2024 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) | 979-8-3503-5663-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/CIBCB58642.2024.10702169

Artificial Neural Networks for the Rapid Prediction of Possible Ferroptosis Inducers Using the GPx4 Enzyme

Jesus Antonio Alvarado-Huayhuaz
Facultad de Ciencias
Universidad Nacional de Ingeniería
Lima, Peru
jalvaradoh@uni.pe

Miquéias Amorim Santos Silva
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
kekoamorimsilva@gmail.com

Elvis Mauricio Jimenez Peña
Departamento de Fisicoquímica
Universidad Nacional Mayor de San Marcos
Lima, Peru
elvis.jimenez@unmsm.edu.pe

Joel Claudio Rengifo Maraví
Departamento de Química
Universidad Nacional Agraria La Molina
Lima, Peru
jcrengifo@lamolina.edu.pe

Angella M. Cordova-Muñoz
Departamento de Farmacología
Universidad Nacional de Trujillo
Trujillo, Peru
acordovam@unitru.edu.pe

Karina dos Santos Machado
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
karina.machado@furg.br

***Abstract*—Cancer affects 20 million people worldwide; therefore, accelerated and urgent pathways are needed for the search for new drugs. The reactivation of cell death induced by the accumulation of iron-dependent lipid peroxides (Ferroptosis) is a potential mechanism for cancer treatment. One pathway to trigger Ferroptosis involves the inactivation of the glutathione-dependent peroxidase (GPx4) protein. Molecules with high affinity for iron are also being studied against cancer through chelation therapy. Therefore, the rapid detection of such molecules in molecular databases is of interest. Catechols exhibit a high iron-chelating power and can be synthesized by microorganisms as a chemical strategy for acquiring Fe^{3+}. They can also be found in natural products. Recently, the natural product database in Peru (PeruNPDB) has been published where we found catechols; however, these have not yet been studied in GPx4. To expedite the search for catechols with affinity for the active site of GPx4, in large databases, we have developed a prediction model for scoring based on artificial neural networks. Our initial database consisted of the 83 catechols and 1,024 molecular descriptors. The target is the docking score obtained by ligand-receptor GPx4 coupling with the Autodock Vina software. Emphasizing the application of recursive feature elimination to improve the model's performance and provides both the model and the procedure to be applied and improved by the scientific community. Our program enables the experimenter to rapidly filter molecules with high affinity specifically for the active site of GPx4, in the search for potential inhibitors.**

***Index Terms*—natural products, catechol, GPx4, molecular docking, artificial neural networks**

I. Introduction

The fight against cancer represents one of the most significant challenges in the field of global health, given the devastating impact this disease has on the lives of millions of people [19]. With its increasing prevalence and the urgent need to discover effective treatments, the search for new drugs has become a research priority. In this context, chelation therapy has emerged as a promising avenue of investigation. This therapeutic approach, involving the removal of toxic metals from the body, has garnered the scientific community's attention due to its potential in cancer treatment [12].

A particularly intriguing area of research within chelation therapy is its potential connection with a cellular process known as ferroptosis. The identification of ferroptosis, a type of programmed cell death that relies on iron and is marked by the peroxidation of lipids, has introduced new opportunities in the treatment of cancer [8]

A key regulator of ferroptosis is the protein glutathione peroxidase 4 (GPx4), and inhibiting its function has emerged as a promising approach to therapy. Nonetheless, the intricate mechanisms governing ferroptosis and the presence of pathways independent of GPx4 highlight the need for a more comprehensive understanding and the development of new therapeutic compounds [13].

Siderophores are low molecular weight organic molecules, biosynthesized by microorganisms as a strategy for iron acquisition, owing to their high chelating power for Fe^{3+} [16]. In recent years, drugs targeting cancer through chelation therapy have been under development. Therefore, the search for siderophores for this purpose is crucial [20] because these molecules form coordination compounds (iron-siderophore) with high stability values, as for example, in the case of catechols synthesized by *Enterobactin, Bacillibactin*, among others [15].

