Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

RANDOM FOREST MODELING OF MOLECULAR DESCRIPTORS OF COX-2-TARGETED NON-STEROIDAL ANTI-INFLAMMATORY DRUGS (NSAIDS)

Liza Tybaco Billones[1]*, Alex Cerbito Gonzaga[1]

1. *Department of Physical Sciences and Mathematics, College of Arts and Sciences University of the Philippines Manila, Padre Faura, Ermita, Manila, 1000 Philippines.*

ARTICLE INFO

Received:
30 Aug 2022
Received in revised form:
16 Dec 2022
Accepted:
16 Dec 2022
Available online:
28 Dec 2022

Keywords: Molecular descriptors, NSAID, COX-2 inhibitors, Random forest, Anti-inflammatory

ABSTRACT

The discovery of next-generation non-steroidal anti-inflammatory drugs (NSAIDs) remains an active area of research as over a billion people suffer from pain and inflammation. A strategic approach in this endeavor is establishing a quantitative relationship between the anti-inflammatory activity and the molecular descriptors of inhibitors of cyclooxygenase-2 (COX-2) that will streamline and expedite the discovery and the subsequent development of novel NSAIDs devoid of side effects associated with COX-1 inhibition. In this work, Random Forest (RF) technique was implemented to formulate a robust quantitative model that predicts the inhibitory activity of compounds on COX-2. The model established in this work displayed excellent predictive performance on compound classification with 93% accuracy and 0.98 AUC. Upon application to two external sets, 759 newly designed derivatives of COX-2 inhibitors and 188 structurally similar compounds were predicted active; 19 of them were found to be promising leads as COX-2-acting anti-inflammatory drugs. The top 2 hits with the highest probability of being active were also found to have the strongest binding affinity with COX-2 and are superior to the known COX-2 selective inhibitors. The RF model is likewise conservative in identifying compounds as active making it all the more beneficial as it helps avoid costly failures at the later stages of the drug discovery phase.

To Cite This Article: Billones LT, Gonzaga AC. Random Forest Modeling of Molecular Descriptors of COX-2-Targeted Non-Steroidal Anti-inflammatory Drugs (NSAIDs). Pharmacophore. 2022;13(6):106-14. https://doi.org/10.51847/OkYCPAEXPR

Introduction

Inflammation is a serious public health concern affecting over 1.5 billion people worldwide [1]. Its symptoms include heat, pain, redness, swelling, and loss of function [2]. It is associated with many chronic diseases, such as diabetes, cancer, cardiovascular, respiratory, and autoimmune diseases [3-6]. These debilitating conditions have consequential ramifications on the patient's quality of life [7, 8].

One of several mechanisms of action of anti-inflammatory drugs involves the inhibition of arachidonic acid metabolism, which is mediated by cyclooxygenase (COX) enzymes, particularly COX-1 and COX-2 [9-12]. These two isozymes are almost identical sequence-wise that only differ in the replacement of isoleucine at position 523 in COX-1 with valine in COX-2 [13]. Isoleucine is bigger than valine and consequently blocks the bulkier molecules (that easily bind with COX-2) from entering the sterically hindered side binding pocket of COX-1.

COX-1 is a constitutive enzyme [14] that is crucial in maintaining tissue homeostasis and is particularly responsible for the production of natural mucus lining that protects the inner stomach [15, 16]. A drug that inhibits COX-1 would likely manifest adverse effects such as gastric ulceration due to reduced production of cytoprotective prostaglandins in the stomach. In contrast, the inducible COX-2 [14] is expressed only in cells with inflammation. Therefore, those drugs that selectively act on COX-2 would not cause the side effects associated with COX-1 inhibition [17].

The traditional NSAIDs are non-selective; that is, they work by inhibiting the activity of both COX-1 and COX-2. The newer NSAIDs, particularly the so-called "coxibs" [18-20], are remarkably selective to COX-2. In general, the available NSAIDs in the market have an array of undesirable side effects specific to a particular drug [21, 22]. Thus, the discovery of new classes of anti-inflammatory compounds with only minimal or mild side effects is still an active area of research.

A prudent technique in drug discovery involves designing or discovering new chemical structures based on known active compounds. It entails the development of quantitative models of biological activity as a function of molecular properties. An

Corresponding Author: Liza Tybaco Billones; Department of Physical Sciences and Mathematics, College of Arts and Sciences University of the Philippines Manila, Padre Faura, Ermita, Manila, 1000 Philippines. E-mail: ltbillones@up.edu.ph.

increasingly becoming popular classification technique is Random Forest (RF) [23]. This machine learning algorithm works by consolidating the outputs of multiple decision trees, i.e. forest, to determine the classification of, for example, compounds as active or inactive against a certain molecular target. RF has been successfully applied in mining biological [24, 25] and medical [26, 27] data. In this work, a Random Forest (RF) model was established and then applied to two external sets of compounds the Derivatives and the Similars, which were derived from or structurally similar to known COX-2-acting NSAIDs, respectively.

