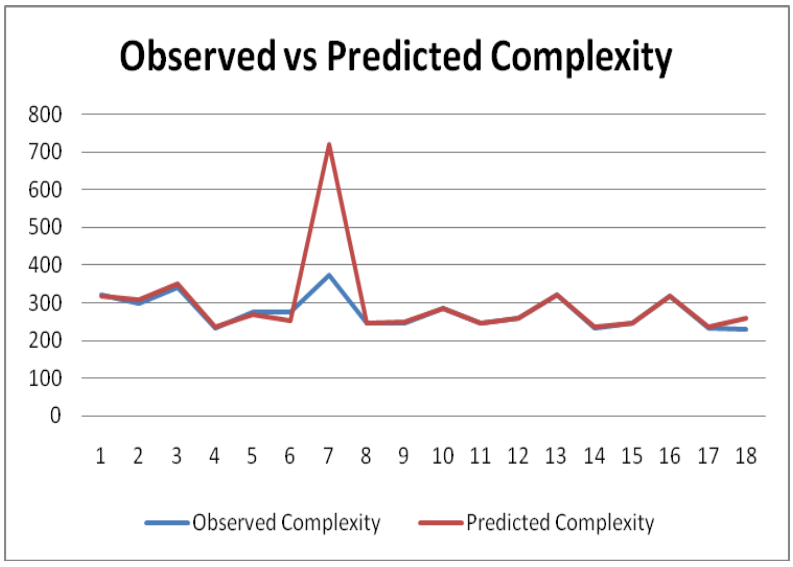

Fig.4 Observed vs Predicted Complexity

V.CONCLUSION

Coronavirus disease, which is called as COVID-19 and that is one of the infectious diseases which is infected by newly found coronavirus. Entire world is affected by the corona virus. This study helps to understand the complexity of the ligand binding with this disease. Since the correlation coefficient value is 0.6238, the model is acceptable. Based on the observed vs predicted values, it has been noted that, the model is valid.

REFERENCES

[1] Pushpalatha, D.P.M., Pushparaj, D.J., Gayathri, R.Correlation analysis between element and aromatic bond count attributes of illicit drugs (2020) Journal of Physics: Conference Series, 1706 (1), art. no. 012032, .

[2] Joseph Pushpa Raj, D., Ponmary Pushpa Latha, D. Dimensionality Reduction of Attributes in order to Predict Parkinson's Disease Using Linear Model (2019) Proceedings of the 2nd International Conference on Smart Systems and Inventive Technology, ICSSIT 2019, art. no. 8987846, pp. 605-608.

[3] Jihad, A.J., Mathew, S.S., Paul, S., Pushpalatha, D.P. Continuous health monitoring using smartphones-A case-study for monitoring diabetic patients in UAE (2017) Proceedings of the 2016 12th International Conference on Innovations in Information

[4] Ponmary Pushpa Latha, D., Poongothai, N. Amino acid exploration in Anopheles gambiae malaria causing target using Association Rule Mining (2015) International Journal of Applied Engineering Research, 10 (20), pp. 16364-16366.

[5] Vasantha Kokilam, K., Ponmary Pushpa Latha, D. A comparative study of parkinson's disease prediction using classification and filtering techniques (2014) International Journal of Applied Engineering Research, 9 (23), pp. 18963-18976.

[6] Ponmary Pushpa Latha, D., Joseph Pushpa Raj, D. Quantitative structure activity relationship study for the prediction of inhibitory concentration 50 for 5-N-acetyl-beta-d-neuraminic acid structurally similar compounds using neural net (2014) Asian Journal of Pharmaceutical and Clinical Research, 7 (4), pp. 173-176.

[7] Ponmary Pushpa Latha, D., Joseph Pushpa Raj, D. Measuring interesting amino acid patterns for alzheimer's disease related studies targets on the binding site using association rule mining (2013) Journal of Applied Pharmaceutical Science, 3 (7), pp. 25-30.

[8] Ponmary Pushpa Latha, D., Jeya Sundara Sharmila, D. Protein-ligand binding data integration for "diabetes disease related studies" in different organism (2013) International Journal of Pharma and Bio Sciences, 4 (2), pp. B51-B61.

[9] Latha, D.P.P., Sharmila, D.J.S. QSAR study for the prediction of half maximal inhibitory concentration of compounds structurally similar to glycerol [Article@Gliserole yapısal olarak benzeyen bileşiklerin yarı maksimal inhibisyon konsantrasyonlarının QSAR çalışması ile tahmini] (2010) Turkish Journal of Biochemistry, 35 (4), pp. 287-292.

[10] Latha, D.P.P., Raj, D.J.P., Sharmila, D.J.S. Generation of unified data structure and data warehouse for Protein Data Banks (2008) Proceedings - International Conference on Computational Intelligence and Multimedia Applications, ICCIMA 2007, 2, art. no. 4426660, pp. 3-7.

[11] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank (2000) Nucleic Acids Research, 28: 235-242.

[12] Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review Samuel Lalmuanawma,a,∗ Jamal Hussain,a and Lalrinfela Chhakchhuakb Chaos Solitons Fractals. 2020 Oct; 139: 110059.

[13] Kim S, Chen J, Cheng T, et al. Pubchem in 2021: new data content and improved web interfaces: (2021), Nucleic Acids Research, 49(D1):D1388-D1395.