Recently, a database of natural products in Peru (PeruNPDB) has been published [1], serving as a valuable resource for research in the field of health sciences. Within this database, analogs of siderophores were identified, such as molecules containing the 1,2-benzenediol group (catechol), for instance. However, the intermolecular interactions of these

979-8-3503-5663-2/24/$31.00 ©2024 IEEE

molecules with the GPx4 receptor have not yet been investigated to assess their inhibitory potential. Several studies using molecular docking have already identified other small molecules (which do not contain catechols) with potential inhibitory effects on GPx4 [4]. This opportunity to investigate the activity of catechols in the context of chelation therapy and ferroptosis applying molecular docking simulations and machine learning presents an exciting and promising avenue for scientific research [18].

Thus, for accelerating the identification of compounds with therapeutic potential against cancer, we have developed a target-specific scoring function based on artificial neural networks ANN. Additionally, we have applied feature selection with recursive feature elimination (RFE) algorithm, to enhance the accuracy and utility of our model. We provide both the model and the procedures used for its development in our GitHub repository [1], hoping that this research can contribute to the advancement of science and the development of new cancer treatments.

II. Methodology

In this work we are proposing a target-specific scoring function (SF) considering as ligands the molecules containing the catechol functional group and as receptor the GPx4 protein. It is important to highlight that the objective is not to generate a scoring function for any organic ligand, but specifically for those containing the catechol chelating group. This focus ensures that we only cover molecules capable of coordinating the ferric ion, as previously explained in the introduction. Figure 1 describes our proposed methodology pipeline for developing a GPx4-specific SF.

First, from the PeruNPDB database[2] [1], we obtained 83 molecules containing the catechol functional group. Using the available SMILES code [21], we generated three-dimensional coordinates and pre-optimized them using the gradient descent algorithm and the MMFF94 force field to obtain structures in PDBQT format with RDKit 2020 [10].

We prepared the GPx4 protein (PDB ID: 2OBI) using the MGLTools 1.5.6 and generated this protein molecular structure in PDBQT format. Molecular docking was performed using the Autodock Vina 1.2.3 [3], and a simulation box with size of 20Å edge was configured, centered at the centroid $(39.368, -17.754, -18.972)$, with an exhaustiveness of 60. The obtained Free Energy of Binding (FEB) of Autodock Vina was considered as the target attribute in our input files for the SF model.

The SMILES codes were also used to obtain 1,024 molecular descriptors using the PaDel-Descriptor 2.21 program [22], which, together with the target attribute (estimated FEB), we constructed the input dataset for generating our proposed SF.

The next step consisted of preprocessing the input dataset to reduce its dimensionality. So, we performed an exploratory analysis after scaling all instances and attributes containing a large number of zeros were removed, resulting in 737 columns. Recursive Feature Elimination (RFE) is a feature selection technique that works by iteratively eliminating the least relevant attributes from a dataset until optimizing a model metric, such as minimizing RMSE. This technique is used to enhance the efficiency and accuracy of predictive models. So, we applied RFE in our input dataset and obtained a reduced dataset with 131 attributes. We conducted a study with and without the application of RFE to compare the effect of its usage.

After selecting 80% of the data for training, we applied 6 hidden layers, consisting of 360, 180, 60, 24, 12, 6 fully connected artificial neurons each. To evaluate the proposed methodology for generating a target-specific SF, we employed R-squared (R^2, Equation 1), Root Mean Square Error ($RMSE$, Equation 2), Mean Squared Error (MSE, Equation 3), and Mean Absolute Error (MAE, Equation 4) metrics:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

Finally, we calculated the R^2, $RMSE$, MSE and MAE for the training and test data where y_i are the observed values, $\hat{y}_i$ are the predicted values, $\bar{y}$ is the average of the observed values, and n is the total number of observations.

In our proposed methodology we have used the following R packages to perform preprocessing, recursive feature elimination (RFE), cross-validation, generate the proposed SF, and obtain the evaluation metrics [14]: *tidyverse*, *doParallel*, *caret*, *VIM*, *neuralnet* [6], and *modelr*.

