**Pharmaceutical Sciences and Drug Design**
**ISSN: 3062-4428**
2022, Volume 2, Page No: 61-70

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

Galaxy Publication

# A Random Forest Approach to Analyzing Molecular Descriptors of COX-2-Selective 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.

***E-mail** ✉ ltbillones@up.edu.ph

**Received:** 24 August 2022; **Revised:** 28 November 2022; **Accepted:** 29 November 2022

## ABSTRACT

The pursuit of next-generation non-steroidal anti-inflammatory drugs (NSAIDs) remains a critical focus in pharmaceutical research, given that over a billion individuals experience pain and inflammation. A key strategy in this effort involves developing a quantitative correlation between the anti-inflammatory potential and the molecular descriptors of cyclooxygenase-2 (COX-2) inhibitors, which will facilitate the identification and advancement of novel NSAIDs that minimize adverse effects associated with COX-1 inhibition. In this study, the random forest (RF) algorithm was used to construct a highly predictive quantitative model to assess the inhibitory activity of various compounds targeting COX-2. The resulting model demonstrated an outstanding classification accuracy of 93% with an AUC of 0.98. When applied to external datasets, it identified 759 newly designed COX-2 inhibitor derivatives and 188 structurally related compounds as active, with 19 emerging as strong candidates for COX-2-targeted anti-inflammatory agents. Among these compounds, the top two compounds showed the highest probability of activity and exhibited superior binding affinity to COX-2 compared to existing selective inhibitors. Furthermore, the RF model proved to be conservative in predicting active compounds, reducing the risk of late-stage failures in drug discovery and increasing the efficiency of the development process.

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

**How to Cite This Article:** Billones LT, Gonzaga AC. A Random Forest Approach to Analyzing Molecular Descriptors of COX-2-Selective Non-Steroidal Anti-Inflammatory Drugs (NSAIDs). Pharm Sci Drug Des. 2022;2:61-70. https://doi.org/10.51847/qi0KURgQFv

## Introduction

Inflammation is a major health concern worldwide, affecting over 1.5 billion people [1]. It manifests as pain, redness, swelling, heat, and functional impairment [2]. This biological response is linked to the development of chronic conditions such as cardiovascular disease, diabetes, autoimmune disorders, cancer, and respiratory illnesses [3-6], all of which significantly impact patients' well-being [7, 8].

A key mechanism of inflammation involves arachidonic acid metabolism, regulated by cyclooxygenase (COX) enzymes, particularly COX-1 and COX-2 [9-12]. Although these enzymes share structural similarities, a key distinction exists at position 523, where COX-1 contains isoleucine, whereas COX-2 has valine [13]. The bulkier isoleucine in COX-1 restricts access to its active site, whereas COX-2, due to its valine substitution, allows the binding of larger molecules.

COX-1 is constitutively expressed [14] and plays an essential role in physiological functions, such as maintaining gastric mucosal integrity by promoting protective prostaglandins [15, 16]. Inhibition of COX-1 can therefore lead to gastrointestinal complications, including ulceration. On the other hand, COX-2 is an inducible enzyme [14] predominantly upregulated in inflamed tissues. Targeting COX-2 selectively can reduce gastrointestinal risks associated with traditional COX-1 inhibition [17].

Conventional NSAIDs inhibit both COX-1 and COX-2, whereas selective COX-2 inhibitors, such as coxibs, were developed to minimize adverse effects [18-20]. However, current anti-inflammatory drugs still present challenges due to their side effects, necessitating the ongoing search for safer therapeutic alternatives [21, 22].

One approach in drug discovery involves identifying new bioactive compounds based on the structural properties of known inhibitors. A key technique in this process is machine learning, with random forest (RF) being a widely used classification method [23]. This algorithm constructs multiple decision trees to categorize compounds as active or inactive, making it valuable for analyzing biological [24, 25] and medical [26, 27] datasets. In this study, RF modeling was applied to two external datasets: one consisting of newly designed compounds (derivatives) and another containing structurally similar molecules (similars) with potential COX-2 inhibitory activity.