Materials and Methods

The compounds with experimental COX-2 inhibitory data were gathered from online literature using keywords such as COX inhibitors, cyclooxygenase inhibitors, COX1/COX2 compounds, and the like. The range of collected articles was found to have been published during the period from 1997 to 2019. The compounds were pre-assigned to two groups based on their IC_{50} value: Active ($IC_{50} \leq 10$ μM) and Inactive ($IC_{50} > 10$ μM). IC_{50} is the compound concentration at which the original enzyme activity is reduced to half.

The optimization of the compound structures and the calculation of their properties were performed in a PC running on Microsoft Windows 7 Professional 64-bit OS using a 3.50-GHz Intel® Core™ i7-4770K processor with 8.00-GB RAM. The data processing and analysis were performed in a machine with macOS Catalina operating system, 3.1-GHz Dual-Core Intel Core i7 processor, and 16-GB RAM. ChemDraw Professional 16.0 (www.perkinelmerinformayics.com) was used to draw the chemical structures, which were saved as structure-data files (.sdf). Discovery Studio (DS) version 4.0 (Biovia, Inc.) was used to convert the 2D to 3D structures and optimize them at the molecular mechanics level using the Dreiding force field [28]. The molecular descriptors were calculated using Spartan 16 (www.wavefun.com) and DS 4.0 software.

All data modeling and data analyses were done using RapidMiner Studio 9.7.001 (www.rapidminer.com). At the onset, a set of 276 compounds (20% of the dataset) was set aside for the test set and the remainder was allocated for the train set. In training the model a 2-fold validation was employed, i.e., only 80% of the train set was used for fitting the initial model and the remaining 20% for the validation set was used in tuning the hyperparameters. Then the optimized random forest model was trained on the full train set of 1104 compounds. Model accuracy, specificity, sensitivity, and AUC were used as evaluation tools.

The RF model was applied to two external sets, the Derivatives, and Similars, to predict their compound activity. The Derivatives emanated from the scaffold of the top 5 families with the most number of compounds. The Similars are compound hits from the library of bioactive compounds (ChEMBL) and the druglike compounds in the ZINC database as outputs of similarity search using the SwissSimilarity (www.swisssimilarity.ch) tool and with the most active compound in each family of inhibitors as query molecule. The chemical structures of the compounds were generated, their molecular properties calculated, and then their compound classes were predicted by the RF model.

The ADMET (absorption, distribution, metabolism, excretion, toxicity) properties of the predicted active compounds were determined *in silico* using the *ADMET and TOPKAT* protocols in DS. The QED (Quantitative Estimate of Druglikeness) scores were also calculated in DS using the *Calculate Molecular Properties* protocol. SwissADME (http://www.swissadme.ch) was used to determine the Synthetic Accessibility (SA) scores.

The molecular geometry of the top hits was optimized at the semi-empirical PM3 level using the equilibrium conformer as starting structure. Each structure was saved as a pdb file. On the other hand, 100-ns Molecular Dynamics simulations [29] were performed on the COX-2 enzyme target (PDB ID: 5IKR) and the equilibrated structure was used in subsequent Molecular Docking studies with the use of Autodock Vina [30] in PyRx (www.pyrx.sourceforge.io).

Results and Discussion

The similarity in the method [14] of experimental COX-2 activity measurement was the primary consideration in selecting the journal articles from which the compounds were collected. **Table 1** shows the list of families of compounds that were gathered from 66 accounts obtained from 6 prominent scientific journals [31]. The 59 families constituted the 1380 collected compounds, of which 929 (67%) were considered COX-2 actives and 451 (33%) compounds were inactive.

Over 400 molecular descriptors were calculated for each compound. The Discovery Studio suite furnished 397 descriptors of which 333 are 2D and 64 are 3D descriptors. The Spartan 16 software contributed 28 descriptors; 9 molecular, 14 QSAR, and five thermodynamic. After data cleaning and removing descriptors with several *NAN* (not a number) entries and those with practically constant values, the number of variables was reduced to 184.

Table 1. Compounds by Family and by Experimental COX-2 Inhibitory Activity, from literature, published 1997-2019

No.	Family	Actives	Inactives	Total
1	1,2-Diarylpyrroles	23	17	40
2	1,2-Diarylimidazoles	82	13	95
3	1,2-Arylhetero-arylimidazoles	37	11	48
4	1,2-Diarylcyclopentenes	44*	4	48