III. Results

A. Database

The collection of molecules and the generated descriptors are available in our Github repository. From the 1,024 descriptors obtained with Padel, some boxplots are shown in Figure 2 to highlight the distribution of the data according to the characteristics of these chelators. Considering that the molecules in our database have in particular one or more catechol groups, it is expected to find a significant variability in the information for the non-chelating part. For example, in the case of *RotBFrac* descriptor, which indicates the percentage or proportion of rotatable bonds relative to the total bonds in the molecule, according to Figure 2, it shows great variability, including outliers. Since it is not possible to show all the boxplots, Figure 2 presents a sample displaying low, medium and high variability descriptors, with and without

[1]Github repository avaible on https://github.com/inefable12/ExpInSilico24

[2]PeruNPDB database available on https://perunpdb.com.pe/

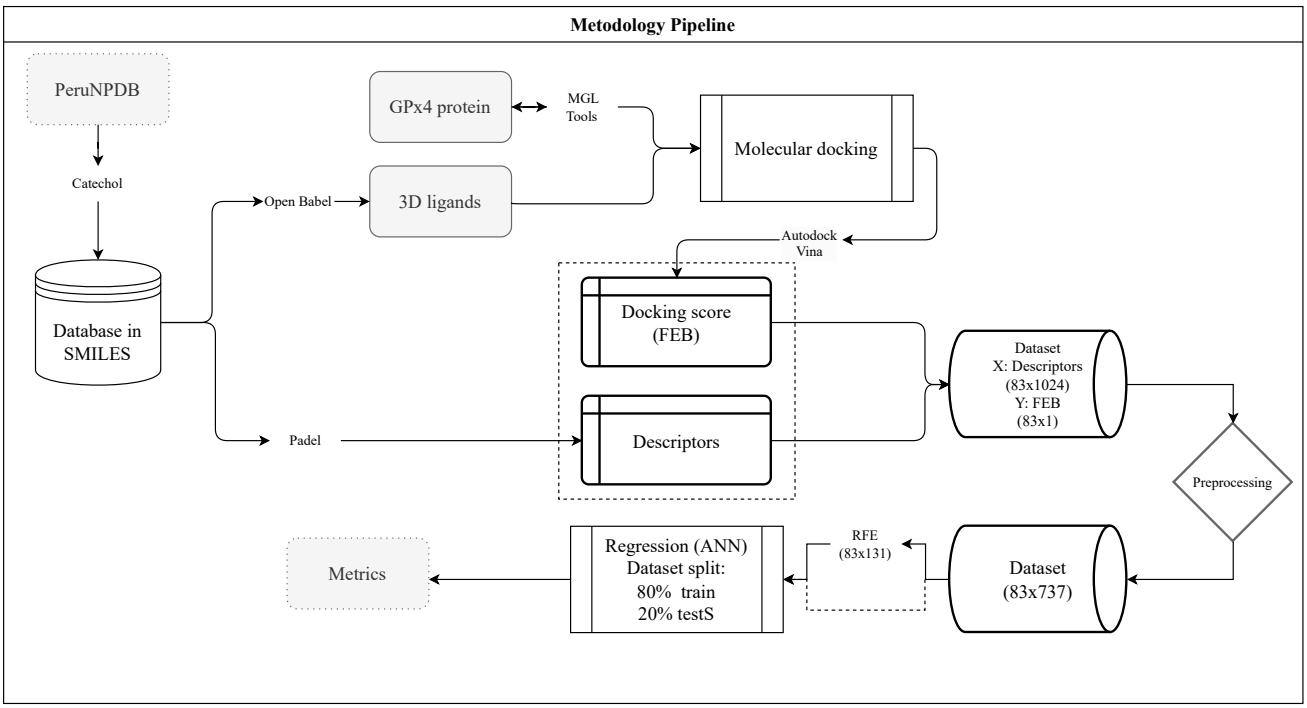

Fig. 1. Proposed methodology pipeline.

outliers (such as in *SpMAD_D*, *VE2_D*, among others), both symmetrical distribution (for example in the target attribute *FEB*) as asymmetric, all of these being considered in the input database.