## Materials and Methods

To gather experimental data on COX-2 inhibition, literature searches spanning 1997 to 2019 were conducted using keywords related to COX inhibition. Compounds were classified based on IC50 values, with active compounds defined as those having IC50 ≤ 10 μM, while those with IC50 > 10 μM were considered inactive.

Molecular structure preparation and computational analyses were performed on a Windows 7 Professional 64-bit system with an Intel® Core™ i7-4770K processor (3.50 GHz) and 8 GB RAM, along with a macOS Catalina machine featuring a 3.1-GHz Dual-Core Intel Core i7 processor and 16 GB RAM. Chemical structures were drawn using ChemDraw Professional 16.0 (www.perkinelmerinformatics.com) and saved in .sdf format. The transition from 2D to 3D structures was executed in Discovery Studio (DS) 4.0 (Biovia, Inc.), utilizing the Dreiding force field for structural optimization [28]. Molecular descriptors were calculated using Spartan 16 (www.wavefun.com) and DS 4.0.

Machine learning analysis was conducted with RapidMiner Studio 9.7.001 (www.rapidminer.com). The dataset was split, with 20% (276 compounds) reserved for testing and the remainder for model training. A 2-fold validation method was implemented, where 80% of the training set was used to build the model, and 20% for hyperparameter tuning. The optimized RF model was then trained on the full training dataset (1,104 compounds). Model performance was assessed using accuracy, sensitivity, specificity, and area under the curve (AUC).

The validated RF model was subsequently applied to external datasets: (1) derivatives, which included novel compounds derived from the most abundant chemical families, and (2) similar, identified from the ChEMBL and ZINC databases through similarity searches using SwissSimilarity (www.swisssimilarity.ch), with the most active inhibitors serving as reference structures. Molecular properties were calculated before classification using the RF model.

Predicted active compounds underwent in silico ADMET analysis using DS modules ADMET and TOPKAT. Drug-likeness was assessed through a quantitative estimate of drug-likeness (QED) scores, while synthetic accessibility (SA) was evaluated via SwissADME (http://www.swissadme.ch).

Molecular geometry optimization of the most promising candidates was performed using PM3 semi-empirical methods, with equilibrium conformers serving as starting structures. The refined structures were saved as pdb files. Furthermore, a 100-ns molecular dynamics simulation [29] was performed on the COX-2 target protein (PDB ID: 5IKR), followed by molecular docking studies using AutoDock Vina [30] in PyRx (www.pyrx.sourceforge.io).

## Results and Discussion

The selection of journal articles for the compound collection was primarily based on the methodological consistency in measuring experimental COX-2 activity [14]. A total of 66 reports from six leading scientific journals contributed to the dataset, as detailed in **Table 1** [31]. From these sources, 59 distinct chemical families were identified, comprising 1,380 compounds in total. Among them, 929 (67%) demonstrated COX-2 activity, while the remaining 451 (33%) were classified as inactive.

For each compound, over 400 molecular descriptors were computed. The Discovery Studio suite generated 397 descriptors, including 333 based on 2D properties and 64 derived from 3D structural features. Additionally, Spartan 16 provided 28 descriptors, distributed across molecular (9), QSAR (14), and thermodynamic (5) categories. To refine the dataset, descriptors with a high frequency of missing values (NaN) or minimal variability were eliminated, reducing the total number of variables to 184.

**Table 1.** Compounds by family and by experimental COX-2 inhibitory activity, from literature, published 1997-2019

| No. | Family | Actives | Inactive | 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

A sequence of trials utilizing 80% of the training dataset, followed by validation on the remaining 20%, revealed that the random forest model attained its highest levels of specificity, accuracy, and sensitivity when the correlation threshold was set at r = 0.75, as illustrated in **Figure 1**. As a result, only variables with correlation coefficients of 0.75 or lower were incorporated into the model, refining the initial 184 descriptors down to 64. The significance of these selected descriptors is depicted in **Figure 2**, arranged in descending order of their influence. Some of the most critical descriptors included molecular weight (wt1), shadow_z-length (sz), which represents molecular shadow extension along the z-axis, the frontier molecular orbitals eho and elu (EHOMO, ELUMO), AM1 energy (am1), polar surface area (psa), and dipole moment (dip).

