International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST21829

Computational in Silico Modelling of Phytochemicals as a Potential Cure

Rachita Kurmi[1], Silpa T S[3], Kavya N R[3], Jennath Sherin A[4]

[1]Shandong University Cheeloo College of Medicine, Jinan, China.

[2]Union Christian College, Aluva, M G University, Kottayam, Kerala, India

[3]Sacred Heart College (Autonomous), Thevara, M G University, Kottayam, Kerala, India.

[4]St.Teresa's College, Ernakulam, M G University, Kottayam, Kerala, India.

Article Info
Volume 8, Issue 2
Page Number : 42-54

Publication Issue
March-April-2021

Article History
Accepted : 01 March 2021
Published : 08 March 2021

ABSTRACT

Urinary tract infection (UTI) is one of the most severe public health problem affecting both sexes but females are more susceptible due to the differences in urogenital and reproductive anatomy, physiology and lifestyle. Due to multi-drug resistant strains and high recurrence rate, UTI has become a major socioeconomic burden. It was found that microbial infections including Enterococcus faecalis, Escherichia coli, Pseudomonas aeruginosa and Proteus mirabilis species are the major causes of UTI with different signs and symptoms including painful urination or dysuria, haematuria, urinary urgency, burning micturition, frequent urination, nausea, and vomiting. Phytochemicals are effective to combat bacterial resistance with high efficacy, and easy availability with minimal or no side effects. For this reason, We present a docking-based screening using a quantum mechanical scoring of a library built from approved drugs and compounds that Curcumin, Anolignan B, Piperine, Carvacrol, Quercetin, Kaempferol, Citral, Allyl Isothiocyanate with Proteins with PDB id's 2N50, 4C4V, 6H1X, 4UU4 could display antibacterial activity against UTI. Clearly, these compounds should be further evaluated in experimental assays and clinical trials to confirm their actual activity against the disease. We hope that these findings may contribute to the rational drug design against UTI.

Keywords : Enterococcus Faecalis, Escherichia Coli, Pseudomonas Aeruginosa, Proteus Mirabilis Species, Urinary Tract Infections

I. INTRODUCTION

Urinary tract infections (UTIs) are the most common bacterial infectious diseases encountered in clinical practice and account for significant morbidity and high medical costs. The prevalence of UTI seems to be a J-shaped distribution, with higher frequency among very young children which gradually increases with age. It is estimated to affect 150 million people each year worldwide, with an annual incidence of 12.6% in women and 3% in men. Although most UTIs can be effectively treated by antibiotics, UTI recurrence is a common problem and sometimes maybe very troublesome. Urinary tract infections may be

asymptomatic, acute, chronic and complicated or uncomplicated, and the clinical manifestations of UTIs depend on the portion of the urinary tract involved.

II. Procedure

1. Ligand Screening

For the initial Ligand screening purposes, a web-based tool named SwissADME (https ://www.swiss adme.ch/) was used to eliminate a few compounds according to Lipinski's rule of five parameters. For a compound to qualify as ligand it should Have < 500 Da molecular weight, a high lipophilicity i.e. value of Log P being less than 5, hydrogen bond acceptors being less than 10 and H-bond donors less than 5. Any compound with more than 2 violations was ruled out for further study (Lipinski2004).

2. Protein Preparation and Active site Determination.

Required protein in pdb format was downloaded from the website **rcsb.org**, commonly known as the **Protein Data Bank**. 3D conformers of the ligand were downloaded from PubChem.

Using **PyMOL (Version 2.4.1)** software water molecules as well as native ligands from the protein were removed, defined as cleaning/purification of the protein for further application. **Using a web server called Deep Site** Active Pockets of the proteins were calculated. The results calculated by the web server were in the form of different ids, centers and scores.

Scoring In deep site was using neural networking based on following instructions using DCNN architecture.

https://academic.oup.com/bioinformatics/article/33/19/3036/3859178 Center values for the grid were selected keeping score greater than 0.98.