B. Data preprocessing

After applying RFE (Figure 3), 131 attributes (molecular descriptors) were selected (Table 1).

C. Artificial neural network

The architecture of the ANN consisted of 131 descriptors as inputs, 6 hidden layers (360, 180, 60, 24, 12, 6 neurons, respectively) and 1 output. Other combinations of hidden layers and neurons were tested as well, however, good metrics were not obtained, likely due to the limited robustness of our model caused by the reduced amount of data (83 catechols from the PeruNPDB). In future work, we will consider new databases of natural products in our region to improve the robustness in our model. However, in this study, we focus on PeruNPDB because it is the first database of natural products in Peru and is newly established. Plotting the FEB data (real value) versus FEB predicted (predict value) by the model with artificial neural networks, we observe the linear trend of these values (Figure 4). The metrics R^2, $RMSE$, MSE and MAE obtained for the linear regression are detailed in Table II where the results for the SF model without feature selection are in the lines 1 and 2 and after applying RFE are in the lines 3 and 4.

TABLE I
DESCRIPTORS SELECTED WITH RFE.

Number	Descriptors
1-5	SHaaCH, SpMin3_Bhm, VE3_Dzi, SpMax2_Bhe, SaaCH,
6-9	SpMax2_Bhp, AATSC2s, MLFER_L, SpMin3_Bhs,
10-14	SpMax2_Bhi, VE2_Dt, SHother, SpDiam_Dt, SpMax2_Bhv,
15-20	GATS2s, hmin, JGI3, SpMax2_Bhm, VR2_Dt, SpMin3_Bhe,
21-25	nBondsM, ATSC2s, ATSC2c, SwHBa, ASP.0,
26-30	minHBa, ETA_Beta_ns, IC5, VR1_Dzm, ATS2v,
31-35	MLogP, SpMin2_Bhs, maxHBa, ATSC4m, ATSC4s,
36-40	AATSC7e, VE3_Dt, VR2_Dzi, ETA_Shape_P, SpMin2_Bhv,
41-45	VP.5, AATS6p, VE1_Dzi, minHother, MATS5e,
46-50	SpMax7_Bhm, VP.6, SpMax3_Bhp, GATS1s, VE3_Dzm,
51-55	VR1_Dze, ATS3v, SpMin2_Bhe, ATS2s, LFER_BO,
56-60	GATS7c, VR1_Dzi, AATSC2c, ATSC1m, MLFER_BH,
61-65	MATS4s, AATSC4c, MATS2s, VE3_Dze, ZMIC4,
66-70	AATSC4s, SaasC, AATSC8e, MAXDP, GATS1e,
71-75	AATS7i, AATSC8s, AATSC5c, AATS8i, EE_Dt,
76-80	AATSC5e, SpMax3_Bhi, ATSC3p, VE3_DzZ, MATS2p,
81-85	SpMAD_D, MATS5m, IC4, VE1_Dt, VR3_Dzp,
86-90	SpMax3_Bhv, ATSC5e, MATS2i, SpMAD_Dzs, MATS7e,
91-95	SpMin3_Bhi, VE2_Dzi, MIC2, AATSC6p, SpMax6_Bhi,
96-100	VR1_DzZ, VE2_DzZ, CIC4, SpMin2_Bhm, AATSC4e,
101-105	VE1_DzZ, VE3_D, MATS3i, ATSC5c, VR3_Dzv,
106-110	AATSC2i, MATS5c, SpMin2_Bhp, SpMin3_Bhp, MAXDP2,
111-115	ATSC7s, MATS5s, SpMin8_Bhs, ATSC3i, VE2_Dze,
116-120	GATS2m, AATSC5v, MIC5, SpMin2_Bhi, ATSC7e,
121-125	SpMax3_Bhm, ATSC8e, BCUTp.1l, AATSC2v, VE1_Dze,
126-131	ATS6p, GATS6p, VR3_Dzi, ATS1s, MATS2v, piPC2