**Figure 3** identifies information gain [32] as the most effective criterion for determining node splits during tree formation. The model achieved its optimal classification performance at a tree depth of 14, where classification errors were minimized.

In decision tree methodologies, node-splitting criteria such as information gain, gain ratio, and Gini index are used to determine the most suitable variable for dividing a node. Information gain measures the reduction in entropy, which signifies the level of impurity within nodes, with lower entropy values indicating nodes that are more homogeneous and thus enhance classification accuracy [33].

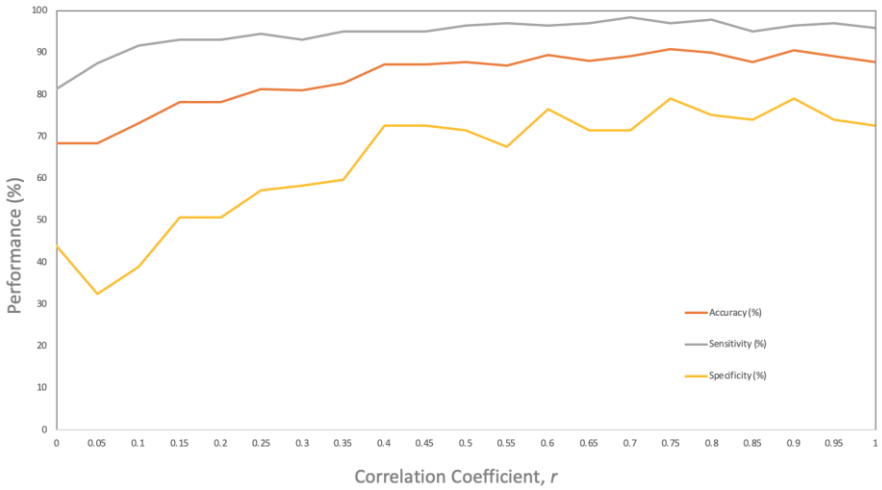


**Figure 1.** The random forest model specificity, sensitivity, and accuracy as determined by the maximum correlation coefficient (r) allowed among the independent variables, using ntree = 100 and maximal depth = 15, with information gain as the splitting criterion.

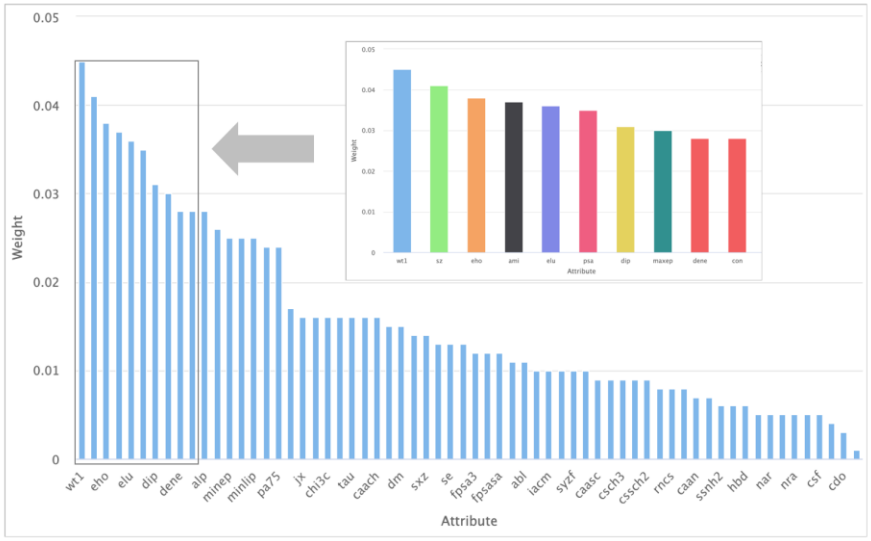


**Figure 2.** The descriptors in the random forest model by their importance in generating the compound class prediction.