5	Terphenyls	42	7	49
6	1,5-Diarylpyrazoles	77	31	108
7	Diarylspiro[2.4]alkenes	33	1	34
8	4,5-Diarylisoxazoles	3	0	3
9	Pyrazoles	12	0	12
10	Pyrazolopyrimidine	18	0	18
11	Celecoxib-Tolmetin hybrids	11	0	11
12	Pyrazole Derivatives	11	9	20
13	Tetrazoles	4	17	21
14	Cyclic imides	16	45	61
15	Dihydropyrazoles	20	7	27
16	Pyrazole-Thiadiazole hybrids	12	6	18
17	Hydrazones, Pyrazoles	11	8	19
18	Pyrazoles, Salicylamides, Pyrazolo[1,2-a]pyridazines	6	5	11
19	Indoles	5	5	10
20	Benzoxazole benzamides	27	3	30
21	Pyrazolones	11	0	11
22	Triarylpyrazolines	16	0	16
23	Quinoline-2-carboxamides	14	0	14
24	Naproxene derivatives	14	0	14
25	Chalcones	12	0	12
26	Indoles, standards	5	1	6
27	Isoindolines	12	0	12
28	Pyrazolo[3,4-b]pyridines	24	0	24
29	Indole-3-glyoxamides	21	0	21
30	Dihydro-pyrazolyl-thiazolinones	15	5	20
31	1,5-diarylpyrazole-Chrysin hybrids	30	0	30
32	2-Imidazolines	15	15	30
33	Tetrahydropyrans	2	5	7
34	Benzenesulfonamides, Benzisothiazolones	14	0	14
35	Pyrazoles	0	8	8
36	Phenylazobenzenes	3	9	12
37	Alkyldiaryl (E)-olefins	4	1	5
38	Mercaptobenzothiazole-oxadiazole hybrids	9	12	21
39	Carboximidamides, Aryloxadiazoles	12	0	12
40	Triazine-4-aminophenyl-morpholine-3-ones	14	8	22
41	Diarylketones, Diarylamines	8	8	16
42	Diarylthiazoles, Diarylimidazoles	6	10	16
43	Carprofen derivatives	1	32	33
44	Benzamides	0	27	27
45	Pyran-2-ones	36	20	56
46	Tetrahydropyrans	18	0	18
47	Chrysin-Indole hybrids	10	0	10
48	Urea-Pyrazole hybrids	13	7	20
49	Nimesulides	15	11	26
50	Phenoxyphenyl pyrrolidines	1	25	26
51	Coxib analogues	6	0	6
52	Isoxasolines	8	2	10
53	Methyl oxazoles	8	3	11
54	Ethanesulfohydroxamic acid esters	3	2	5
55	Benzylidenes	11	11	22
56	Thiadiazoles, Oxadiazoles	14	24	38

57	Diazenium diolates	0	6	6
58	Indomethacin derivatives	14	1	15
59	Propynones	16	9	25
	Total	**929**	**451**	1380

*2 are standards; not cyclopentenes

The series of runs on the 80% chunk of the train set with concomitant validation on the remaining 20% indicated that the specificity (along with accuracy and sensitivity) of the Random Forest model was maximized at r = 0.75, as shown in **Figure 1**. Consequently, only those variables with at most 0.75 correlation coefficient with each other were used for the model construction; that is, only 64 out of the original set of 184 variables were included in the optimized model. The implicit importance of these descriptors to the model is shown in **Figure 2**, in order of decreasing weights. Among the highest are molecular weight (*wt1*), shadow_z-length (*sz*) or the length of the molecular shadow along the z-axis, the frontier orbitals *eho* and *elu* (E_{HOMO}, E_{LUMO}), AM1 energy (*am1*), polar surface area (*psa*), and dipole moment (*dip*). **Figure 3** shows that *Information Gain* [32] is the best node-splitting criterion in creating the trees. The maximal depth is 14, with the minimum classification error at this value.

Information Gain, Gain Ratio, and Gini Index are the criteria that can be used in selecting the variable that would be used in splitting a node. Information Gain is the reduction in information entropy, which measures the impurity of the nodes with lower values indicating lower entropy or purer nodes [33].

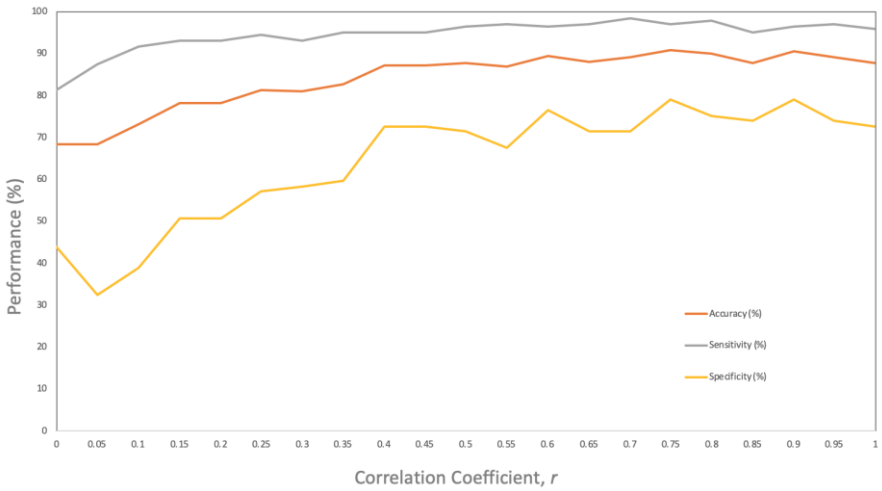

Figure 1. The Random Forest model specificity, sensitivity, and accuracy as determined by the maximum correlation coefficient (r) allowed among the independent variables, using ntree = 100 and maximal depth = 15, with information gain as the splitting criterion.