IV. DISCUSSION

PeruNPDB is the first Peruvian natural product database reported in the literature that contains information about the SMILES code of molecules, as well as the reference articles, among other details. This information has been valuable in

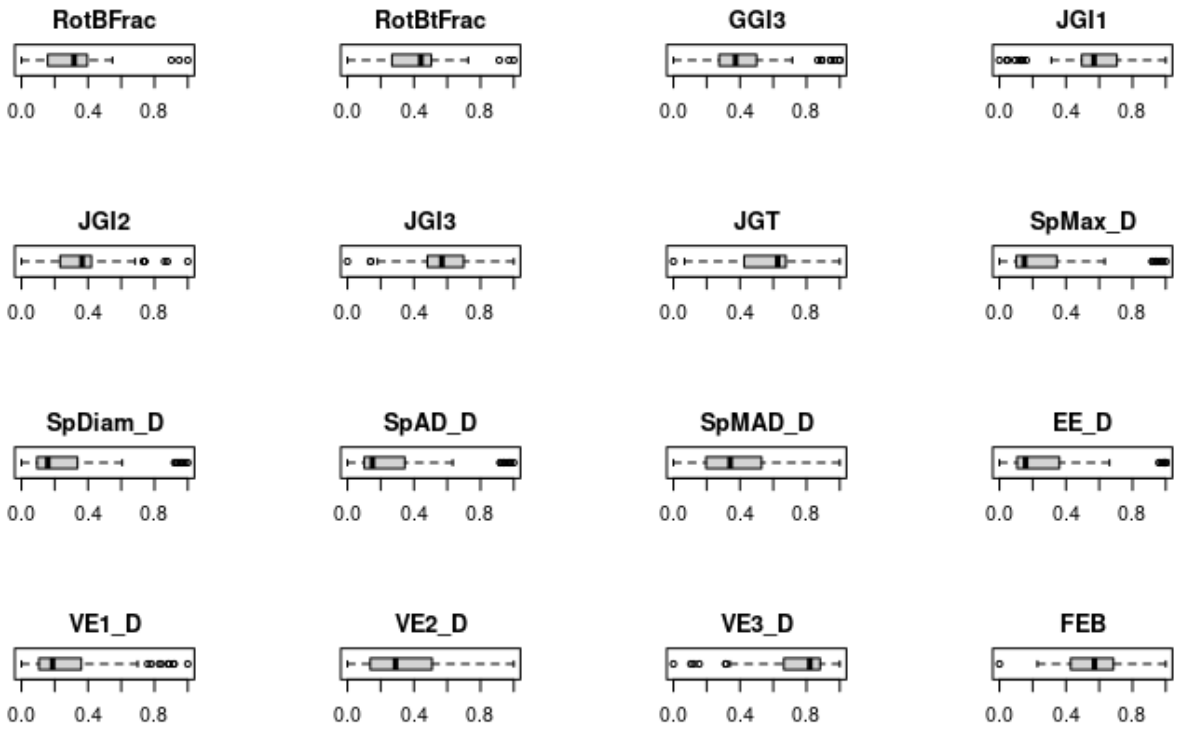

Fig. 2. Box plots of some attributes in our dataset

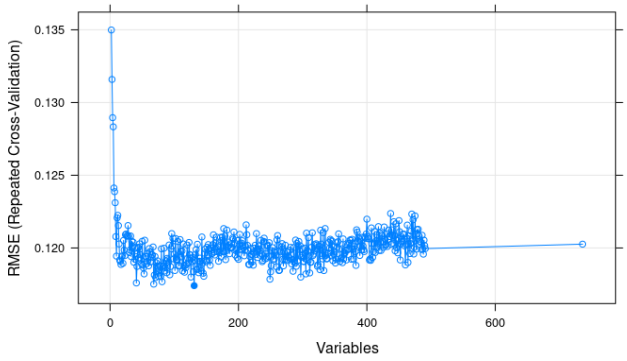

Fig. 3. RMSE and cross validation regarding the number of attributes

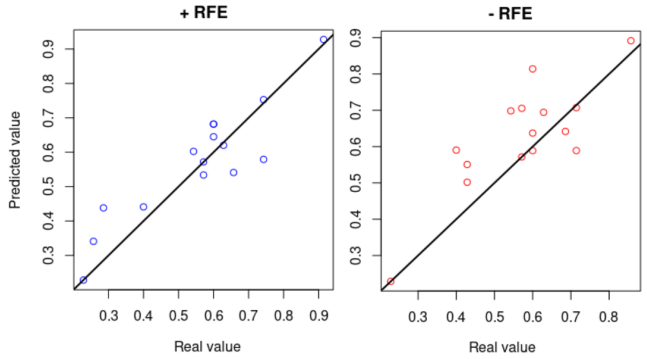

Fig. 4. Test data (X axis) versus the values predicted by the model with ANN (Y axis)

our search for functional groups with high affinity for iron, as it allows the use of SMARTS code for selecting functional groups using RDKit, with bacterial siderophores as a reference, mainly constituted by catechols, hydroxamic acids, and α-hydroxycarboxylates. Although PeruNPDB does not record hydroxamic groups, we found 83 molecules containing the catechol functional group; therefore, these were selected for the purpose of developing the machine learning target-specific SF model.

On the other hand, the interest in finding organic molecules with affinity for GPx4 lies in their potential application against cancer by inducing of ferroptosis through the inhibition of glutathione-dependent peroxidase (GPx4) [2, 4].

Molecules containing the catechol group can form complexes with Fe^{3+} with high thermodynamic stability favored by entropic effects (chelate effect) and electrostatics (Pearson's hard-soft acid-base theory) [16], which in the case of microorganisms, this promotes their recognition in iron-siderophore receptor proteins in the cell membrane and their potential internalization into the periplasm promoted by the Trojan horse effect [9].