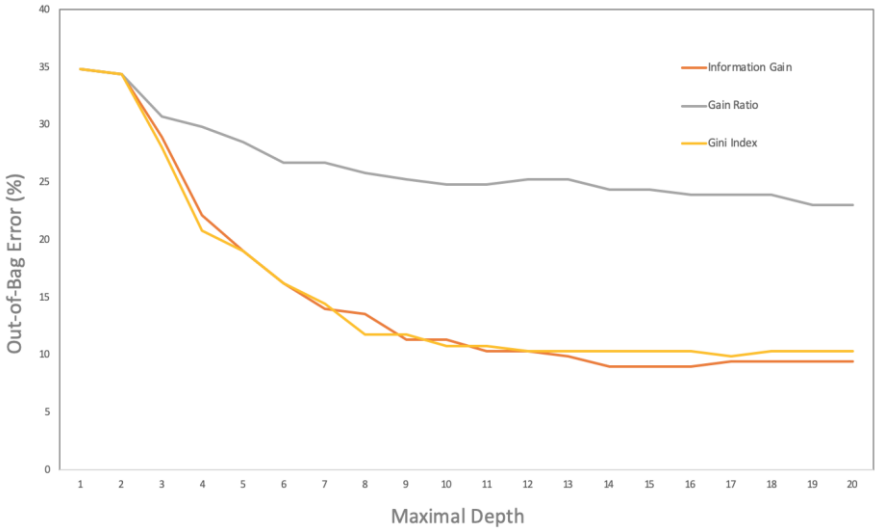


**Figure 3.** The classification error (%) for the different splitting criteria (information gain, gain ratio, and Gini index) is determined by the maximal depth.

Using the Gini index [34] as the splitting criterion, the classification error reached its lowest point at a tree depth of 17, which extended the branching structure by three levels compared to information gain. The Gini index evaluates the degree of value distribution within a node, where smaller values indicate reduced entropy, resulting in purer nodes. On the other hand, when the Gain ratio was applied [35], classification error continued to fluctuate even at a depth of 20, preventing stabilization.

Based on these model refinement analyses, the final model was trained on the complete 80% training dataset, incorporating descriptors that met the condition $r \leq 0.75$. Information Gain was employed as the node-splitting criterion, with the tree depth capped at 14 across all 100 decision trees within the random forest ensemble. Encouragingly, the model demonstrated exceptional predictive accuracy, attaining an overall performance of 93%. It successfully identified 182 out of 186 active compounds (98% sensitivity) and correctly classified 75 out of 90 inactive compounds (83% specificity), as depicted in **Figure 4**.

An additional performance evaluation metric, the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, provides insight into classification effectiveness. A perfectly functioning model achieves an AUC of 1. As illustrated in **Figure 5**, the AUC-ROC for the random forest model reached 0.98, signifying near-perfect discrimination between active and inactive compounds.

Furthermore, a set of compounds labeled as derivatives was systematically designed by introducing modifications to the most active molecule within each of the five selected COX-2 inhibitor classes. This led to the virtual creation of 1,100 compounds derived from the core structural frameworks of cyclopentenes, imidazolyls, difluorobenzenes, furanyl/thiophenyls, and isoxazoles. Additionally, another set, termed similars, consisted of 600 compounds sourced from the ChEMBL bioactive and ZINC drug-like databases using the SwissSimilarity platform. The similarity search queries were based on the most active representative compound from each known COX-2 inhibitor family.