UCSF Chimera (Version 1.14) was used to prepare the receptor using DockPrep function. **Dock Prep** prepared structures for Docking using these functions:

- deleting water molecules
- repairing truncated sidechains
- adding hydrogens
- assigning partial charges
- writing files in Mol2 format

3. In silico Docking Using Auto dock Vina

Auto dock Vina (Version 1.1.2) along with UCSF Chimera (Version 1.14) was used for molecular **Docking Studies**. Center values and size of the grid of different scores were used from **DEEPSITE** calculations done above.

Following Parameters were set in auto dock vina.

Receptor options –

- **Add hydrogens in Chimera** (**true**/false) – whether to add hydrogens in Chimera before calling the script. The receptor prep script will check for hydrogens and add them if they are missing. AutoDock Vina needs the polar (potentially H-bonding) hydrogens to identify atom types for scoring purposes.
- **Merge charges and remove non-polar hydrogens** (**true**/false) – note AutoDock Vina does not use charges or nonpolar hydrogens, so this setting is not expected to affect results except for the presence or absence of nonpolar hydrogens in the processed receptor
- **Merge charges and remove lone pairs** (**true**/false) – note AutoDock Vina does not use charges or lone pairs, so this setting is not expected to affect results except for the presence or absence of lone pairs in the processed receptor (and there may not have been any lone pairs to start with)

- **Ignore waters** (**true**/false)
- **Ignore chains of non-standard residues** (**true**/false) – ignore chains composed entirely of residues other than the 20 standard amino acids.
- **Ignore all non-standard residues** (true/**false**) – ignore all residues other than the 20 standard amino acids.

For Ligands

- **Merge charges and remove non-polar hydrogens** (**true**/false) – note Auto Dock Vina does not use charges or nonpolar hydrogens, so this setting is not expected to affect results except for the presence or absence of nonpolar hydrogens in the ligand output files
- **Merge charges and remove lone pairs** (**true**/false) – note AutoDock Vina does not use charges or lone pairs, so this setting is not expected to affect results except for the presence or absence of lone pairs in the ligand output files (and there may not have been any lone pairs to start with)

Docking parameters

- **Number of binding modes** (1-10, **10**) – maximum number of binding modes to generate
- **Exhaustiveness of search** (1-8, **8**) – thoroughness of search, roughly proportional to time
- **Maximum energy difference (kcal/mol)** (1-3,**3**) – maximum score range; binding modes with scores not within this range of the best score will be discarded.

The docking results were calculated by Auto dock vina using it's Scoring function and results were displayed in the form of Scores and RMSD values. Docking results with the highest value score accompanied by negative sign and least RMSD values were chosen for further studies.

4. Residue Analysis

PyMOL was used for visualization of interactions of the docked structure at the ligand sites. **Discovery Studio 2020** was used to study the ligand interactions and total number of residues. It was also used to plot the 2D structure of the interactions and residues.

1. **Statistical Analysis:** Descriptive, estimation and Hypothesis testing with confidence interval of 95% was applied to data using formula 1 given below.

$$CI = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

CI = confidence interval
$\bar{x}$ = sample mean
z = confidence level value
s = sample standard deviation
n = sample size

Formula 1 used for calculation of confidence interval

III. RESULTS AND DISCUSSION

Molecular Docking:

The docking result was obtained from Auto dock vina in the form of Dock score for all the four proteins docked with above mentioned ligands.