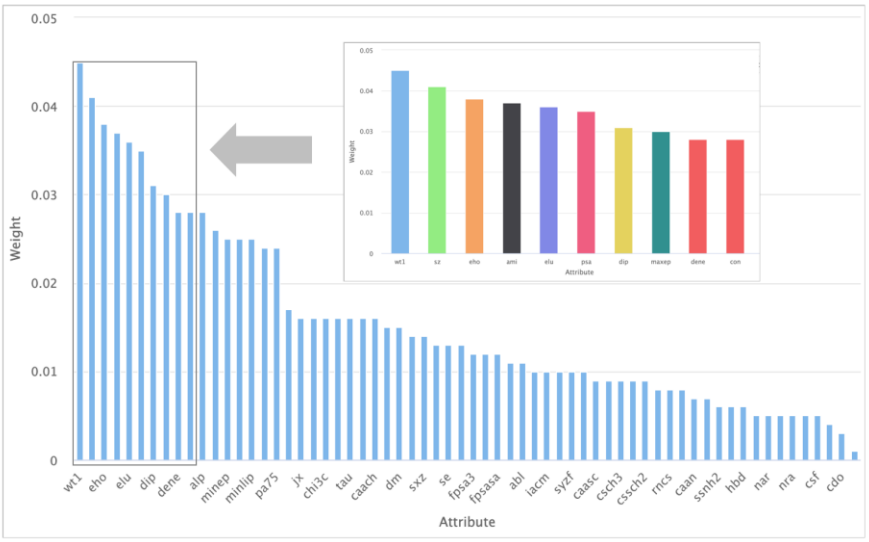

Figure 2. The descriptors in the Random Forest model by their importance in generating the compound class prediction.

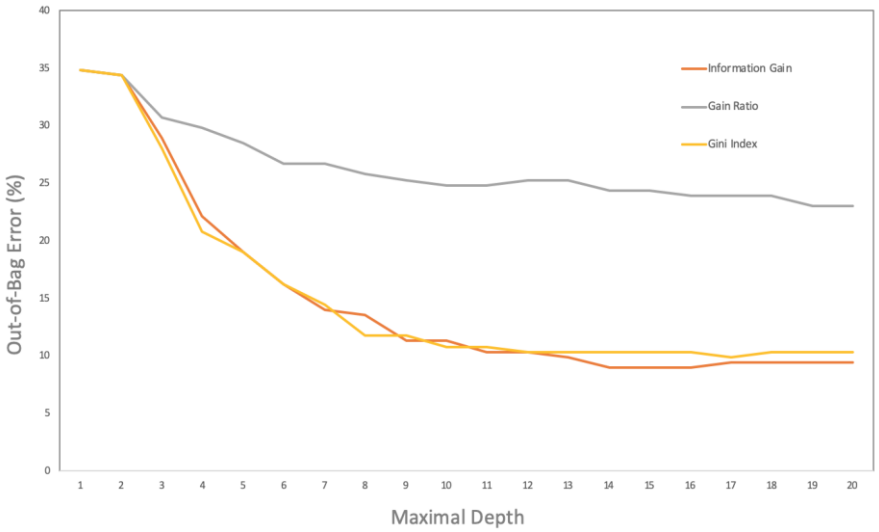

Figure 3. The classification error (%) for the different splitting criteria (information gain, gain ratio, and gini index) is determined by the maximal depth.

With Gini Index [34], the minimum classification error was achieved at the maximal depth of 17, which is three levels more tree branching compared to that of the Information Gain. The Gini Index measures the inequality of the values in the node, smaller values of the Gini Index indicate lower entropy or purer nodes. As regards the Gain Ratio criterion [35], the classification error did not stabilize even at a tree depth of 20.

Considering the results of the model optimization studies above, the final model was generated in the full train set comprised of 80% of the dataset, involving the descriptors that satisfy $r \leq 0.75$, with Information Gain as a splitting criterion at the maximal depth of 14 for all the 100 decision trees in the forest. Delightfully, the resulting Random Forest model exhibited excellent prediction performance (93% overall accuracy). It correctly classified 182 of the 186 actives (98% sensitivity) and 75 of the 90 inactive (83% specificity) as shown in **Figure 4**.