**Figure 4.** The random forest model class prediction of the test set of compounds.

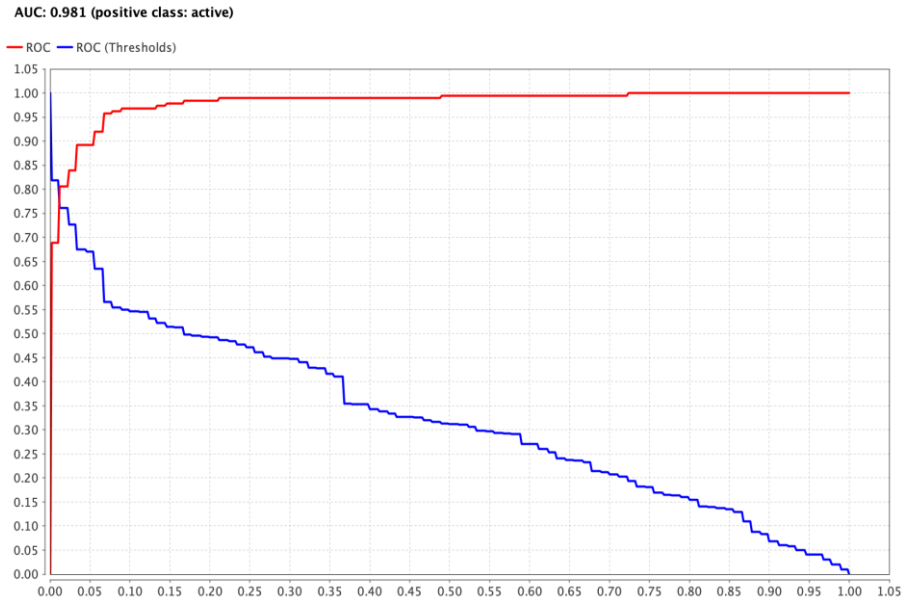


**Figure 5.** The random forest model receiver operator characteristic (ROC) plot.

When the RF model was applied to the set of derivatives, 69% (759 out of 1100) of the compounds were predicted to exhibit COX-2 activity. Among these, cyclopentene derivatives (compounds 1–300) showed the highest activity, whereas difluorobenzenes (compounds 501–700) were predominantly inactive. It's worth noting that the RF model tends to predict a higher rate of inactivity compared to the multiple logistic regression (MLogR) model [31], which could be advantageous in the early stages of drug discovery by allowing for the exclusion of compounds that are likely to be inactive.

For the Similars, the RF model predicted 31% (188 out of 600) to be active against COX-2. The average similarity score of these active compounds was 0.38, indicating considerable structural differences from the query structures used in the search. This accounts for the relatively small number of active compounds in the similar group.
These compounds present the potential for developing new scaffolds for COX-2 selective NSAIDs, and the large proportion of inactive similars helps reduce attrition rates and lower the overall cost of drug discovery.
Further in silico analysis of the predicted active compounds was conducted to assess their drug-likeness and synthetic accessibility. Over 93% of the active derivatives scored above 0.5 on the quantitative estimate of drug-likeness (QED) scale [36], indicating they are likely to be drug-like. Their average synthetic accessibility (SA) score was 3.3, within the acceptable range of 1–6 [37], and they showed low to optimal solubility and moderate to good intestinal absorption. Most were non-mutagenic and non-carcinogenic (89%), although all were hepatotoxic. Likewise, the active similars also had a wide range of synthetic accessibility scores (2.1 to 5.6) and 55% were considered drug-like. These compounds also exhibited favorable characteristics such as non-mutagenicity (84%) and non-carcinogenicity (81%), although only 23% showed good intestinal absorption.
The top hits from these groups were selected based on specific criteria, including a PA greater than 0.7, QED above 0.5, and other parameters such as acceptable synthetic accessibility (SAS), solubility (AS), and non-toxicity. Only 13 Derivatives and 6 Similars met these stringent requirements, as shown in **Figure 6**.