Urinary Tract Infection caused by four main pathogenic bacteria-

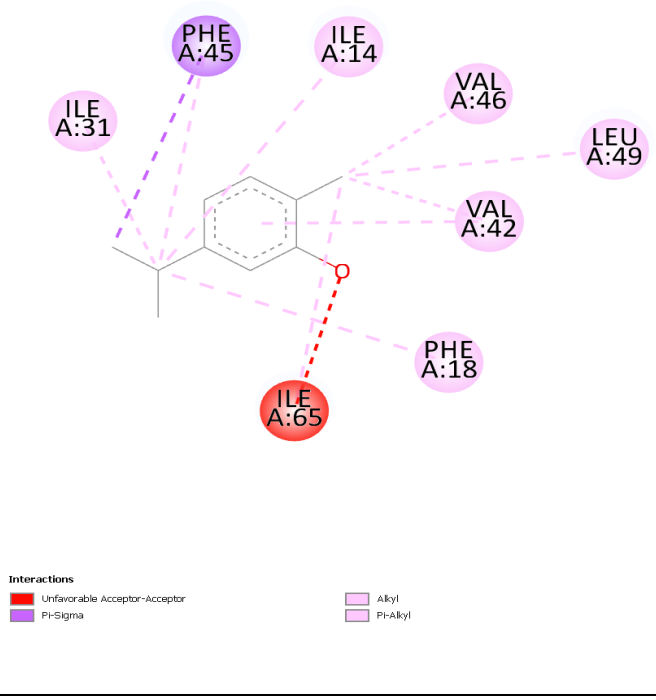
Enterococcus faecalis Target Protein Docking Results:

PDB-ID 2N50

For 2N50, out of the three active sites the 3rd active site was selected with a Deep site score of 0.900. The selection was made on the basis of the highest binding energy of the ligand-receptor. The docking results before statistics are shown in Table 1 and Table 2 shows the post statistical docking scores with Ligand Protein Interactions.

LIGANDS	DOCKSCORE
Citral	-6.2
Allyl isothiocyanate	-4.0
Curcumin	-6.0
Carvacrol	-7.8
Kaempferol	-6.7
Quercetin	-6.9
Piperine	-5.9
Anolignan B	-4.8

Table 1

Ligands	Dockscore	Interactions
Carvacrol	-7.8	

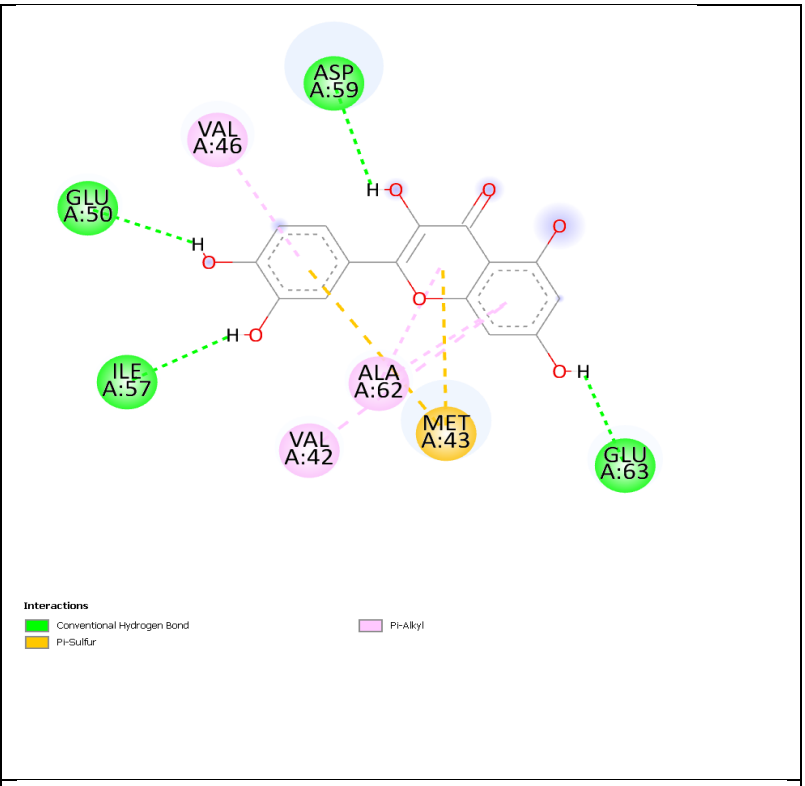
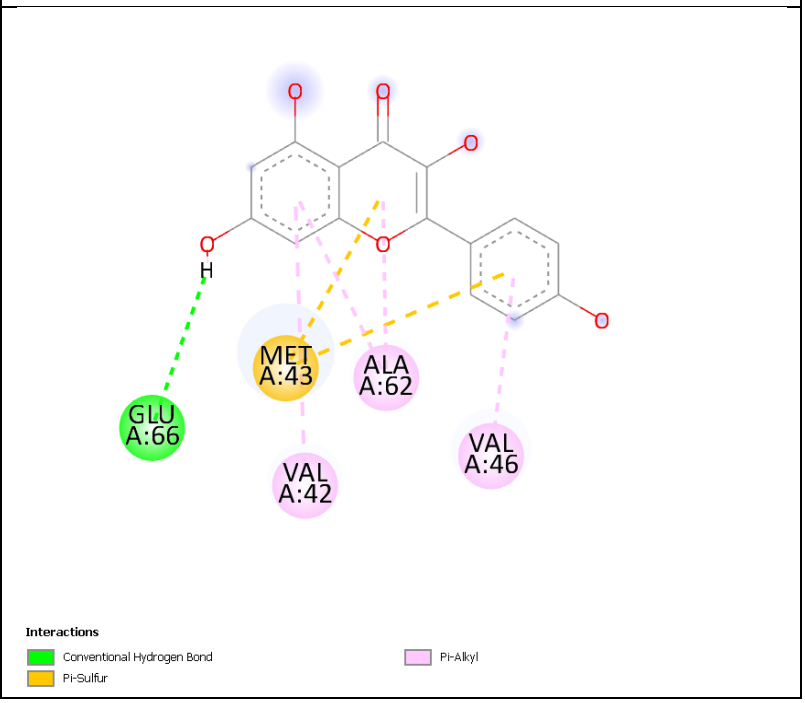
Quercetin	-6.9	
Kaempferol	-6.7	

