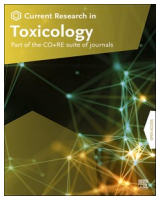
Current Research in Toxicology 6 (2024) 100158

Contents lists available at ScienceDirect

Current Research in Toxicology

journal homepage: www.journals.elsevier.com/current-research-in-toxicology



Identification of estrogen receptor agonists among hydroxylated polychlorinated biphenyls using classification-based quantitative structure–activity relationship models

Lukman K. Akinola [a,b,*], Adamu Uzairu [a], Gideon A. Shallangwa [a], Stephen E. Abechi [a], Abdullahi B. Umar [a]

[a] Department of Chemistry, Ahmadu Bello University, Zaria, Nigeria
[b] Department of Chemistry, Bauchi State University, Gadau, Nigeria

ARTICLE INFO

Keywords:
Autocorrelation descriptor
Binary logistic regression
Estrogen receptor
Hydroxylated polychlorinated biphenyl
Quantitative structure–activity relationship

ABSTRACT

Identification of estrogen receptor (ER) agonists among environmental toxicants is essential for assessing the potential impact of toxicants on human health. Using 2D autocorrelation descriptors as predictor variables, two binary logistic regression models were developed to identify active ER agonists among hydroxylated polychlorinated biphenyls (OH-PCBs). The classifications made by the two models on the training set compounds resulted in accuracy, sensitivity and specificity of 95.9 %, 93.9 % and 97.6 % for ERα dataset and 91.9 %, 90.9 % and 92.7 % for ERβ dataset. The areas under the ROC curves, constructed with the training set data, were found to be 0.985 and 0.987 for the two models. Predictions made by models I and II correctly classified 84.0 % and 88.0 % of the test set compounds and 89.8 % and 85.8% of the cross-validation set compounds respectively. The two classification-based QSAR models proposed in this paper are considered robust and reliable for rapid identification of ERα and ERβ agonists among OH-PCB congeners.

Introduction

The demand and supply of new chemicals by industrialized society has resulted in the release of large amount of diverse chemicals into the natural environment. Among these environmental contaminants, polychlorinated biphenyls (PCBs) stand out because of their resistance to biodegradation (Borja et al., 2005). Due to their lipophilic character, numerous congeners of PCBs accumulate in the tissues of organisms at successively higher levels within the food web, potentially posing detrimental health risk to human beings (Fernández-González et al., 2015; Zhu et al., 2015). PCBs are synthetic organic compounds that were first synthesized in the early 1880s (Pentyala et al., 2011). Because of their diverse applications, millions of tons of PCBs were commercially produced globally for over six decades but their production was thereafter banned and discontinued in later years due to their persistence and potential for bioaccumulation (Lallas, 2001; Warmuth and Ohno, 2013). Although several studies have shown that PCBs by themselves pose significant risks to human health, the discovery that PCBs are capable of undergoing hydroxylation via biotic and abiotic means to form hydroxylated polychlorinated biphenyls (OH-PCBs) has raised serious environmental concern because OH-PCBs are several orders of magnitude more toxic than the parent PCBs (Tehrani and Aken, 2014).

Identification of nuclear receptor agonists and antagonists among environmental toxicants is fundamental in assessing the potential impact of toxicants on human health. This is because many environmental toxicants are known to cause adverse health effects in humans by inappropriately interacting with nuclear receptors (Toporova and Balaguer, 2020). Although some *in vitro* experiments have reliably been used by researchers to investigate the mechanisms of endocrine disruption via nuclear receptor binding (Baker, 2001), routine use of these experimental techniques for large scale screening of environmental toxicants is considered expensive, laborious and time-consuming (Shukla et al., 2010; Tukker et al., 2016; Lang et al., 2018; Sakkiah et al., 2019). *In vitro* studies to measure the nuclear receptor binding activities of some OH-PCB congeners were reported in literature (Arulmozhiraja et al., 2005; Takeuchi et al., 2011; Kamata et al., 2019). To accelerate the time and reduce the cost and effort required for toxicity testing in future experimental studies, computational methodologies designed to

* Corresponding author at: Department of Chemistry, Bauchi State University, Gadau, Nigeria.
E-mail address: lkakinola@basug.edu.ng (L.K. Akinola).

https://doi.org/10.1016/j.crtox.2024.100158
Received 23 November 2023; Received in revised form 22 February 2024; Accepted 22 February 2024
Available online 23 February 2024

Table 1
Names and agonistic activities of 17β-estradiol and 99 hydroxylated polychlorinated biphenyls in ERα and ERβ datasetsa,b.

Compound ID	Chemical name	REC_{20}(M)	
		ERα	ERβ
	17β-estradiol	$2.5x10^{-12}$	$5.3x10^{-12}$
P1	2-chlorobiphenyl-4-ol	$4.4x10^{-8}$	$2.9x10^{-8}$
P2	3-chlorobiphenyl-4-ol	–	$2.1x10^{-6}$
P3	5-chlorobiphenyl-2-ol	$4.9x10^{-6}$	$3.0x10^{-6}$
P4	2′,3′-dichlorobiphenyl-2-ol	$3.4x10^{-6}$	$4.6x10^{-6}$
P5	2′,5′-dichlorobiphenyl-2-ol	$4.2x10^{-6}$	$2.1x10^{-6}$
P6	2′,5′-dichlorobiphenyl-3-ol	$4.4x10^{-7}$	$7.3x10^{-7}$
P7	2′,5′-dichlorobiphenyl-4-ol	$5.9x10^{-8}$	$1.0x10^{-8}$
P8	3′,4′-dichlorobiphenyl-2-ol	$5.3x10^{-6}$	$3.7x10^{-6}$
P9	3,5-dichlorobiphenyl-4-ol	–	$6.7x10^{-6}$
P10	2,2′,5′-trichlorobiphenyl-3-ol	$4.0x10^{-7}$	$5.7x10^{-7}$
P11	2,2′,5′-trichlorobiphenyl-4-ol	$7.0x10^{-8}$	$2.6x10^{-8}$
P12	2′,3,3′-trichlorobiphenyl-2-ol	$4.1x10^{-6}$	–
P13	2′,3,3′-trichlorobiphenyl-4-ol	$4.0x10^{-6}$	$1.3x10^{-6}$
P14	2′,3,4′-trichlorobiphenyl-2-ol	$4.2x10^{-6}$	$5.5x10^{-6}$
P15	3′,4,6-trichlorobiphenyl-2-ol	–	–
P16	2,3′,4-trichlorobiphenyl-3-ol	–	–
P17	2′,3,4′-trichlorobiphenyl-4-ol	$1.3x10^{-6}$	$4.8x10^{-7}$
P18	3′,4,6-trichlorobiphenyl-3-ol	–	–
P19	3,3′,6-trichlorobiphenyl-2-ol	–	–
P20	2′,3,5′-trichlorobiphenyl-4-ol	$1.3x10^{-6}$	$4.5x10^{-7}$
P21	2,3′,5-trichlorobiphenyl-4-ol	$2.2x10^{-6}$	$1.8x10^{-6}$
P22	2′,5,5′-trichlorobiphenyl-2-ol	–	–
P23	2,4,4′-trichlorobiphenyl-3-ol	$6.0x10^{-6}$	$3.9x10^{-6}$
P24	2′,4′,6′-trichlorobiphenyl-2-ol	$2.2x10^{-6}$	$1.1x10^{-6}$
P25	2′,4′,6′-trichlorobiphenyl-3-ol	$1.2x10^{-7}$	$7.3x10^{-8}$
P26	2′,4′,6′-trichlorobiphenyl-4-ol	$3.4x10^{-9}$	$1.7x10^{-9}$
P27	2′,4,5′-trichlorobiphenyl-3-ol	–	–
P28	2,4′,5-trichlorobiphenyl-4-ol	$4.1x10^{-6}$	$8.0x10^{-7}$
P29	3,4′,6-trichlorobiphenyl-2-ol	–	–
P30	2,3′,4′-trichlorobiphenyl-4-ol	$6.7x10^{-8}$	$2.3x10^{-8}$
P31	3′,4′,6-trichlorobiphenyl-3-ol	$4.0x10^{-7}$	$1.1x10^{-7}$
P32	3′,5′,6-trichlorobiphenyl-3-ol	$8.3x10^{-7}$	$4.8x10^{-7}$
P33	3,3′,4′-trichlorobiphenyl-2-ol	–	–
P34	3,3′,4′-trichlorobiphenyl-4-ol	$4.5x10^{-6}$	$6.6x10^{-7}$
P35	3′,4′,5-trichlorobiphenyl-2-ol	–	–
P36	3,3′,5-trichlorobiphenyl-2-ol	–	–
P37	3,3′,5′-trichlorobiphenyl-4-ol	$3.0x10^{-6}$	$1.0x10^{-6}$
P38	3,3′,5′-trichlorobiphenyl-2-ol	–	–
P39	3′,5,5′-trichlorobiphenyl-2-ol	–	–
P40	3′,4,4′-trichlorobiphenyl-3-ol	–	–
P41	3,4′,5-trichlorobiphenyl-2-ol	–	–
P42	3′,4,5′-trichlorobiphenyl-2-ol	–	–
P43	3,4′,5-trichlorobiphenyl-4-ol	–	–
P44	2,2′,3′,5-tetrachlorobiphenyl-4-ol	$2.1x10^{-6}$	$6.0x10^{-6}$
P45	2,2′,3′,6-tetrachlorobiphenyl-3-ol	$2.5x10^{-7}$	$3.0x10^{-7}$
P46	2′,4,5′,6-tetrachlorobiphenyl-2-ol	–	–
P47	2,2′,4′,6′-tetrachlorobiphenyl-4-ol	$1.3x10^{-9}$	$3.4x10^{-9}$
P48	2,2′,5,5′-tetrachlorobiphenyl-3-ol	$6.6x10^{-6}$	–
P49	2,2′,5′,6-tetrachlorobiphenyl-3-ol	$1.6x10^{-6}$	–
P50	2′,3,3′,4-tetrachlorobiphenyl-2-ol	–	–
P51	2,3,3′,4′-tetrachlorobiphenyl-4-ol	$4.0x10^{-6}$	$5.2x10^{-7}$
P52	2′,3,3′,5-tetrachlorobiphenyl-2-ol	–	–
P53	3′,5,5′,6-tetrachlorobiphenyl-2-ol	–	–
P54	2′,3′,4′,5′-tetrachlorobiphenyl-2-ol	$2.5x10^{-6}$	$1.4x10^{-6}$
P55	2′,3′,4′,5′-tetrachlorobiphenyl-3-ol	$2.4x10^{-7}$	$1.4x10^{-7}$
P56	2′,3′,4′,5′-tetrachlorobiphenyl-4-ol	$1.9x10^{-7}$	$1.1x10^{-7}$
P57	2′,3′,5′,6′-tetrachlorobiphenyl-2-ol	$9.4x10^{-6}$	–
P58	2′,3′,5′,6′-tetrachlorobiphenyl-3-ol	$1.0x10^{-6}$	$5.7x10^{-7}$
P59	2′,3′,5′,6′-tetrachlorobiphenyl-4-ol	$8.2x10^{-8}$	$4.7x10^{-9}$
P60	2′,3,4,4′-tetrachlorobiphenyl-2-ol	$6 .2x10^{-6}$	$5.8x10^{-7}$
P61	2,3′,4,4′-tetrachlorobiphenyl-3-ol	–	–
P62	3′,4,4′,6-tetrachlorobiphenyl-3-ol	–	$5.3x10^{-6}$
P63	2′,3,4′,5-tetrachlorobiphenyl-2-ol	–	–
P64	2,3′,4,5′-tetrachlorobiphenyl-3-ol	–	–
P65	2′,3,4′,5-tetrachlorobiphenyl-4-ol	–	–
P66	3′,4,5′,6-tetrachlorobiphenyl-3-ol	–	–
P67	3′,4,5′,6-tetrachlorobiphenyl-2-ol	–	–
P68	2′,3,4′,6′-tetrachlorobiphenyl-4-ol	$2.7x10^{-7}$	$9.1x10^{-9}$