The area under the curve (AUC) of the ROC curve (or AUC-ROC) is another classification model performance metric. A model that perfectly performs in the classification task will have an AUC equal to 1. **Figure 5** shows that the AUC-ROC for the RF model is 0.98 indicating that the model can almost perfectly distinguish active from inactive.

Meanwhile, a set of compounds so-called Derivatives were designed by modifying the most active compound in the five selected classes of COX-2 inhibitors. A total of 1100 compounds were virtually created based on the structural motifs of cyclopentenes, imidazolyls, difluorobenzenes, furanyl/thiophenyls, and isoxazoles. The other set, Similars, is a collection of 600 hits from ChEMBL bioactives and ZINC Drug-like databases on the SwissSimilarity website. The query structures used in the similarity search were the most active compound in each family of known COX-2 inhibitors.

Figure 4. The Random Forest model class prediction of the test set of compounds.

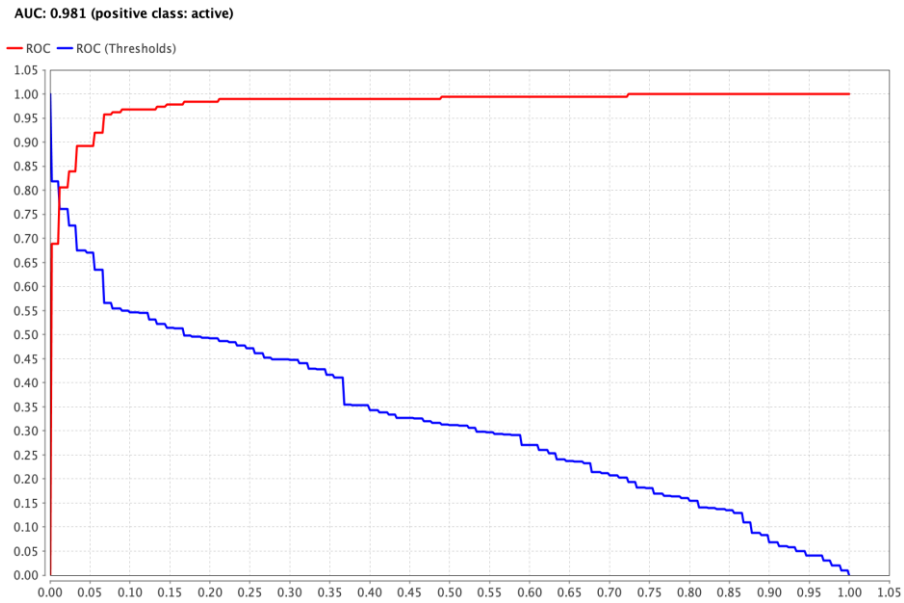

Figure 5. The Random Forest model receiver operator characteristic (ROC) plot.

When the RF model was applied to the Derivatives, 69% (759 out of 1100) of the designed structures were predicted to be active against COX-2. It can be observed that the cyclopentene variants (compounds 1 – 300) were the ones predicted to be the most active, whereas the diflourobenzenes (compounds 501 – 700) were the most inactive. Notably, the RF model gives a conservative estimate of the probability that a compound will be active as compared to the Multiple Logistic Regression (MLogR) model [31]. Nevertheless, identifying more derivatives as inactive (vis-à-vis the MLogR account) might be considered advantageous in drug discovery efforts that employ early-stage elimination of potentially inactive compounds.

For the Similars, only 31% (188 out of 600) were predicted to be active. The average similarity score of the Similars that were predicted active was only 0.38 indicating remarkable structural differences between the query structures and the top hits from ChEMBL and ZINC databases. It is therefore not surprising that relatively fewer Similars were classified as active.

The group of compounds provided new scaffolds that may pave the way for the discovery of new classes of COX-2 selective NSAIDs. Additionally, the tagging of the majority of Similars as COX-2 inactive is beneficial as it reduces the attrition rate, thereby diminishing the cost of drug discovery and development.

The predicted active compounds were further evaluated *in silico* to determine their drug-likeness and synthetic accessibility scores, and also, other drug development parameters. Over 93% of the predicted active Derivatives have a Quantitative Estimate of Druglikeness or QED score [36] above 0.5, i.e., druglike. They are relatively easy to synthesize having average synthetic accessibility (SA) score of 3.3, and all within the 1-6 acceptable range [37]. Most of them have low to optimal aqueous solubility and good to moderate intestinal absorption. All are non-inhibitors of CYP2D6 and non-mutagens, mostly non-carcinogens (89%), although all are hepatotoxic. Likewise, the active Similars are all synthetically accessible having SA scores that range from 2.1 to 5.6. The majority (55%) possess druglike properties, i.e., QED score above 0.5, have acceptable solubility (75%), are non-carcinogens (81%), non-mutagens (84%), and non-inhibitors of CYP2D6 (92%). Although only 44 compounds (23%) have good intestinal absorption.