Table 2

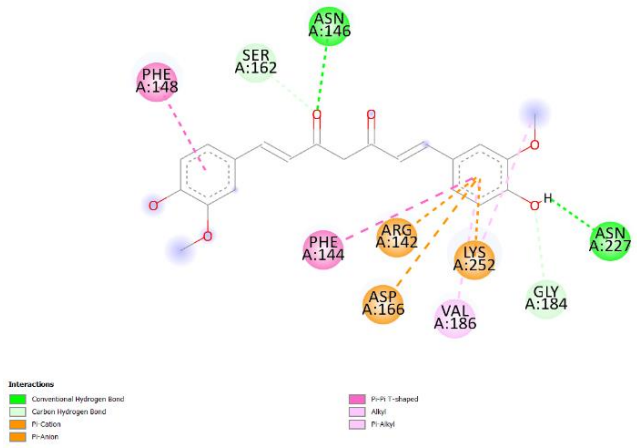
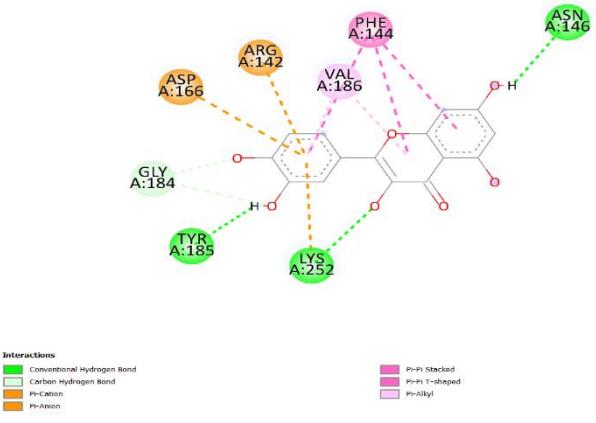
Escherichia coli Target Protein Docking Results:

PDB-ID 4C4V

For 4C4V, out of the three active sites the 2nd active site was selected with a Deep site score of 0.999. The selection was made on the basis of the highest binding energy of the ligand-receptor. The docking results before statistics are shown in Table 1 and Table 2 shows the post statistical docking scores with Ligand Protein Interactions.