Table 1 (*continued*)

Compound ID	Chemical name	REC_{20}(M)	
		ERα	ERβ
P69	2′,4′,5,6′-tetrachlorobiphenyl-2-ol	–	–
P70	2,3′,4′,5-tetrachlorobiphenyl-4-ol	–	–
P71	3,3′,4′,6-tetrachlorobiphenyl-2-ol	–	–
P72	2′,3,5,5′-tetrachlorobiphenyl-2-ol	–	–
P73	2′,3,5,5′-tetrachlorobiphenyl-4-ol	–	–
P74	2,3′,5,5′-tetrachlorobiphenyl-4-ol	–	–
P75	3,3′,4,4′-tetrachlorobiphenyl-2-ol	–	–
P76	3,3′,4′,5-tetrachlorobiphenyl-2-ol	–	–
P77	3,3′,4′,5-tetrachlorobiphenyl-4-ol	–	–
P78	3,3′,5,5′-tetrachlorobiphenyl-2-ol	–	–
P79	2′,3,3′,5,6-pentachlorobiphenyl-2-ol	–	–
P80	2,2′,3′,4′,5′-pentachlorobiphenyl-4-ol	$6.5x10^{-8}$	$3.8x10^{-8}$
P81	2,2′,3′,5′,6′-pentachlorobiphenyl-4-ol	$1.8x10^{-7}$	$1.0x10^{-7}$
P82	2,2′,3′,4,6-pentachlorobiphenyl-3-ol	$1.3x10^{-6}$	$5.3x10^{-6}$
P83	2′,3,3′,4′,5′-pentachlorobiphenyl-4-ol	$6.3x10^{-7}$	$4.6x10^{-7}$
P84	2′,3′,4′,5,5′-pentachlorobiphenyl-2-ol	–	–
P85	2,3,3′,4′,5-pentachlorobiphenyl-4-ol	–	–
P86	3,3′,4′,5,6-pentachlorobiphenyl-2-ol	–	–
P87	2,3′,4′,5,6-pentachlorobiphenyl-3-ol	$9.4x10^{-6}$	$5.0x10^{-6}$
P88	3,3′,5,5′,6-pentachlorobiphenyl-2-ol	–	–
P89	2′,3,3′,5′,6′-pentachlorobiphenyl-4-ol	–	$6.2x10^{-6}$
P90	2′,3′,5,5′,6′-pentachlorobiphenyl-2-ol	–	–
P91	2,3′,5,5′,6-pentachlorobiphenyl-3-ol	–	–
P92	2,3′,4,4′,6-pentachlorobiphenyl-3-ol	–	–
P93	2′,3,4′,5,6′-pentachlorobiphenyl-2-ol	–	–
P94	2,3′,4,5′,6-pentachlorobiphenyl-3-ol	–	–
P95	2′,3,4′,5,6′-pentachlorobiphenyl-4-ol	–	–
P96	2,2′,3,4′,5,5′-hexachlorobiphenyl-4-ol	–	–
P97	2′,3,3′,4′,5,5′-hexachlorobiphenyl-4-ol	–	–
P98	2′,3,3′,5,5′,6′-hexachlorobiphenyl-4-ol	–	–
P99	2,2′,3,4′,5,5′,6-heptachlorobiphenyl-4-ol	–	–

[a] REC_{20} values taken fromTakeuchi et al. (2011).
[b] REC_{20} value denotes 20 % relative effective concentration.

prioritize the selection of OH-PCB congeners that could act as agonists or antagonists of nuclear receptors are required.

Quantitative structure–activity relationship (QSAR) modeling is a computational approach that establishes a correlation between biological activities and theoretically-computed molecular descriptors (or experimentally-measured properties) of a series of chemical compounds. QSAR modeling relies on the assumption that changes in the molecular structures of chemical compounds reflect corresponding changes in the observed biological activities (Rogers and Hopfinger, 1994). QSAR models can either be classification-based or regression-based, depending on whether the response variables are discrete class labels or continuous quantities (Ambure et al., 2019). Classification-based QSAR models are particularly useful for rapid identification of environmental toxicants that can act as nuclear receptor agonists or antagonists. Consequently, classification-based QSAR models have the potential to be employed for rational selection and prioritization of chemicals in nuclear receptor-mediated toxicity studies. To the best of our knowledge, apart from the two classification-based QSAR models reported by Akinola et al. (2023) for identification of thyroid hormone receptor agonists among OH-PCB congeners, no classification-based QSAR model is currently available in the literature to identify OH-PCB congeners that could act as agonists or antagonists of other nuclear receptors. The objective of the present study was to develop classification-based QSAR models that can be utilized for rapid identification of ERα and ERβ agonists among OH-PCB congeners using 2D autocorrelation descriptors as predictor variables.