The top hits were determined from the pool of predicted actives based on the following criteria: (a) PA > 0.7, (b) QED > 0.5, (c) $1 \leq SAS \leq 6$, (d) $2 \leq AS \leq 4$, (e) $0 \leq IA \leq 1$, (f) Non-Carcinogen, (g) Non-Mutagen, and (h) DTP Non-Toxic, and (i) CYP2D6 Non-inhibitor. Only 13 from the Derivatives and 6 from the Similars passed the requirements **(Figure 6)**.

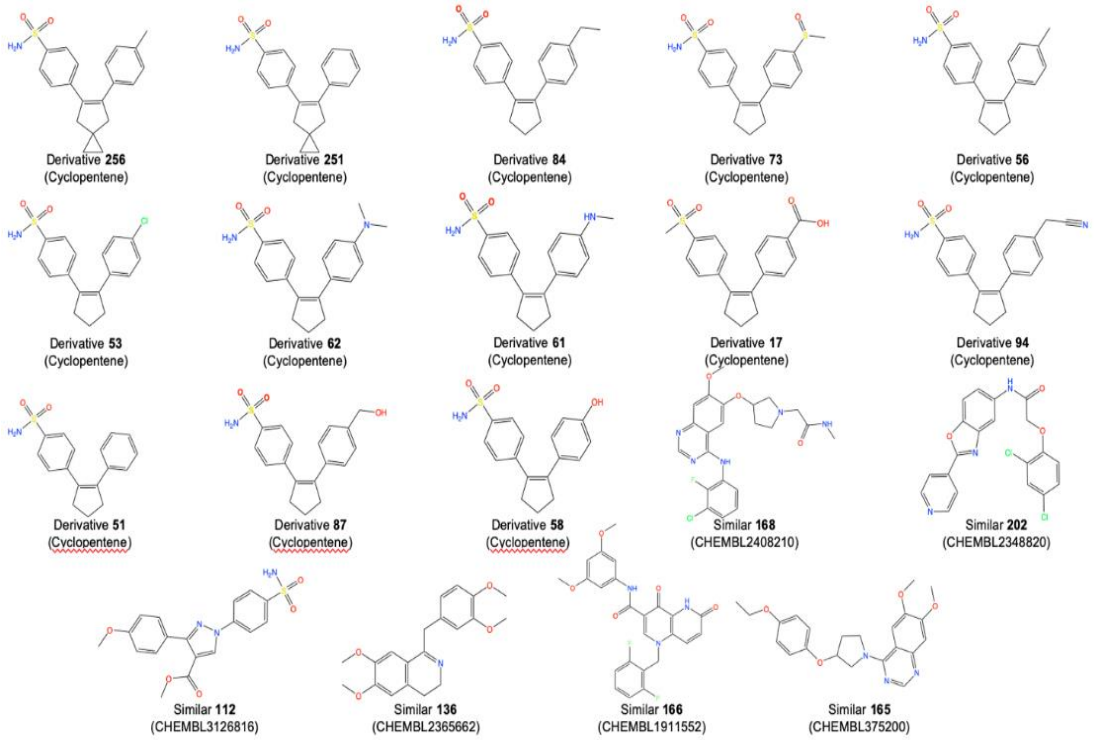

Figure 6. The molecular structures of the top hit from the Derivatives and Similars.

Ten of the top 13 Derivatives were also the top hits found in related MLogR studies [31]. Although they are grouped among the cyclopentane derivatives, the top 2 hits are diarylspiroheptenes D256 and D251. The three additional compounds in the list are D84, D61, and D87, which are also derivatives of cyclopentenes. On the other hand, the top hits from Similars are unique and constitute a different set of compounds compared to the MLogR hits [31].

The above criteria for identifying the hits further lowered the false positive error rate by only including compounds with a high probability of being active (PA>0.7). Moreover, by considering only those with QED scores greater than 0.5, *i.e.* closer to 1 (drug-like) than to 0 (non-drug-like), these top hits likely possess the desirable properties of a drug [36]. They are relatively easy to synthesize in an organic chemistry laboratory, i.e. SAS within the 2-4 range, and have excellent aqueous solubility and intestinal absorption. They do not potentially cause cancers (non-carcinogens) and mutations (non-mutagens), are safe for a developing fetus in pregnant women (except most Similars), and can be taken with other drugs (non-CYP2D6 inhibitors). Molecular docking studies on these hits were also conducted and the results are very promising. All the top 13 hits from the Derivatives and one from the Similars (S202) have binding energies greater than that of Etoricoxib (BE = –7.8 kcal/mol), a known COX-2 selective drug; and have comparable to superior BE values compared to the co-crystallized ligand [38] mefenamic acid (BE = –8.6 kcal/mol).

Impressively, the top 2 hits (i.e. D256 and D251) with the highest probability of being active are also the ones with the greatest BE values or strongest affinities with COX-2 among the top hits identified in this work**.** These two compounds have very high chances of reaching the last stages of the drug discovery and development process.