LIGANDS	DOCKSCORE
Citral	-5.3
Allyl isothiocyanate	-3.1
Curcumin	-7.8
Carvacrol	-5.9
Kaempferol	-7.7
Quercetin	-8.1
Piperine	-7.9
Anolignan B	-7.3

Table 3

Ligands	Dockscore	Interactions
Curcumin	-7.8	
Quercetin	-8.1	

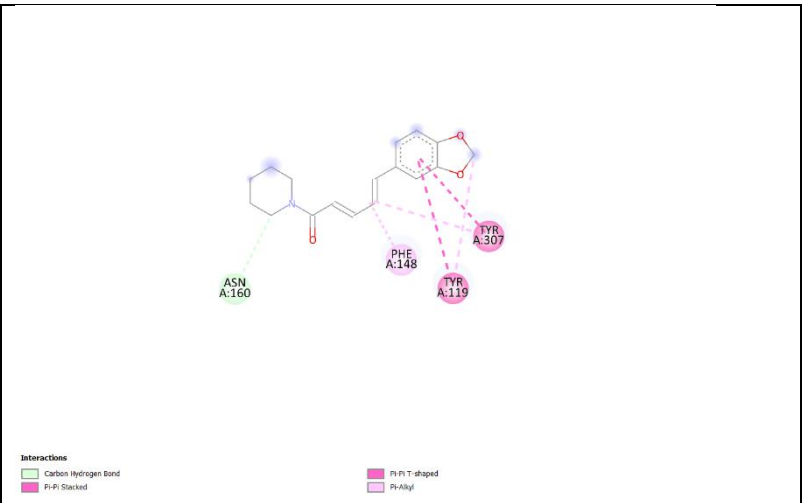
Piperine	-7.9	

Table 4

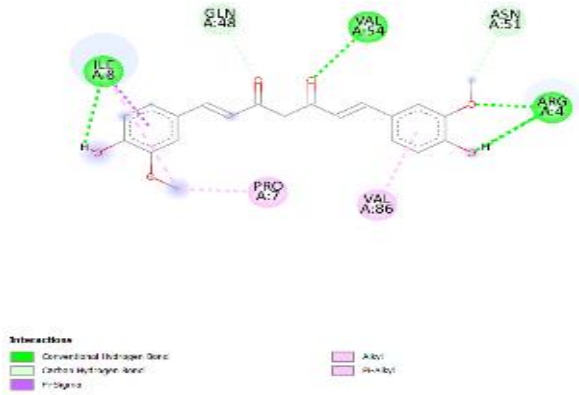
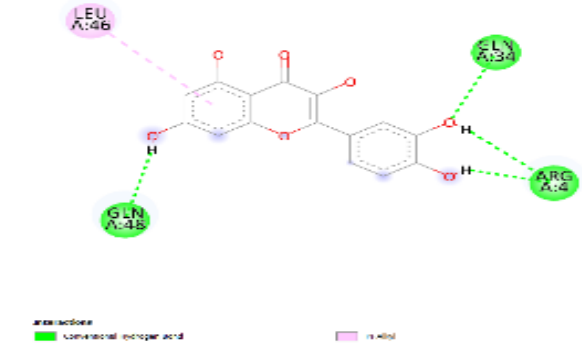
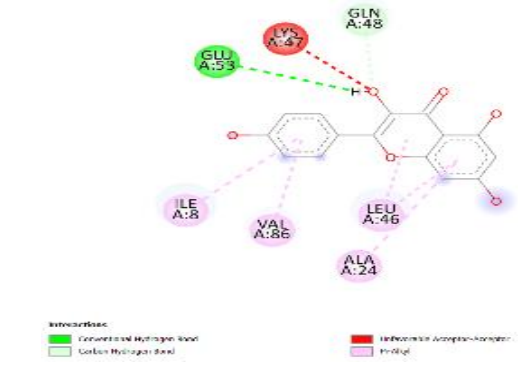
Pseudomonas aeruginosa Target Protein Docking Results:

PDB-ID 4UU4

For 4UU4, out of the four active sites the 1st active site was selected with a Deep site score of 0.991. The selection was made on the basis of the highest binding energy of the ligand-receptor. The docking results before statistics are shown in Table 1 and Table 2 shows the post statistical docking scores with Ligand Protein Interactions.