Materials and methods

Datasets

The datasets used for developing and validating the binary logistic regression models reported in this paper were obtained from literature (Takeuchi et al., 2011). These datasets were generated in an *in vitro* investigation involving measurement of agonistic activities of 99 monohydroxylated polychlorinated biphenyls (OH-PCBs) against human estrogen receptor α (ERα dataset) and human estrogen receptor β (ERβ dataset) using reporter gene assays. According to Takeuchi et al. (2011), the estrogenic activity of a test compound (OH-PCB) in the *in vitro* reporter gene assay was defined as the concentration of the test compound (OH-PCB) that produced a response that equals 20 % of the maximal response produced by 17β-estradiol in assays conducted under similar condition. Of the 99 OH-PCB congeners tested in the *in vitro* reporter gene assays, 44 and 55 congeners were observed to be active and inactive estrogen receptor agonists respectively in both ERα and ERβ datasets. Table S1 (Supplementary Material) shows the names, 2D structures, CAS registry numbers and agonistic activities of the 99 OH-PCBs in ERα and ERβ datasets. An abridged version of Table S1 is shown in Table 1.

Calculation and preprocessing of molecular descriptors

Two-dimensional structure of each of the 99 OH-PCB molecules listed in Table 1 was drawn using the 2D sketch palette in Spartan '14 software (Shao et al., 2006). These 2D structures were converted into 3D structures and then optimized using semi-empirical AM1 model as implemented in Spartan '14 software (Shao et al., 2006). Geometry optimization of the 3D structures of OH-PCB molecules is required in order to minimize the energy of the structures. These optimized structures were then imported into PaDEL-Descriptor software and a total of 346 2D autocorrelation descriptors were calculated for each OH-PCB molecule in the dataset (Yap, 2011). Intercorrelated descriptors (redundant descriptors) and descriptors with constant or nearly constant values (irrelevant descriptors) were eliminated from the pool of 346 2D autocorrelation descriptors calculated by PaDEL-Descriptor software (Yap, 2011). In this paper, intercorrelated descriptors with correlation coefficient exceeding 0.90 and constant-value descriptors with variance lower than 0.0001 were removed using V-WSP algorithm (Ballabio et al., 2014) as implemented in V-WSP tool (version 1.2) developed by Ambure et al. (2015). Correlation matrix was constructed to verify the absence of multicollinearity in the final 2D autocorrelation descriptors selected for model building.

Dataset division

The 44 OH-PCB congeners listed as active compounds in Table 1 were divided into training and test sets, with the training set being 75 % of the total active compounds and the test set being 25 % of the total active compounds. The 55 inactive compounds listed in Table 1 were also divided into training and test sets, with the training set being 75 % of the entire inactive compounds and the test set being 25 % of the entire inactive compounds. The dataset division procedure described above was implemented in Dataset Division GUI 1.2 developed by Ambure et al. (2015) using Kennard-Stone algorithm (Kennard and Stone, 1969; Snarey et al., 1997; Martin et al., 2012). The 33 active compounds and the 41 inactive compounds assigned to the training set were then combined to form 74 training set compounds. These 74 training set compounds were used to develop the binary logistic regression models reported in this paper. Similarly, the 11 active compounds and the 14 inactive compounds assigned to the test set were also combined to form 25 test set compounds. These 25 test set compounds were reserved for external validation of the developed models. The 74 OH-PCB congeners assigned to the training sets in the ERα and ERβ datasets, along with the values of 2D autocorrelation descriptors selected for model building in these datasets, are shown in Tables S2 and S3 respectively (Supplementary Material). Similarly, the 25 OH-PCB congeners assigned to the test sets in the ERα and ERβ datasets, along with the values of 2D autocorrelation descriptors used for model validation, are shown in Tables S4 and S5 respectively (Supplementary Material).

Development of binary logistic regression models

The two classification-based QSAR models reported for both ERα and ERβ datasets in this paper were developed using binary logistic regression as implemented in IBM® SPSS® Statistics (version 26). In this multivariate statistical method, the 2D autocorrelation descriptors shown in Tables S2 and S3 (Supplementary Material) for the training set compounds (74 OH-PCB congeners) were used as input independent variables while the coded values of discrete class labels of compounds in the training set (1 for active OH-PCBs and 0 for inactive OH-PCBs) were used as the outcome variable. Feature selection was carried out using forward conditional procedure as implemented in IBM® SPSS® Statistics (version 26). The statistical significance of each of the chosen predictor variables was assessed through Wald test (Hosmer et al., 2013). A predictor variable is considered statistically significant, and thus included in a binary logistic regression model, if the p-value obtained from its Wald test is less than 0.05 (Hosmer et al., 2013). Additionally, the developed models were evaluated for their goodness-of-fit using both Omnibus test and Hosmer-Lemeshow test (Bewick et al., 2005; Goeman and le Cessie, 2006; Stoltzfus, 2011). In the context of binary logistic regression, goodness-of-fit serves as an indicator of how well a model fit the data used in building the model (Goeman and le Cessie, 2006; Stoltzfus, 2011). Furthermore, Nagelkerke R square and Cox & Snell R square were calculated to assess the proportion of the total variation in the outcome variable that could be explained by the predictor variables (Sarma and Vardhan, 2019). Nagelkerke R square is an adjusted variant of Cox & Snell R square that adjusts the scale of the statistic to cover the full range from 0 to 1 (Bewick et al., 2005; Sapra, 2014). Finally, the binary logistic regression models generated for both ERα and ERβ datasets were used to compute the logit values and posterior probabilities of group memberships for all the training set compounds.

Model validation

The 2D autocorrelation descriptors selected for model building in ERα and ERβ datasets were also calculated for test set compounds selected from ERα and ERβ datasets. Utilizing these 2D autocorrelation descriptors, logit values and posterior probabilities of group memberships were calculated for the test set compounds in both ERα and ERβ datasets. The computed posterior probabilities were then used to categorize the test set compounds into either active or inactive estrogen receptor agonists based on a predetermined optimal decision threshold. Compounds with predicted probabilities that equal or greater than the optimal decision threshold were classified as active and those with predicted probabilities less than the optimal decision threshold as inactive. The predictions made on the test set compounds were then used to evaluate the predictive abilities of the developed QSAR models using the performance metrics described later in this paper. To provide a more robust estimate of predictive capacities of the developed models, K-fold cross-validation was also employed to evaluate the performance of the models. In this approach, each of the datasets was partitioned into K subsets of approximately equal size, iteratively training the model on K-1 folds while testing on the held-out fold (Yadav and Shukla, 2016). This process was repeated K times, ensuring each fold served as training set in K-1 times and as validation set once. Value of $K = 5$ was chosen in the cross-validation procedure described above, maintaining a 4:5 ratio of active to inactive compounds in each fold (imbalance ratio in both ERα and ERβ datasets is 1.25). The performance metrics described later

in this paper were then calculated for each of the cross-validation sets and the average performance metrics across folds provided a robust estimate of models' predictive performance.

Determination of optimal decision threshold

To categorize the compounds in the training set, test set and cross-validation sets into active and inactive compounds, the probabilities predicted for compounds in each subgroup were compared to a pre-determined optimum decision threshold. For a well-balanced dataset, the default decision threshold is always set at 0.5 (Esposito et al., 2021). This implies that a compound is categorized as active when its predicted probability equals or exceeds 0.5, and as inactive when its predicted probability is less than 0.5. However, this default decision threshold may not be applicable when dealing with imbalanced datasets, as is the case in the present study (Esposito et al., 2021). Imbalanced datasets exhibit a disproportionate distribution of classes, with one class being significantly more prevalent than the other (He and Garcia, 2009). To determine the optimal threshold for each of the datasets used in this paper, values of F_1 score (Eq. (8), Youden's index (Eq. 9) and geometric mean (Eq. 10) were calculated at different probability thresholds. The threshold corresponding to the point of maximum F_1 score, maximum Youden's index and maximum geometric mean was selected as the optimal decision threshold for the classification (Schisterman et al., 2005; Zou et al., 2016; Hancock et al., 2022). By employing these three approaches, the determination of the optimal decision threshold was effectively ensured, thereby overcoming the challenges posed by the imbalanced nature of the datasets utilized in the present study.