Conclusion

Random Forest modeling was performed on a dataset consisting of 1380 compounds with known experimental COX-2 activity and 184 molecular descriptors. The RF model possesses excellent predictive performance scores, i.e., 93% accuracy, 98% sensitivity, 83% specificity, and 0.98 AUC.

The quantitative relationship, i.e., RF model, between the relevant structural features and compound classification whether active or inactive against the COX-2 enzyme was applied to two sets of compounds with no known COX-2 inhibitory activities: the Derivatives, variants of the most active member of the 5 largest families of known COX-2 inhibitors designed by isosteric elaboration; and the Similars, a collection of structurally similar compounds from ChEMBL and ZINC databases that were obtained through SwissSimilarity platform. The model classified 759 Derivatives and 188 Similars as active compounds against the COX-2 enzyme.

The top hits composed of 13 Derivatives and 6 Similars have outstanding drug-likeness and toxicity profiles and are relatively easy to synthesize. These 19 compounds are rational choices for further drug development to produce new COX-2-acting medicinal agents. The 13 top Derivatives and one top Similar have comparable or even superior binding affinities with the drug target compared to the control ligands. The derivatives D256 and D251 are the top two choices having the highest probability of being active and the strongest binding affinity with COX-2. The RF model established in this work is

conservative in identifying compounds as active, thus, is beneficial in avoiding costly failures at the later stages of the drug discovery phase.

Acknowledgments: This work was fully supported by the Office of the Vice President of Academic Affairs, the University of the Philippines System through the Faculty Reps Administrative Staff Development Program (FRASDP).

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Global Industry Analysts, Inc. Global pain management market to reach US$60 billion by 2015. According to a new report by Global Industry Analysts, Inc. 2011 [cited 2022 Nov 1]. Available from: https://www.prweb.com/releases/2011/1/prweb8052240.htm
2. Ferrero-Miliani L, Nielsen OH, Andersen PS, Girardin SE. Chronic inflammation: importance of NOD2 and NALP3 in interleukin-1beta generation. Clin Exp Immunol. 2007;147(2):227-35. doi:10.1111/j.1365-2249.2006.03261.x
3. Mallbris L, Akre O, Granath F, Yin L, Lindelöf B, Ekbom A, et al. Increased risk for cardiovascular mortality in psoriasis inpatients but not in outpatients. Eur J Epidemiol. 2004;19(3):225-30. doi:10.1023/b:ejep.0000020447.59150.f9
4. Kolb H, Mandrup-Poulsen T. The global diabetes epidemic as a consequence of lifestyle-induced low-grade inflammation. Diabetologia. 2010;53(1):10-20. doi:10.1007/s00125-009-1573-7
5. Miller AH, Maletic V, Raison CL. Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression. Biol Psychiatry. 2009;65(9):732-41. doi:10.1016/j.biopsych.2008.11.029
6. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell. 2010;140(6):883-99. doi:10.1016/j.cell.2010.01.025
7. National Fibromyalgia & Chronic Pain Association. Pain facts: an overview of American Pain Surveys, 2015 [cited 2019 July]. Available from: http://chronicpainaware.org/pain-101/pain-survey-results
8. American Academy of Pain Association. Facts and figures on pain, 2016 [cited 2019 July]. Available from: http://www.painmed.org/files/facts-and-figures-on-pain.pdf
9. Litalien C, Beaulieu P. Molecular mechanisms of drug actions: from receptors to effectors. In Fuhrman BP, Zimmerman JJ (eds.). Pediatric critical care (4th Ed.). Philadelphia, PA: Elsevier Saunders; 2011. pp. 1553-68.
10. Soboleva MS, Loskutova EE, Kosova IV, Amelina IV. Problems and the Prospects of Pharmaceutical Consultation in the Drugstores. Arch Pharm Pract. 2020;11(2):154-9.
11. Vo TH, Dang TN, Nguyen TT, Nguyen DT. An Educational Intervention to Improve Adverse Drug Reaction Reporting: An Observational Study in a Tertiary Hospital in Vietnam. Arch Pharma Pract. 2020;11(3):32-7.
12. Nakagawa N. Comparative study between formative assessment and flipped classroom lectures in a drug information course. J Adv Pharm Educ Res. 2021;11(2):5-10.
13. Fu JY, Masferrer JL, Siebert K, Raz A, Needleman PJ. The induction of prostaglandin-H2 synthase (cyclooxygenase) in human monocytes. J Biol Chem. 1990;265(28):16737-40.
14. Gierse JK, Hauser SD, Creely DP, Koboldt C, Rangwala SH, Isakson PC, et al. Expression and selective inhibition of the constitutive and inducible forms of human cyclo-oxygenase. Biochem J. 1995;305(Pt 2)(Pt 2):479-84. doi:10.1042/bj3050479
15. Chandrasekharan NV, Dai H, Roos KL, Evanson NK, Tomsik J, Elton TS, et al. COX-3, a cyclooxygenase-1 variant inhibited by acetaminophen and other analgesic/antipyretic drugs: cloning, structure, and expression. Proc Natl Acad Sci U S A. 2002;99(21):13926-31. doi:10.1073/pnas.162468699
16. Laine L, Takeuchi K, Tarnawski A. Gastric mucosal defense and cytoprotection: bench to bedside. Gastroenterology. 2008;135(1):41-60. doi:10.1053/j.gastro.2008.05.030
17. Kurumbail RG, Kiefer JR, Marnett LJ. Cyclooxygenase enzymes: catalysis and inhibition. Curr Opin Struct Biol. 2001;11(6):752-60. doi:10.1016/s0959-440x(01)00277-9
18. Penning TD, Talley JJ, Bertenshaw SR, Carter JS, Collins PW, Docter S, et al. Synthesis and biological evaluation of the 1,5-diarylpyrazole class of cyclooxygenase-2 inhibitors: identification of 4-[5-(4-methylphenyl)-3-(trifluoromethyl)-1H-pyrazol-1-yl]benze nesulfonamide (SC-58635, celecoxib). J Med Chem. 1997;40(9):1347-65.
19. Riendeau D, Percival MD, Brideau C, Charleson S, Dubé D, Ethier D, et al. Etoricoxib (MK-0663): preclinical profile and comparison with other agents that selectively inhibit cyclooxygenase-2. J Pharmacol Exp Ther. 2001;296(2):558-66.
20. Talley JJ, Bertenshaw SR, Brown DL, Carter JS, Graneto MJ, Kellogg MS, et al. N-[[(5-methyl-3-phenylisoxazol-4-yl)-phenyl]sulfonyl]propanamide, sodium salt, parecoxib sodium: A potent and selective inhibitor of COX-2 for parenteral administration. J Med Chem. 2000;43(9):1661-3.