LIGANDS	DOCKSCORE
Citral	-4.9
Allyl isothiocyanate	-2.9
Curcumin	-7.1
Carvacrol	-5.3
Kaempferol	-6.7
Quercetin	-6.9
Piperine	-6.0
Anolignan B	-6.7

Table 5

Ligands	Dockscore	Interactions
Curcumin	-7.1	
Quercetin	-6.9	
Kaempferol	-6.7	

Anolignan B	-6.7	

Table 6

<u>Proteus mirabilis</u> Target Protein Docking Results:

PDB-ID 6H1X

For 6H1X, out of the four active sites the 2nd active site was selected with a Deep site score of 0.937. The selection was made on the basis of the highest binding energy of the ligand-receptor. The docking results before statistics are shown in Table 1 and Table 2 shows the post statistical docking scores with Ligand Protein Interactions.

LIGANDS	DOCKSCORE
Citral	-3.7
Allyl isothiocyanate	-2.5
Curcumin	-5.5
Carvacrol	-4.0
Kaempferol	-5.7
Quercetin	-5.6
Piperine	-5.3
Anolignan B	-5.2

Table 7

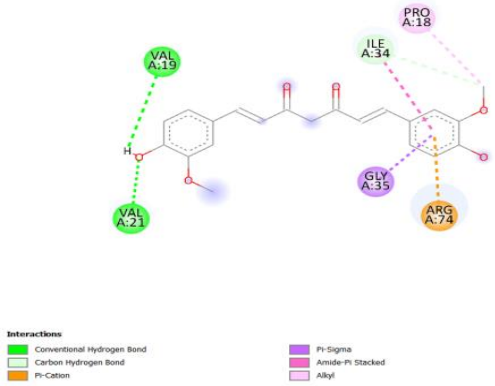
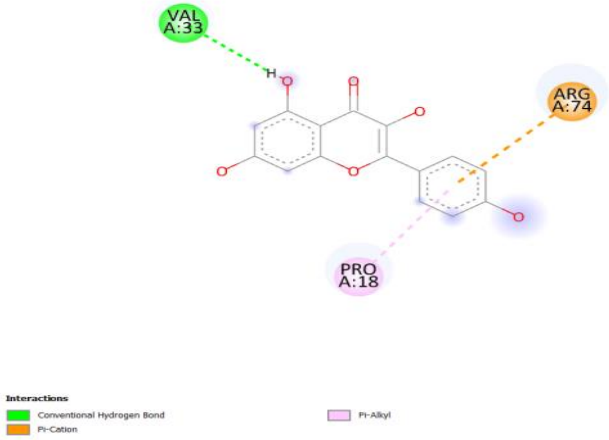
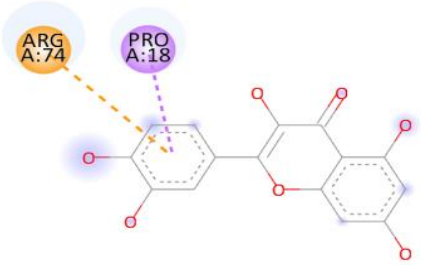
Ligands	Dockscore	Interactions
Curcumin	-5.5	
Kaempferol	-5.7	
Quercetin	-5.6	

Table 8

Ligands	2N50	4C4V	4UU4	6H1X	Number of Proteins interacted
Citral	No	No	No	No	0
Allyl isothiocyanate	No	No	No	No	0
Curcumin	No	Yes	Yes	Yes	3
Carvacrol	Yes	No	No	No	1
Kaempferol	Yes	No	Yes	Yes	3
Quercetin	Yes	Yes	Yes	Yes	4
Piperine	No	Yes	No	No	1
Anolignan B	No	No	Yes	No	1

Table 9. Summarizes the results showing ligands and their interacted proteins that were considered in the study