Assessment of model performance

The classifications obtained for the compounds assigned to training, test and cross-validation sets in both ERα and ERβ datasets were organized in a specific table layout known as confusion matrix that allows easy calculation of true positive (TP), true negative (TN), false positive (FP) and false negative (FN). In this paper, TP and TN refer to the number of active and inactive OH-PCB congeners that were correctly classified by the models as active and inactive estrogen receptor agonists respectively while FP and FN refer to the number of inactive and active OH-PCB congeners that were misclassified by the models as active and inactive estrogen receptor agonists respectively. From the values of TP, TN, FP and FN obtained, performance metrics such as accuracy (ACC), sensitivity or recall or true positive rate (TPR), specificity or true negative rate (TNR), precision or positive predictive value (PPV), negative predictive value (NPV), F_1 score (F_1), balanced accuracy (BA) and Matthews correlation coefficient (MCC) were calculated for the prediction made on training set compounds, test set compounds and cross-validation set compounds using the formulae shown in Eqs. (1)–(8). The performance of the developed models vis-à-vis the classifications made on the training set compounds in both ERα and ERβ datasets were also evaluated graphically using receiver operating characteristic (ROC) curves.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

$$TNR = \frac{TN}{TN + FP} \tag{3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

$$NPV = \frac{TN}{TN + FN} \tag{5}$$

Table 2
Symbols and definitions of molecular descriptors utilized in building models I and II.

Symbol	Definition	Type	Class
ATS8m	Moreau-Broto autocorrelation–lag 8/ weighted by mass	2D	Autocorrelation
ATS6e	Moreau-Broto autocorrelation–lag 6/ weighted by Sanderson electronegativity	2D	Autocorrelation
ATSC3e	Centered Moreau-Broto autocorrelation–lag 3/weighted by Sanderson electronegativity	2D	Autocorrelation
ATSC3p	Centered Moreau-Broto autocorrelation–lag 3 weighted by polarizability	2D	Autocorrelation
MATS5p	Moran autocorrelation–lag 4/weighted by polarizability	2D	Autocorrelation
GATS7c	Geary autocorrelation–lag 7/weighted by charge	2D	Autocorrelation
GATS4s	Geary autocorrelation–lag 4/weighted by intrinsic state	2D	Autocorrelation

$$BA = \frac{TPR + TNR}{2} \tag{6}$$

$$MCC = \frac{TPxTN - FPxFN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{7}$$

$$F_1 = \frac{2xPrecisionxRecall}{Precision + Recall} \tag{8}$$

$$Youden's\ index(J) = Sensitivity + Specificity - 1 \tag{9}$$

$$Geometric\ mean = \sqrt{Sensitivity \times Specificity} \tag{10}$$

Results

A total of seven 2D autocorrelation descriptors were collectively selected for building the two binary logistic regression models reported in this paper. The symbols and definitions of these autocorrelation descriptors are shown in Table 2. Of the seven autocorrelation descriptors listed in Table 2, five descriptors (ATS8m, ATS6e, ATSC3e, ATSC3p and GATS4s) were used to develop the structure–activity relationship model for ERα dataset. Similarly, four of the seven autocorrelation descriptors listed in Table 2 (ATS6e, ATSC3p, MATS5p and GATS7c) were employed in building the structure–activity relationship model for ERβ dataset. The values of the five descriptors computed for compounds assigned to the training set and test set in ERα dataset are presented in Tables S2 and S4 respectively (Supplementary Material). Similarly, the values of the four descriptors computed for compounds assigned to the training set and test set in ERβ dataset are presented in Tables S3 and S5 respectively (Supplementary Material). The correlation matrices computed to verify absence of multicollinearity in the autocorrelation descriptors selected for ERα and ERβ datasets are shown in Tables S6 and S7 respectively (Supplementary Material). The absolute values of the correlation coefficients shown in Tables S6 and S7 were found to be very low, suggesting absence of multicollinearity among the autocorrelation descriptors selected for building the two binary logistic regression models reported in this paper.

Using the values of ATS8m, ATS6e, ATSC3e, ATSC3p and GATS4s shown in Table S2 (Supplementary Material) as predictor variables and the coded values of the discrete class labels (1 for active OH-PCBs and 0 for inactive OH-PCBs) of the training set compounds in Table S2 as outcome variable, application of binary logistic regression method produced the logistic regression coefficients (B), their standard errors (S.E.), the p-values of Wald tests, the odds ratios (Exp(B)) and the 95 % confidence intervals of the odds ratios listed in Table 3. From the values of the logistic regression coefficients listed in Table 3, the binary logistic regression model (Model I) displayed in Eq. (11) was constructed. Similarly, application of binary logistic regression method, using the

Table 3
Logistic regression coefficients and odds ratios of 2D autocorrelation descriptors utilized in building model I.

	B	S.E.	Wald	df	p-value	Exp(B)	95 % C.I. for Exp(B)	
							Lower	Upper
ATS8m	−0.001	0.001	4.007	1	0.045	0.999	0.997	1.000
ATS6e	−0.216	0.091	5.643	1	0.018	0.806	0.674	0.963
ATSC3e	−4.799	2.185	4.821	1	0.028	0.008	0.000	0.597
ATSC3p	2.853	0.935	9.303	1	0.002	17.342	2.773	108.471
GATS4s	−7.303	3.152	5.368	1	0.021	0.001	0.000	0.325
Constant	68.055	26.382	6.654	1	0.010	3.597×10^{29}		

Table 4
Logistic regression coefficients and odds ratios of 2D autocorrelation descriptors utilized in building model II.

	B	S.E.	Wald	df	p-value	Exp(B)	95 % C.I. for Exp(B)	
							Lower	Upper
ATS6e	−0.415	0.137	9.239	1	0.002	0.660	0.505	0.863
ATSC3p	2.607	0.860	9.194	1	0.002	13.561	2.514	73.148
MATS5p	−8.932	3.755	5.658	1	0.017	0.000	0.000	0.208
GATS7c	−7.307	2.871	6.479	1	0.011	0.001	0.000	0.186
Constant	120.178	39.664	9.180	1	0.002	1.558×10^{52}		

values of ATS6e, ATSC3p, MATS5p and GATS7c shown in Table S3 (Supplementary Material) as predictor variables and the coded values of the discrete class labels (1 for active OH-PCBs and 0 for inactive OH-PCBs) of the training set compounds in Table S3 as response variable, produced the logistic regression coefficients (B), their standard errors (S.E.), the p-values of Wald tests, the odds ratios (Exp(B)) and the 95 % confidence intervals of the odds ratios listed in Table 4. From the values of the logistic regression coefficients listed in Table 4, the binary logistic regression model (Model II) displayed in Eq. (12) was constructed. The p-value displayed in Tables 3 and 4 for each of the predictor variables indicates that the strength of the relationship between the outcome variable and each of the predictor variables was statistically significant at $p < 0.05$. In Tables 3 and 4, the odds ratio of ATSC3p was found to be greater than one. This indicates that OH-PCB congener with higher value of ATSC3p has higher likelihood of being classified as active ERα agonist or active ERβ agonist. Conversely, the odds ratios of less than one reported in Tables 3 and 4 for the other molecular descriptors indicate that OH-PCB congener with higher value of each of these molecular descriptors has lower likelihood of being classified as active ERα agonist or active ERβ agonist. The results of the Omnibus tests of model coefficients and Hosmer-Lemeshow test to assess the goodness-of-fit of Models I and II are presented in Tables S8 and S9 (Supplementary Material) respectively. The results of the Omnibus tests of model coefficients shown in Table S8 indicate that there were significant improvements in fit ($p < 0.05$) for Models I and II when compared to the null models constructed without any predictor variable. The results of the Hosmer-Lemeshow test presented in Table S9 show no significant difference between the observed outcomes and the outcomes predicted by the models, indicating that both Model I ($\chi^2(8) = 5.187$, $p = 0.737$) and Model II ($\chi^2(8) = 2.872$, $p = 0.942$) adequately fit the data in the training sets. In Table S10 (Supplementary Material), two pseudo-R-squared values were presented for each of Model I and Model II. The Nagelkerke R^2, which is an adjusted version of the Cox and Snell R^2, was used to explain the results presented in Table S10. As shown in Table S10, the Nagelkerke R^2 value of 0.892 reported for Model I and the Nagelkerke R^2 value of 0.881 reported for Model II indicate that 89.2 % of variation in the outcome variable in Model I and 88.1 % of variation in the outcome variable in Model II can be accounted for by the predictor variables in Models I and II respectively.

Table 5
Confusion matrix for the predictions made by models I and II on training set compounds*.