21. Bally M, Dendukuri N, Rich B, Nadeau L, Helin-Salmivaara A, Garbe E, et al. Risk of acute myocardial infarction with NSAIDs in real world use: bayesian meta-analysis of individual patient data. BMJ. 2017;357:j1909. doi:10.1136/bmj.j1909
22. Lanas A, Chan FKL. Peptic ulcer disease. Lancet. 2017;390(10094):613-24. doi:10.1016/S0140-6736(16)32404-7
23. Breiman L. Random forests. In: Machine learning, Kluwer Academic Publishers, The Netherlands. 2001;45(1):5-32.
24. Boulesteix AL, Janitza S, Kruppa J, König IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdiscip Rev Data Min Knowl Discov. 2012;2(6):493-507.
25. Hsueh HM, Zhou DW, Tsai CA. Random forests-based differential analysis of gene sets for gene expression data. Gene. 2013;518(1):179-86. doi:10.1016/j.gene.2012.11.034
26. Lind AP, Anderson PC. Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. PLoS One. 2019;14(7):e0219774. doi:10.1371/journal.pone.0219774
27. Tetschke F, Schneider U, Schleussner E, Witte OW, Hoyer D. Assessment of fetal maturation age by heart rate variability measures using random forest methodology. Comput Biol Med. 2016;70:157-62. doi:10.1016/j.compbiomed.2016.01.020
28. Mayo SL, Olafson BD, Goddard WA. Dreiding: a generic force field for molecular simulations. J Phys Chem. 1990;94(26):8897-909.
29. Macalino SJY. Molecular dynamics simulation of human COX-2. Unpublished work, 2021.
30. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-61. doi:10.1002/jcc.21334
31. Billones LT, Gonzaga AC. Multiple Logistic Regression Modeling of Compound Class (Active/Inactive) and Prediction on Designed Coxib Derivatives and Compounds Similar to Known COX-2 Inhibitors. Chem-Bio Inform J. 2022;22:63-87. doi:10.1273/cbij.22.63
32. Kullback S, Leibler RA. On information and sufficiency. Annals Math Stat. 1951;22(1):79-86.
33. Billones LT, Morales NB, Billones JB. Logistic regression and random forest unveil key molecular descriptors of druglikeness. Chem-Bio Inform J. 2021;21:39-58.
34. Gini C. On the measure of concentration with special reference to income and statistics. Colorado College Publication, General Series. 1936;208(1):73-9.
35. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1(1):81-106.
36. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. Nat Chem. 2012;4(2):90-8. doi:10.1038/nchem.1243
37. Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J Cheminform. 2009;1(1):8. doi:10.1186/1758-2946-1-8
38. Orlando BJ, Malkowski MG. Substrate-selective Inhibition of Cyclooxygeanse-2 by Fenamic Acid Derivatives Is Dependent on Peroxide Tone. J Biol Chem. 2016;291(29):15069-81. doi:10.1074/jbc.M116.725713