After iron internalization, the ligand is separated from the

TABLE II
EVALUATION METRICS FOR THE SF GENERATED WITHOUT FEATURE SELECTION (LINES 1 AND 2) AND CONSIDERING RFE FEATURE SELECTION (LINES 3 AND 4).

	R2	RMSE	MSE	MAE
Train	0.9894	0.0186	0.0003	0.0042
Test	0.6292	0.1062	0.0112	0.0809
Train + RFE	0.9881	0.0189	0.0003	0.0057
Test + RFE	0.8267	0.0787	0.0061	0.0596

metal either by destabilizing the complex, due to the reduction of the metal to Fe^{2+} or by enzymatic degradation of the organic part of the complex. Thus, the presence of chelating groups in the intracellular medium can alter homeostasis through different pathways.

In cancer cells, a prominent characteristic is cellular proliferation, which is associated with a high demand for iron and the overexpression of iron receptors, such as lipocalins [7]. There is evidence of the promising anticancer activity of siderophores such as desferrioxamine (DFO), exochelin-MS, and mycobactin [5], a derivative of DFO with caffeine (DFCAF) [11], and also with catechol groups, such as enterobactin [17]. This evidence strongly justifies the need to find molecules with chelating groups such as catechol.

We registered an approximate 24 hours of computational calculation for obtaining the FEB with Autodock Vina (exhaustiveness: 60), among the 83 ligands and the GPx4 protein, from the SMILES representation available in PeruNPDB and the extraction plus pre-treatment of the protein from the PDB, respectively, versus approximately 1 minute used by the trained model for obtaining the FEB, also from the SMILES representation. We highlight this rapid FEB prediction as an efficient and fast tool for searching for organic molecules with affinity for GPX4 in large databases.