	Class	Predicted group membership		Total
		Active (1)	Inactive (0)	
Model I (ERα dataset)	Active (1)	31	2	33
	Inactive (0)	1	40	41
Model II (ERβ dataset)	Active (1)	30	3	33
	Inactive (0)	3	38	41

Model I (ERα dataset)

$$\ln\left(\frac{P}{1-P}\right) = 68.055 - 0.001\,\mathrm{ATS8m} - 0.216\,\mathrm{ATS6e} - 4.799\,\mathrm{ATSC3e} + 2.853\mathrm{ATSC3p} - 7.303\mathrm{GATS4s} \quad (11)$$

Model II (ERβ dataset)

$$\ln\left(\frac{P}{1-P}\right) = 120.178 - 0.415\,\mathrm{ATS6e} + 2.607\,\mathrm{ATSC3p} - 8.932\,\mathrm{MATS5p} - 7.307\,\mathrm{GATS7c} \quad (12)$$

Having established the fitness of the binary logistic regression models displayed in Eq. (11) (Model I) and Eq. (12) (Model II), the two models were then used to calculate the values of logit for the training set compounds. The probabilities of allotting the training set compounds to active class were then calculated from these logit values. Tables S11 and S12 (Supplementary Material) show the values of logit and predicted probabilities calculated for the training set compounds selected from ERα and ERβ datasets respectively. In order to convert the predicted probabilities obtained in this study to class labels (1 for active and 0 for inactive), an optimal decision threshold was determined for each dataset using the values of F_1 score, Youden's index and geometric mean computed at various probability thresholds in Tables S17 and S18 (Supplementary Material). As shown in Table S17, a maximum F_1 score of 0.818 was obtained at probability thresholds of 0.1, 0.2, 0.3, 0.4, 0.5 and 0.6 for ERα dataset and a maximum F_1 score of 0.857 was obtained at probability thresholds of 0.4 and 0.5 for ERβ dataset. Given that the precisions and recalls reported in Table S17 were indistinguishable at these multiple probability thresholds, an optimal decision threshold of 0.5 was selected for both ERα and ERβ datasets. The result presented above was corroborated by the results shown in Table S18 where maximum Youden's index and maximum geometric mean were also obtained at probability thresholds of 0.1, 0.2, 0.3, 0.4, 0.5 and 0.6 for ERα dataset and at probability thresholds of 0.4 and 0.5 for ERβ dataset.

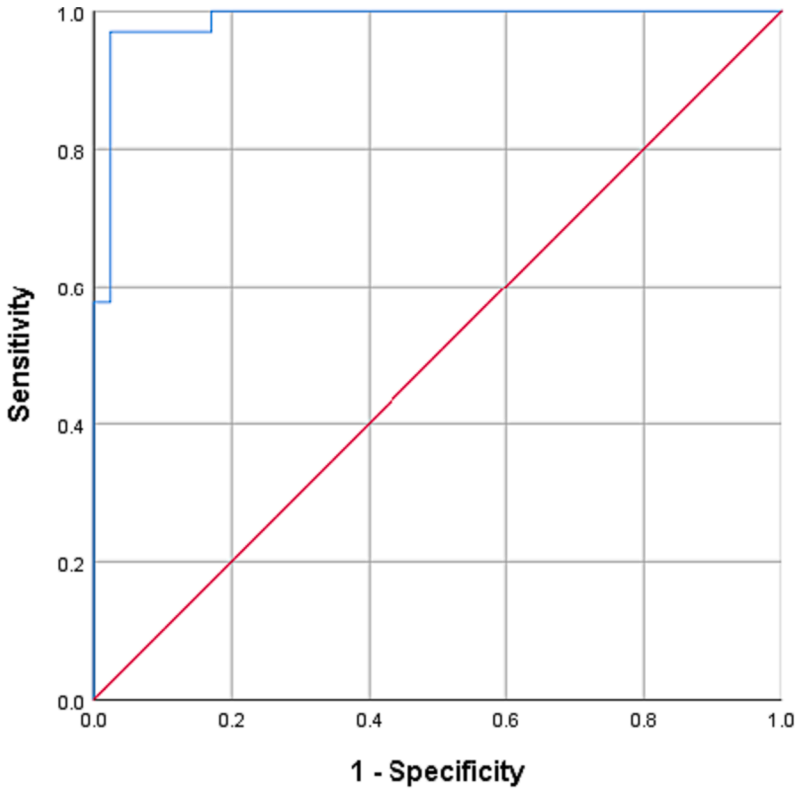


Fig. 1. Receiver operating characteristic (ROC) curve for evaluating the performance of model I on training set compounds selected from ERα dataset.

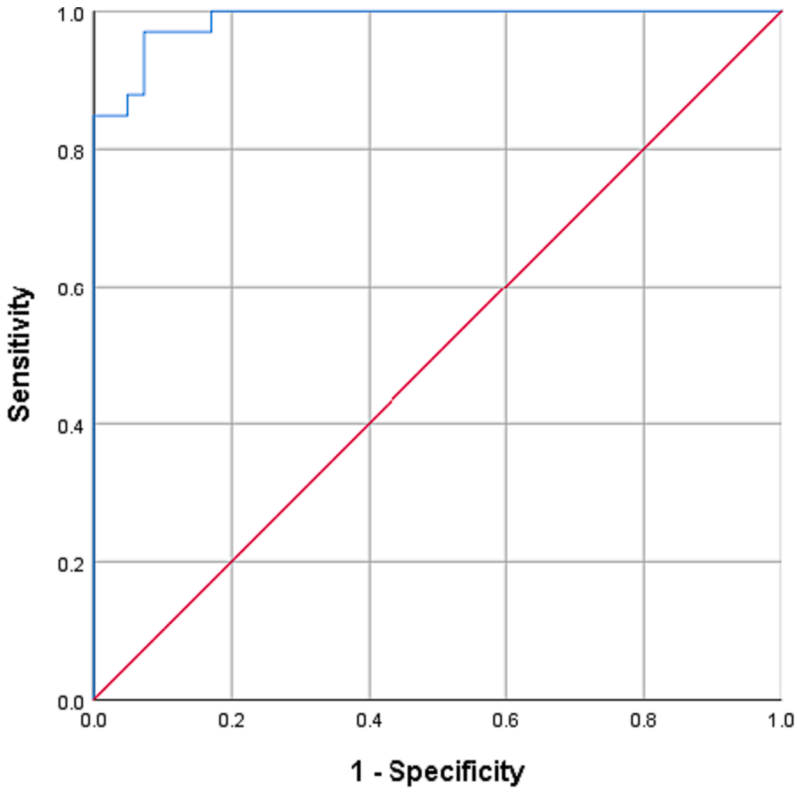


Fig. 2. Receiver operating characteristic (ROC) curve for evaluating the performance of model II on training set compounds selected from ERβ dataset.

Table 6
Confusion matrix for the predictions made by models I and II on test set compounds*.

	Class	Predicted group membership		Total
		Active (1)	Inactive (0)	
Model I (ERα dataset)	Active (1)	9	2	11
	Inactive (0)	2	12	14
Model II (ERβ dataset)	Active (1)	9	2	11
	Inactive (0)	1	13	14

Having established the optimal decision threshold at 0.5, all investigated OH-PCB congeners in this study were classified as active if their predicted probabilities equal or exceed the 0.5 threshold. The classifications of the training set compounds in Tables S11 and S12 by Models I and II are summarized in the confusion matrix shown in Table 5. As shown in Table 5, exactly 31 out of the 33 active compounds and 40 out of the 41 inactive compounds in the ERα training set were correctly classified by Model I (Eq. (11). Table 5 also shows that 30 out of the 33 active compounds and 38 out of the 41 inactive compounds in the ERβ training set were correctly classified by Model II (Eq. (12). Evaluating the performance of the predictions made on the training set compounds by Models I and II, using the metrics listed in Eqs. (1)–(8), resulted in the values of the performance metrics listed in Table 9. As shown in Table 9, the overall accuracy, sensitivity, specificity, precision, negative predictive value, F_1 score, balanced accuracy and Matthews correlation coefficient recorded for the classification made by Model I on the training set compounds obtained from ERα dataset were 95.9 %, 93.9 %, 97.6 %, 96.9 %, 95.2 %, 95.4 %, 95.8 % and 91.8 % respectively. Table 9 also shows that the overall accuracy, sensitivity, specificity, precision, negative predictive value, F_1 score, balanced accuracy and Matthews correlation coefficient recorded for the classification made by Model II on the training set compounds obtained from ERβ dataset were 91.9 %,

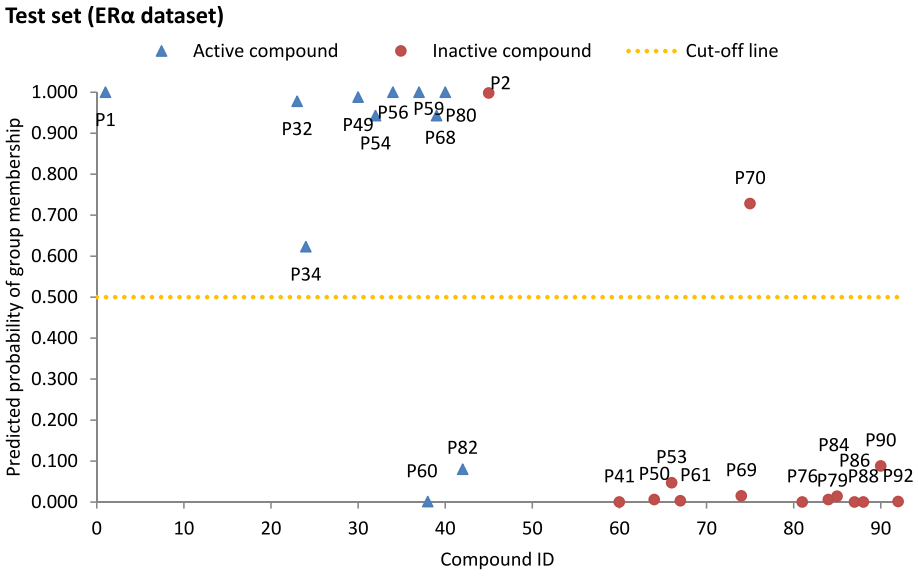


Fig. 3. Graphical representation of classification predicted by model I on test set compounds selected from ERα dataset.