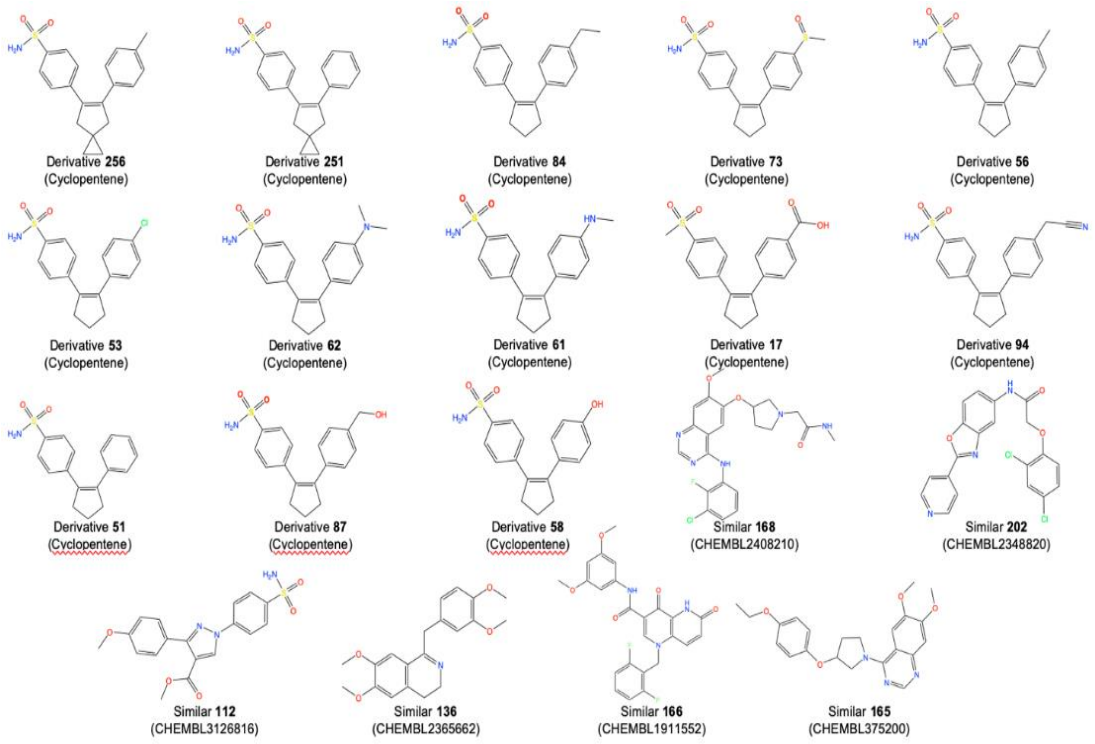


**Figure 6.** The molecular structures of the top hit from the derivatives and similar.

Among the top thirteen derivatives identified, 10 compounds were also highlighted in related MLogR studies [31]. These compounds mainly belong to the cyclopentane derivatives category, but the two highest-ranking hits, D256 and D251, are diarylspiroheptenes. Other prominent compounds on the list include D84, D61, and D87, all derivatives of cyclopentenes. In contrast, the top hits from the Similars group are distinct and differ from those found in MLogR studies [31].
The selection criteria for these hits significantly reduced the false positive rate by focusing only on compounds with a high probability of activity (PA > 0.7). Additionally, selecting compounds with QED scores above 0.5 further ensured they were closer to drug-like characteristics [36]. These compounds were deemed easy to synthesize, with an SAS range of 2-4, and showed favorable properties such as good solubility, efficient intestinal absorption, and no evidence of carcinogenic or mutagenic effects. They are also suitable for use with other drugs due to their non-inhibition of CYP2D6. Molecular docking studies demonstrated promising results as well, with all the top 13 derivatives and one of the similar (S202) showing stronger binding energies than etoricoxib (BE =

–7.8 kcal/mol), a recognized COX-2 selective drug, and comparable to or exceeding the binding energy of mefenamic acid (BE = –8.6 kcal/mol) [38].
D256 and D251, the top two hits, exhibit the strongest binding to COX-2 among the identified compounds, which increases their potential for progressing to later stages of drug development.

## Conclusion

Random forest (RF) modeling was employed on a dataset comprising 1380 compounds with known COX-2 activity, alongside 184 molecular descriptors. The RF model achieved excellent predictive performance with 93% accuracy, 98% sensitivity, 83% specificity, and an AUC of 0.98.
The model was applied to two sets of compounds with no previous COX-2 activity: the derivatives, designed by modifying the most active compound from five major COX-2 inhibitor families, and the similar, which were selected from ChEMBL and ZINC databases via the SwissSimilarity platform. The RF model identified 759 Derivatives and 188 Similars as active against COX-2.
The identified top compounds, including thirteen derivatives and six similars, demonstrated favorable drug-likeness properties, manageable toxicity profiles, and ease of synthesis. These 19 compounds represent strong candidates for further drug development as COX-2 inhibitors. Moreover, these compounds showed binding affinities comparable to or even stronger than the reference drugs. D256 and D251 were the top two compounds with the highest probability of being active and exhibited the most potent binding to COX-2. The conservative nature of the RF model in classifying compounds as active minimizes the risk of costly failures in the later stages of the drug discovery process.

**Acknowledgments:** This research was fully supported by the Office of the Vice President for Academic Affairs at the University of the Philippines, 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. p. 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. Gen Ser. 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