Although the model effectively fulfills its purpose with the training data, it presented overfitting as observed in the four metrics detailed in Table II. These were improved by previously applying the RFE with ANN, achieving an increase in the coefficient of determination by approximately 20% and reducing overfitting considerably. It is worth mentioning that the MSE increased by 0.0109 from training data to test data, while considering RFE, there is a smaller increase of 0.0058, which demonstrates an improvement in the model due to the reduction of MSE. The same trend is observed for MAE, with increases of 0.0767 without the application of RFE and 0.0539 considering RFE for feature selection.

The entire database used and the model are available in our GitHub repository, with the aim of opening it to the scientific community for its application in databases or for model improvement.

V. CONCLUSIONS

Cancer represents one of the gravest challenges in global health, impacting millions worldwide. The increasing incidence and the imperative for effective treatments have prioritized the search for new therapeutic agents. In this context, chelation therapy, particularly in relation to ferroptosis and the inhibition of Glutathione peroxidase 4 (GPx4), has emerged as a significant area of investigation. The identification of compounds within the Peru Natural Products Database (PeruNPDB) that have not been previously associated with GPx4 activity, specifically the catechols, presents a novel avenue for research.

To achieve the objectives of this research, a comprehensive methodology was employed. A predictive model was developed using artificial neural networks (ANN) to ascertain the docking scores between catechol-containing molecules and the GPx4 receptor. Data from the PeruNPDB was utilized, with molecular descriptors generated and optimized for model training. The application of recursive feature elimination (RFE) further refined the model, enhancing its predictive accuracy and efficiency.

The findings underscored the model's capability in swiftly predicting the Free Energy of Binding (FEB), highlighting its potential as a rapid screening tool for compounds with therapeutic efficacy against cancer. The improvement in model performance through RFE, evidenced by the improved metrics, signifies a step forward in computational drug discovery. The entire dataset and the developed model have been made accessible via a GitHub repository, encouraging further exploration and refinement within the scientific community.

This study not only advances the understanding of potential cancer treatments but also exemplifies the integration of computational and biological research methodologies in the quest for effective cancer therapeutics.

ACKNOWLEDGMENT

JAAH thanks the financial support of the Management Agreement No. 237-2015-FONDECYT to the Vice-Rectorate for Research of the Universidad Nacional de Ingeniería (Peru). MASS and KSM acknowledge the financial support given by CAPES Financial Code 001, FAPERGS processes 22/2551-0000385-0 and 22/2551-0000390-7 and CNPq grants 405745/2021-4 and 440363/2022-5.

REFERENCES

[1] Haruna L Barazorda-Ccahuana, Lena Gálvez Ranilla, Mayron Antonio Candia-Puma, Eymi Gladys Cárcamo-Rodriguez, Angela Emperatriz Centeno-Lopez, Gonzalo Davila-Del-Carpio, José L Medina-Franco, and Miguel Angel Chávez-Fumagalli. Perunpdb: The peruvian natural products database for in silico drug screening. *Scientific Reports*, 13(1):7577, 2023.

[2] Yahui Ding, Xiaoping Chen, Can Liu, Weizhi Ge, Qin Wang, Xin Hao, Mengmeng Wang, Yue Chen, and Quan Zhang. Identification of a small molecule as inducer of ferroptosis and apoptosis through ubiquitination of gpx4 in triple negative breast cancer cells. *Journal of hematology & oncology*, 14:1–21, 2021.

[3] Jerome Eberhardt, Diogo Santos-Martins, Andreas F Tillack, and Stefano Forli. Autodock vina 1.2. 0: New

docking methods, expanded force field, and python bindings. *Journal of chemical information and modeling*, 61(8):3891–3898, 2021.

[4] Ani Fatonah, Usman Sumo Friend Tambunan, Wilis Okti Pamungkas, Gana Lahirin Dewanto, and Ig Satrio Wicaksono. Discovery of gpx4 inhibitor by molecular docking simulation as a potential ferroptosis inducer. *Biointerface Research in Applied Chemistry*, 10(1):4929–4933, 2020.

[5] Karuna Gokarn, Vishwas Sarangdhar, and Ramprasad B Pal. Effect of microbial siderophores on mammalian non-malignant and malignant cell lines. *BMC complementary and alternative medicine*, 17:1–11, 2017.