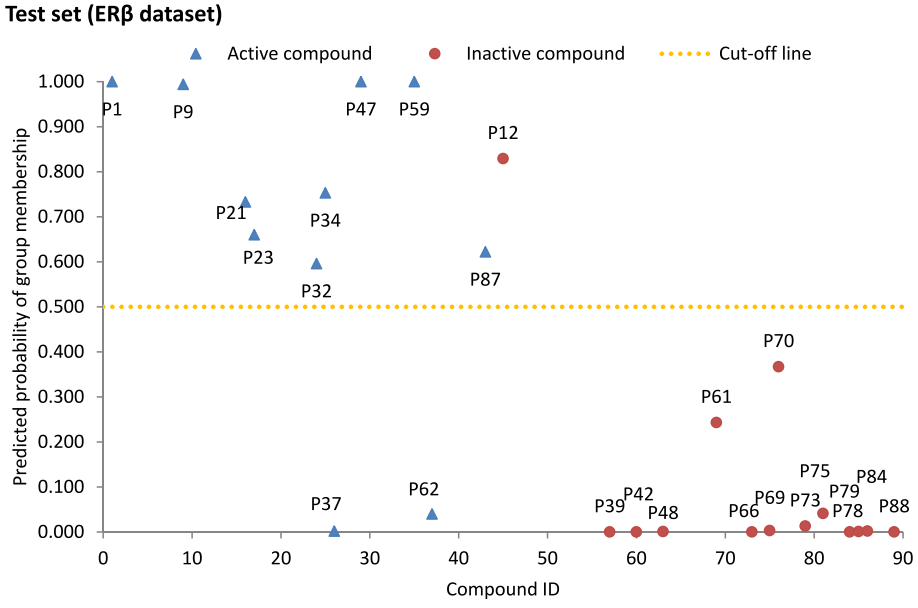


Fig. 4. Graphical representation of classification predicted by model II on test set compounds selected from ERβ dataset.

Table 7
Confusion matrix for validation sets derived from K-fold cross-validation of ERα dataset.

	Class	Predicted group membership		Total
		Active (1)	Inactive (0)	
Validation set 1	Active (1)	8	1	9
	Inactive (0)	0	11	11
Validation set 2	Active (1)	9	0	9
	Inactive (0)	0	11	11
Validation set 3	Active (1)	7	2	9
	Inactive (0)	1	10	11
Validation set 4	Active (1)	7	2	9
	Inactive (0)	1	10	11
Validation set 5	Active (1)	6	2	8
	Inactive (0)	1	10	11

Table 8
Confusion matrix for validation sets derived from K-fold cross-validation of ERβ dataset.

	Class	Predicted group membership		Total
		Active (1)	Inactive (0)	
Validation set 1	Active (1)	9	0	9
	Inactive (0)	3	8	11
Validation set 2	Active (1)	7	2	9
	Inactive (0)	3	8	11
Validation set 3	Active (1)	8	1	9
	Inactive (0)	0	11	11
Validation set 4	Active (1)	8	1	9
	Inactive (0)	1	10	11
Validation set 5	Active (1)	5	3	8
	Inactive (0)	0	11	11

90.9 %, 92.7 %, 90.9 %, 92.7 %, 90.9 %, 91.8 % and 83.6 % respectively. From the values of the performance metrics reported for Models I and II, it can be seen that the classification predicted by Model I in ERα dataset was slightly better than the classification predicted by Model II in ERβ dataset. Overall, the values of the performance metrics reported in Table 9 indicate satisfactory classifications of the training set compounds by the two binary logistic regression models displayed in Eq. (11) (Model I) and Eq. (12) (Model II). The performance of the classifications predicted by Models I and II on the training set compounds in ERα and ERβ datasets was also evaluated graphically using the receiver operating characteristic (ROC) curves shown in Fig. 1 and Fig. 2 respectively. The areas under these ROC curves (AUC) were found to be 0.985 for Model I and 0.987 for Model II. The high values of AUC reported in Fig. 1 and Fig. 2 suggest excellent discriminating abilities of the two classification-based QSAR models displayed in Eq. (11) and Eq. (12).

Finally, the predictive abilities of Models I and II were evaluated using OH-PCB congeners that were not part of the compounds used for model building. To accomplished this task, the two binary logistic regression models displayed in Eq. (11) and Eq. (12) were used to calculate the logit values and the probabilities of allotting the test set compounds to active class using the values of the 2D autocorrelation descriptors shown in Tables S4 and S5 (Supplementary Material). The results of the classifications made by Models I and II on test set compounds selected from ERα and ERβ datasets were shown in Tables S13 and S14 respectively (Supplementary Material). The results presented in Tables S13 and S14 are summarized as confusion matrix in Table 6 and depicted graphically in Fig. 3 and Fig. 4 for better visualization. In Figs. 3 and 4, compounds above the horizontal cut-off lines were classified as active while compounds below the horizontal cut-off lines were classified as inactive. As shown in Fig. 3, of the 25 OH-PCB congeners assigned to test set in ERα dataset, only compounds P2, P60, P70 and P82 were misclassified by Model I. In Fig. 4, among the 25 OH-PCB congeners assigned to test set in ERβ dataset, only compounds P12, P37, and P62 were misclassified by Model II. The proportions of active and inactive compounds that were correctly classified in Fig. 3 by Model I were 81.8 % and 85.7 % respectively. In Fig. 4, the proportions of active and inactive compounds that were correctly classified by Model II were 81.8 % and 92.8 % respectively. The performance of the classifications shown in Table 6 for the test set compounds was further assessed using the values of the performance metrics listed in Table 9. The values of the performance metrics shown in Table 9 for the predictions made by Model I (Eq. (11) and Model II (Eq. (12) on the test set compounds suggest that the two QSAR models have good predictive abilities when applied to new OH-PCB congeners that were not part of the compounds used for model building. To provide a more robust estimate of the predictive capacities of the developed models, K-fold cross-validation was also employed to evaluate the predictive abilities of the developed models. Tables S15 and S16 (Supplementary Material) show the predicted probabilities and predicted group memberships for the validation sets utilized in the K-fold cross-validation of ERα dataset and ERβ dataset respectively. For easy evaluation of model performance vis-à-vis the predictions made on validation set compounds, the results displayed in Tables S15 and S16 (Supplementary Material) are summarized as confusion matrices in Tables 7 and 8 respectively. The average values of the performance metrics derived from Tables 7 and 8 are shown in Table 9. As shown in Table 9, the average values of the performance metrics reported for the cross-validation sets indicate satisfactory predictions, affirming the robustness of the classification-based QSAR models developed in this paper.