IV. CONCLUSION

All eight ligands were studied using bioavailability radar. Our results proposed that Quercetin shows the best docking results with all the four target proteins. Curcumin and Kaempferol also depicts successful interaction with proteins with 2N50 and 4C4V as exceptions respectively. Carvacrol, Piperine and Anolignan B shows specific interaction with one protein each-2N50, 4C4V and 4UU4 respectively. Citral and Allyl Isothiocyanate didn't show standardized results with any of the proteins included in the study. To find the effectiveness and to propose the exact mechanism in-vitro studies can be encouraged on Quercetin, Curcumin and Kaempferol targeting respective pathogenic bacteria that are discussed above to understand the mechanism and a potential cure for Urinary Tract Infections.

V. ACKNOWLEDGMENT

We would like to thank our supervisor/Guide Bharat Kwatra, from Invenzion Labs Inc. whose expertise was invaluable in formulating the research questions, methodology and drawing conclusions. His insightful feedback and guidance pushed us to sharpen our thinking and brought our work to a higher level

VI. REFERENCES

[1]. Xuefang Xu, Yaguo Lei, Zeda Li(2020), "An Incorrect Data Detection Method for Big Data Cleaning of Machinery Condition Monitoring" published in IEEE transactions on industrial electronics.

[2]. Mohammad Mahdavi, Felix Neutatz, Larysa Visengeriyeva, and Ziawasch Abedjan,(2019), "Towards Automated Data Cleaning Workflows" published in in IEEE transactions on industrial electronics.

[3]. William E Winkler, "Data Cleaning Methods", U.S. Bureau of the Census Statistical Research, Washington, DC.

[4]. Raju Dara,Dr.Ch. Satyanarayana, Dr.A.Govardhan(2013), "Front End Data Cleaning And Transformation In Standard

Printed Form Using Neural Models" published in International Journal on Computational Sciences & Applications (IJCSA).

[5]. Jaya Bajpai Pravin S. Metkewar(2016), "Data Quality Issues and Current Approaches to Data Cleaning Process in Data Warehousing" published in GRD Journals- Global Research and Development Journal for Engineering.

[6]. H. R. Bhapkar, Parikshit N. Mahalle, Nilanjan Dey, and K. C. Santosh(2020), "Revisited COVID-19 Mortality and Recovery Rates: Are we missing Recovery Period" published in Journal of Medical Systems, Springer.

[7]. https://stackoverflow.com/questions/47230817/plotly-notebook-mode-with-google-colaboratory

[8]. https://plotly.com/python/multiple-axes/

[9]. GDP datasets for prediction and forecasting: https://www.macrotrends.net/countries/IND/india/gdp-gross-domestic-product

[10]. Recovery and Death Numbers Dataset: https://api.covid19india.org/documentation/csv/

[11]. Shinde, G. R., Kalamkar, A. B., Mahalle, P. N., Dey, N., Chaki, J., & Hassanien, A. E. (2020). Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. SN Computer Science, 1(4). doi:10.1007/s42979-020-00209-9.

[12]. Simon James Fong1, Gloria Li , Nilanjan Dey, Rubén González Crespo , Enrique Herrera-Viedma5.Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 6, No.1, DOI: 10.9781/ijimai.2020.02.002.

[13]. Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics, 15(1), 276. doi:10.1186/1471-2105-15-276.

[14]. Neeraj Poonia, Sarita Azad. Short-term forecasts of COVID-19 spread across Indian states until 1 May 2020.

[15]. Chimmula, V. K. R., & Zhang, L. (2020). Time Series Forecasting of COVID-19 transmission in Canada Using LSTM Networks. Chaos, Solitons & Fractals, 109864. doi:10.1016/j.chaos.2020.109864.

[16]. Yoshiro Suzuki, Ayaka Suzuki, Shun Nakamura, Toshiko Ishikawa, Akira Kinoshita.Machine learning model estimating number of COVID-19 infection cases over coming 24 days in every province of South Korea (XGBoost and MultiOutputRegressor).