[6] Frauke Günther and Stefan Fritsch. Neuralnet: training of neural networks. *R J.*, 2(1):30, 2010.

[7] Qianqian Guo, Liwen Li, Shanshan Hou, Ziqiao Yuan, Chenhui Li, Wenzhou Zhang, Lufeng Zheng, and Xiaoman Li. The role of iron in cancer progression. *Frontiers in Oncology*, 11:778492, 2021.

[8] Matthew D Hall, Travis S Marshall, Alexandra DT Kwit, Lisa M Miller Jenkins, Andrés E Dulcey, James P Madigan, Kristen M Pluchino, Andrew S Goldsborough, Kyle R Brimacombe, Gary L Griffiths, et al. Inhibition of glutathione peroxidase mediates the collateral sensitivity of multidrug-resistant cells to tiopronin. *Journal of Biological Chemistry*, 289(31):21473–21489, 2014.

[9] Jesus Antonio Alvarado Huayhuaz, Hector Aguilar Vitorino, Othon Souto Campos, Silvia Helena Pires Serrano, Telma Mary Kaneko, and Breno Pannia Espósito. Desferrioxamine and desferrioxamine-caffeine as carriers of aluminum and gallium to microbes via the trojan horse effect. *Journal of trace elements in medicine and biology*, 41:16–22, 2017.

[10] Greg Landrum. Rdkit documentation. *Release*, 1(1-79):4, 2013.

[11] Bin Li, Breno Pannia Espósito, Shunhao Wang, Jie Zhang, Ming Xu, Shuping Zhang, Zhihong Zhang, and Sijin Liu. Desferrioxamine-caffeine shows improved efficacy in chelating iron and depleting cancer stem cells. *Journal of Trace Elements in Medicine and Biology*, 52:232–238, 2019.

[12] Jie Li, Feng Cao, He-liang Yin, Zi-jian Huang, Zhi-tao Lin, Ning Mao, Bei Sun, and Gang Wang. Ferroptosis: past, present and future. *Cell death & disease*, 11(2):88, 2020.

[13] Tianyu Ma, Jingtong Du, Yufeng Zhang, Yuyao Wang, Bingxuan Wang, and Tianhong Zhang. Gpx4-independent ferroptosis—a new strategy in disease's therapy. *Cell death discovery*, 8(1):434, 2022.

[14] Scot Mente and Max Kuhn. The use of the r language for medicinal chemistry applications. *Current Topics in Medicinal Chemistry*, 12(18):1957–1964, 2012.

[15] Daniel James Raines, Thomas James Sanderson, AK Duhme-Klair, and Ellis Wilde. Siderophores. In *Elsevier Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*. Elsevier, 2015.

[16] Kenneth N Raymond, Benjamin E Allred, and Allyson K Sia. Coordination chemistry of microbial iron transport. *Accounts of chemical research*, 48(9):2496–2505, 2015.

[17] Piu Saha, Beng San Yeoh, Xia Xiao, Rachel M Golonka, Sivarajan Kumarasamy, and Matam Vijay-Kumar. Enterobactin, an iron chelating bacterial siderophore, arrests cancer cell proliferation. *Biochemical pharmacology*, 168:71–81, 2019.

[18] Fernanda I Saldivar-González, Eli Fernández-de Gortari, and José L Medina-Franco. Inteligencia artificial en el diseño de fármacos: hacia la inteligencia aumentada. *Educación química*, 34(2):17–25, 2023.

[19] Rebecca L Siegel, Angela N Giaquinto, and Ahmedin Jemal. Cancer statistics, 2024. *CA: a cancer journal for clinicians*, 74(1):12–49, 2024.

[20] Fredson Torres Silva and Breno Pannia Espósito. Intracellular iron binding and antioxidant activity of phytochelators. *Biological Trace Element Research*, 200(8):3910–3918, 2022.

[21] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

[22] Chun Wei Yap. Padel-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, 32(7):1466–1474, 2011.