Discussion

The present study set out to develop classification-based QSAR models for categorizing OH-PCB congeners into active and inactive estrogen receptor agonists. Application of binary logistic regression method on training set compounds selected from ERα and ERβ datasets, using 2D autocorrelation descriptors as predictor variables, led to the development of two classification-based QSAR models for predicting ERα and ERβ agonists among OH-PCB congeners. Evaluating the performance of the classifications made by the two models on test set compounds selected from ERα and ERβ datasets revealed that the two classification models have good predictive abilities and can reliably be used for identification and prioritization of new ERα and ERβ agonists among OH-PCB congeners. Activation of estrogen receptors via binding to environmental toxicants has been identified as a molecular initiating event in several apical adverse outcomes of toxicant exposure in humans (Shanle and Xu, 2011). For instance, human exposure to endocrine disrupting chemicals has been shown to be associated with reproductive dysfunction, systemic lupus erythematosus, endometrial carcinoma, breast cancer, ovarian cancer and female precocious puberty (Li and McMurray, 2009; Chighizola and Meroni, 2012; Bourguignon and Parent, 2012; Darbre and Williams, 2015; Mallozzi et al., 2017; Ben-Jonathan, 2019; Tam et al., 2022; Caserta et al., 2022). Previous experiments conducted to establish the link between estrogen receptor binding to chemicals and the adverse effects of chemicals in humans were found to be laborious, expensive and time-consuming. Use of classification-based QSAR models for active compound selection and prioritization in experimental studies involving toxicity testing of chemicals can drastically reduce the cost, effort and time required to

Table 9
Values of performance metrics for the predictions made by models I and II on training, test and cross-validation sets.

Performance metric		Training set		Test set		Cross-validation set*	
Metrics	Symbol	Model I	Model II	Model I	Model II	Model I	Model II
Accuracy	ACC	0.959	0.919	0.840	0.880	0.898	0.858
Sensitivity (or true positive rate)	TPR	0.939	0.909	0.818	0.818	0.839	0.836
Specificity (or true negative rate)	TNR	0.976	0.927	0.857	0.929	0.945	0.873
Precision (or positive predictive value)	PPV	0.969	0.909	0.818	0.900	0.921	0.868
Negative predictive value	NPV	0.952	0.927	0.857	0.867	0.883	0.882
F_1 score	F_1	0.954	0.909	0.818	0.857	0.878	0.839
Balanced accuracy	BA	0.958	0.918	0.838	0.873	0.892	0.854
Matthews correlation coefficient	MCC	0.918	0.836	0.675	0.757	0.794	0.728

*Average values of five validation sets.

conduct these experiments. Using different molecular descriptors and different machine learning algorithms for modeling, some classification-based QSAR models to identify agonists or antagonists of thyroid hormone receptor (Bai et al., 2018; Wang and Xing, 2019), estrogen receptor (Liu et al., 2007; Roncaglioni et al., 2008; Wang et al., 2021) and androgen receptor (Piir et al., 2021) among diverse groups of chemical compounds suspected to be endocrine disruptors were reported in the literature. The two classification-based QSAR models reported in this paper were developed using binary logistic regression method. Some of the main attractions of using binary logistic regression algorithm for building classification-based QSAR models include easy implementation of the algorithm, easy interpretation of the resulting models, no assumption about distribution of classes in feature space is required, the algorithm is less inclined to over-fitting, and it is one of the most efficient algorithms when the different outcomes represented by the dataset are linearly separable (Sperandei, 2014; Sarma and Vardhan, 2019). The performance of the two binary logistic regression models developed in the present paper was found to be comparable with the performance of the classification-based QSAR models previously reported in the literature (Liu et al., 2007; Roncaglioni et al., 2008; Bai et al., 2018; Wang and Xing, 2019; Piir et al., 2021; Wang et al., 2021). The two binary logistic regression models developed in this paper are therefore considered suitable for rapid identification of ERα and ERβ agonists among OH-PCB congeners. It is crucial to highlight that in QSAR studies, the modeling approach can be either predictive or descriptive, depending on the research goal (Gramatica, 2020). While predictive QSAR models aim to accurately predict the biological activities of untested compounds, descriptive QSAR models focus on understanding the relationship between structural features and biological activities of compounds within the training dataset without necessarily prioritizing predictive capabilities (Zefirov and Palyulin, 2001; Fujita and Winkler, 2016). Descriptive QSAR models mostly utilize interpretable molecular descriptors, whereas predictive QSAR models often use molecular descriptors that can be challenging to interpret (Zefirov and Palyulin, 2001; Fujita and Winkler, 2016). A limitation of the present study is the absence of explanation for the underlying structural features in OH-PCBs responsible for estrogen receptor binding. This limitation is due to lack of interpretability of the 2D autocorrelation descriptors selected for building the QSAR models reported in this paper. Nevertheless, the choice of these molecular descriptors is justified since the primarily focus of the present study was to develop predictive QSAR models for rapid identification of ERα and ERβ agonists among untested OH-PCB congeners.

Conclusion

Investigation of OH-PCBs in nuclear receptor-mediated toxicities in previous experimental studies was found to be time-consuming and resource-intensive. To expedite the identification of active nuclear receptor agonists and antagonists among OH-PCB congeners in future experimental endeavors, it is imperative to develop and apply classification-based QSAR models. In this study, two binary logistic regression models were successfully developed to predict active ERα and ERβ agonists among OH-PCB congeners using 2D autocorrelation descriptors as predictor variables. Through comprehensive internal and external validation procedures, the robustness, reliability and predictivity of the proposed QSAR models were established. The two classification-based QSAR models developed in this paper are considered suitable for rapid identification of active ERα and ERβ agonists among OH-PCB congeners, offering a promising approach for prioritizing OH-PCBs in toxicity testing and regulatory consideration.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

CRediT authorship contribution statement

Lukman K. Akinola: Conceptualization, Data curation, Writing (original draft), Writing (review & editing), Visualization, Investigation, Formal analysis, Methodology. **Adamu Uzairu:** Conceptualization, Data curation, Writing (review & editing), Formal analysis, methodology, Supervision. **Gideon A. Shallangwa:** Conceptualization, Data curation, Writing (review & editing), Formal analysis, Methodology, Supervision. **Stephen E. Abechi:** Conceptualization, Data curation, Writing (review & editing), Formal analysis, Methodology, Supervision. **Abdullahi B. Umar:** Conceptualization, Data curation, Writing (review & editing), Formal analysis, Methodology, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All the data used in the paper can be found within the paper and the Supplementary Material

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crtox.2024.100158.

References

Akinola, L.K., Uzairu, A., Shallangwa, G.A., Abechi, S.E., 2023. Development of binary classification models for grouping hydroxylated polychlorinated biphenyls into active and inactive thyroid hormone agonists. SAR QSAR Environ. Res. 34, 267–284. https://doi.org/10.1080/1062936X.2023.2207039.

Ambure, P., Aher, R.B., Gajewicz, A., Puzyn, T., Roy, K., 2015. "NanoBRIDGES" software: open access tools to perform QSAR and nano-QSAR modeling. Chemom. Intell. Lab. Syst. 147, 1–13. https://doi.org/10.1016/j.chemolab.2015.07.007.

Ambure, P., Halder, A.K., Diaz, H.G., Cordeiro, M.N.D.S., 2019. QSAR-co: An open source software for developing robust multitasking or multitarget classification-based QSAR models. J. Chem. Inf. Model. 59, 2538–2544. https://doi.org/10.1021/acs.jcim.9b00295.

Arulmozhiraja, S., Shiraishi, F., Okumura, T., Iida, M., Takigami, H., Edmonds, J.S., Morita, M., 2005. Structural requirements for the interaction of 91 hydroxylated polychlorinated biphenyls with estrogen and thyroid hormone receptors. Toxicol. Sci. 84, 49–62. https://doi.org/10.1093/toxsci/kfi063.

Bai, X., Yan, L., Ji, C., Zhang, Q., Dong, X., Chen, A., Zhao, M., 2018. A combination of ternary classification models and reporter gene assays for the comprehensive thyroid hormone disruption profiles of 209 polychlorinated biphenyls. Chemosphere 210, 312–319. https://doi.org/10.1016/j.chemosphere.2018.07.023.

Baker, V.A., 2001. Endocrine disrupters — testing strategies to assess human hazard. Toxicol. Vitro 15, 413–419. https://doi.org/10.1016/S0887-2333(01)00045-5.

Ballabio, D., Consonni, V., Mauri, A., Claeys-Bruno, M., Sergent, M., Todeschini, R., 2014. A novel variable reduction method adapted from space-filling designs. Chemom. Intell. Lab. Syst. 136, 147–154. https://doi.org/10.1016/j.chemolab.2014.05.010.

Ben-Jonathan, N., 2019. Endocrine disrupting chemicals and breast cancer: the saga of bisphenol A. In: Zhang, X. (Ed.), Estrogen Receptor and Breast Cancer. Springer, Switzerland, pp. 343–377. https://doi.org/10.1007/978-3-319-99350-8_13.

Bewick, V., Cheek, L., Ball, J., 2005. Statistics review 14: Logistic regression. Crit. Care 9, 112–118. https://doi.org/10.1186/cc3045.

Borja, J., Taleon, D.M., Auresenia, J., Gallardo, S., 2005. Polychlorinated biphenyls and their biodegradation. Process Biochem. 40, 1999–2013. https://doi.org/10.1016/j.procbio.2004.08.006.