[17]. Sujath, R., Chatterjee, J. M., & Hassanien, A. E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. Stochastic Environmental Research and Risk Assessment. doi:10.1007/s00477-020-01827-8.

[18]. Pai, C., Bhaskar, A., & Rawoot, V. (2020). Investigating the dynamics of COVID-19 pandemic in India under lockdown. Chaos, Solitons & Fractals, 138, 109988. doi:10.1016/j.chaos.2020.109988.

[19]. Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., Du, J., Mohammed, S., … Mukherjee, B. (2020). Predictions, Role of Interventions and Effects of a Historic National Lockdown in India's Response to the the COVID-19 Pandemic: Data Science Call to Arms. Harvard Data Science Review. https://doi.org/10.1162/99608f92.60e08ed5

[20]. Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, Nanda C, Sharma S, Sharma YD, Rabaan AA, Rahmani J, Kumar P. Prediction of the COVID-19 Pandemic for the Top 15 Affected Countries: Advanced Autoregressive Integrated Moving Average (ARIMA) Model. JMIR Public Health Surveill 2020;6(2):e19115.doi: 10.2196/19115

[21]. AutoRegressive Integrated Moving Average – https://www.investopedia.com/terms/a/autoregr

essive-integrated-moving-average-arima.asp#:~:text=What%20Is%20an%20Autoregressive%20Integrated,or%20to%20predict%20future%20trends.

[22]. Introduction to ARIMA: nonseasonal models - https://people.duke.edu/~rnau/411arim.htm

[23]. Understanding LSTM Networks - https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[24]. Borchani, H., Varando, G., Bielza, C., & Larrañaga, P. (2015). A survey on multi-output regression. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5(5), 216–233. doi:10.1002/widm.1157

[25]. An End-to-End guide to Understand the Math behind XGBOOST – https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/

[26]. Marcek, D. and M. Marcek, 2006. Neural Networks and their Applications. EDIS Publ., Slovakia.

[27]. Yunhai Wang, Fubo Han, Lifeng Zhu, Oliver Deussen, and Baoquan Chen, "Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series" , IEEE Transactions on Visualization and Computer Graphics 2017.

[28]. Benjamin Born, Alexander M. Dietrich, Gernot J. Müller, "The lockdown effect: A counterfactual for Sweden", July 2020, CEPR Discussion Papers 14744, C.E.P.R. Discussion Papers.

[29]. Mahipal Jadeja, Kesha Shah "TREE-MAP: A VISUALIZATION TOOL FOR LARGE DATA" Published in GSB@SIGIR 2015.

[30]. Niyazi ARI, "Matplotlib In Python", 2014 11th International Conference on Electronics, Computer and Computation (ICECCO), doi: 10.1109/ICECCO.2014.6997585

[31]. "What is data visualization" https://www.klipfolio.com/resources/articles/what-is-data-visualization#DataViz2

[32]. "The impact of COVID-19 — data visualization using Plotly and comparative analysis with SARS" https://towardsdatascience.com/the-impact-of-covid-19-data-analysis-and-visualization-560e54262dc

[33]. "What is Data Visualization and Why Is It Important?" https://www.import.io/post/what-is-data-visualization/

[34]. "Treemaps in python" https://plotly.com/python/treemaps/

[35]. "Polar chart" https://doc-archives.microstrategy.com/producthelp/10.6/AdvancedReportingGuide/WebHelp/Lang_1033/Content/AdvancedReporting/Polar_chart.htm

[36]. Website-investopedia.com/terms/d/datamining.asp

[37]. Gunther Schuh, Jan-Philipp Prote,Philipp Hunnekes(2019),"Data Mining Methods For Macro Level Process Planning",13th CIRP Conference on Intelligent Computation in Manufacturing Engineering.

Cite this article as :

Rachita Kurmi, Silpa T S, Kavya N R, Jennath Sherin A, "Computational in silico modelling of phytochemicals as a potential cure", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 2, pp. 42-54, March-April 2021. Available at doi : https://doi.org/10.32628/IJSRST21829
Journal URL : https://ijsrst.com/IJSRST21829