Bourguignon, J.P., Parent, A.S., 2012. The impact of endocrine disruptors on female pubertal timing. In: Diamanti-Kandarakis, E., Gore, A.C. (Eds.), Endocrine Disruptors and Puberty. Springer, New York, pp. 325–337. https://doi.org/10.1007/978-1-60761-561-3_13.

Caserta, D., De Marco, M.P., Besharat, A.R., Costanzi, F., 2022. Endocrine disruptors and endometrial cancer: molecular mechanisms of action and clinical implications, a systematic review. Int. J. Mol. Sci. 23, 2956. https://doi.org/10.3390/ijms23062956.

Chighizola, C., Meroni, P.L., 2012. The role of environmental estrogens and autoimmunity. Autoimmun. Rev. 11, A493–A501. https://doi.org/10.1016/j.autrev.2011.11.027.

Darbre, P.D., Williams, G., 2015. Endocrine disruption and cancer of reproductive tissues. In: Darbre, P.D. (Ed.), Endocrine Disruption and Human Health. Elsevier, Amsterdam, pp. 177–200. https://doi.org/10.1016/B978-0-12-801139-3.00010-7.

Esposito, C., Landrum, G.A., Schneider, N., Stiefl, N., Riniker, S., 2021. GHOST: adjusting the decision threshold to handle imbalance data in machine learning. J. Chem. Inf. Model. 61, 2623–2640. https://doi.org/10.1021/acs.jcim.1c00160.

Fernández-González, R., Yebra-Pimentel, I., Martínez-Carballo, E., Simal-Gándara, J., 2015. A critical review about the human exposure to polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs) and polychlorinated biphenyls (PCBs) through foods. Crit. Rev. Food Sci. Nutr. 55 (11), 1590–1617. https://doi.org/10.1080/10408398.2012.710279.

Fujita, T., Winkler, D.A., 2016. Understanding the roles of the "two QSARs". J. Chem. Inf. Model. 56 (2), 269–274. https://doi.org/10.1021/acs.jcim.5b00229.

Goeman, J.J., le Cessie, S., 2006. A goodness-of-fit test for multinomial logistic regression. Biometrics 62, 980–985. https://doi.org/10.1111/j.1541-0420.2006.00581.x.

Gramatica, P., 2020. Principles of QSAR modeling: comments and suggestions from personal experience. Int. J. Quant. Struct. Prop. Relatsh. 5 (3), 61–97. https://doi.org/10.4018/IJQSPR.20200701.oa1.

Hancock, J., Johnson, J.M., Khoshgoftaar, T.M. (2022) A comparative approach to threshold optimization for classifying imbalance data. *IEEE 8th International Conference on Collaboration and Internet Computing (CIC)*, Atlanta, GA, USA, pp. 135–142. https://doi.org/10.1109/CIC56439.2022.00028.

He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21 (9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239.

Hosmer, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression. John Wiley & Sons Inc, Hoboken, pp. 10–15.

Kamata, R., Nakajima, D., Shiraishi, F., 2019. Measurement of the agonistic activities of monohydroxylated polychlorinated biphenyls at the retinoid X and retinoic acid receptors using recombinant yeast cells. Toxicol. Vitro 57, 9–17. https://doi.org/10.1016/j.tiv.2019.01.022.

Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137–148. https://doi.org/10.1080/00401706.1969.10490666.

Lallas, P.L., 2001. The Stockholm convention on persistent organic pollutants. Am. J. Int. Law 95 (3), 692–708. https://doi.org/10.2307/2668517.

Lang, A., Volkamer, A., Behm, L., Roblitz, S., Ehrig, R., Schneider, M., Geris, L., Wichard, J., Buttgereit, F., 2018. *In silico* methods—computational alternative to animal testing. ALTEX—Alternat. Anim. Exp. 35 (1), 126–128. https://doi.org/10.14573/altex.1712031.

Li, J., McMurray, R.W., 2009. Effects of chronic exposure to DDT and TCDD on disease activity in murine systemic lupus erythematosus. Lupus 18, 941–949. https://doi.org/10.1177/0961203309104431.

Liu, H., Papa, E., Walker, J.D., Gramatica, P., 2007. *In silico* screening of estrogen-like chemicals based on different nonlinear classification models. J. Mol. Graph. Model. 26, 135–144. https://doi.org/10.1016/j.jmgm.2007.01.003.

Mallozzi, M., Leone, C., Manurita, F., Bellati, F., Caserta, D., 2017. Endocrine disrupting chemicals and endometrial cancer: an overview of recent laboratory evidence and epidemiological studies. Int. J. Environ. Res. Public Health 14, 334. https://doi.org/10.3390/ijerph14030334.

Martin, T.M., Harten, P., Young, D.M., Muratov, E.N., Golbraikh, A., Zhu, H., Tropsha, A., 2012. Does rational selection of training and test sets improve the outcome of QSAR modeling? J. Chem. Inf. Model. 52, 2570–2578. https://doi.org/10.1021/ci300338w.

Pentyala, S.N., Rebecchi, M., Mishra, S., Rahman, A., Stefan, R., Rebecchi, J., Kodavanti, P.R.S., 2011. Polychlorinated biphenyls: in situ bioremediation from the environment. In: Reddy, G.R., Flora, S.J.F., Basha, R.M. (Eds.), Environmental Pollution: Ecology and Human Health. Narosa Publishing House, New Delhi, pp. 249–262.

Piir, G., Sild, S., Maran, U., 2021. Binary and multi-class classification for androgen receptor agonists, antagonists and binders. Chemosphere 262, 128313. https://doi.org/10.1016/j.chemosphere.2020.128313.

Rogers, D., Hopfinger, A.J., 1994. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J. Chem. Inf. Comput. Sci. 34, 854–866. https://doi.org/10.1021/ci00020a020.

Roncaglioni, A., Piclin, N., Pintore, M., Benfanati, E., 2008. Binary classification models for endocrine disrupter effects mediated through the estrogen receptor. SAR QSAR Environ. Res. 19, 697–733. https://doi.org/10.1080/10629360802550606.

Sakkiah, S., Kusko, R., Tong, W., Hong, H. (2019) Applications of molecular dynamics simulations in computational toxicology. In: Hong H (ed.) *Advances in Computational Toxicology: Methodologies and Applications in Regulatory Science*. Springer Nature, Switzerland, pp. 181–212. https://doi.org/10.1007/978-3-030-16443-0_10.

Sapra, R.L., 2014. Using R2 with caution. Curr. Med. Res. Pract. 4, 130–134. https://doi.org/10.1016/j.cmrp.2014.06.002.

Sarma, K.V.S., Vardhan, R.V., 2019. Multivariate Statistics Made Simple: A Practical Approach. Taylor & Francis Group, Boca Raton, pp. 169–184.

Schisterman, E.F., Perkins, N.J., Liu, A., Bondell, H., 2005. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. Epidemiology 16, 73–81. https://doi.org/10.1097/01.ede.0000147512.81966.ba.

Shanle, E.K., Xu, W., 2011. Endocrine disrupting chemicals targeting estrogen receptor signaling: Identification and mechanisms of action. Chem. Res. Toxicol. 24, 6–19. https://doi.org/10.1021/tx100231n.

Shao, Y., Molnar, L.F., Jung, Y., Kussmann, J., Ochsenfeld, C., Brown, S.T., Gilbert, A.T., Slipchenko, L.V., Levchenko, S.V., O'Neill, D.P., DiStasio Jr, R.A., Lochan, R.C., Wang, T., Beran, G.J.O., Besley, N.A., Herbert, J.M., Lin, C.Y., van Voorhis, T., Chien, S.H., Sodt, A., Steele, R.P., Rassolov, V.A., Maslen, P.E., Korambath, P.P., Adamson, R.D., Austin, B., Baker, J., Byrd, E.F.C., Dachsel, H., Doerksen, R.J., Dreuw, A., Dunietz, B.D., Dutoi, A.D., Furlani, T.R., Gwaltney, S.R., Heyden, A., Hirata, S., Hsu, C.P., Kedziora, G., Khalliulin, R.Z., Klunzinger, P., Lee, A.M., Lee, M.S., Liang, W., Lotan, I., Nair, N., Peters, B., Proynov, E.I., Pieniazek, P.A., Rhee, Y.M., Ritchie, J., Rosta, E., Sherrill, C.D., Simmonett, A.C., Subotnik, J.E., Woodcock, H.L., Zhang, W., Bell, A.T., Chakraborty, A.K., Chipman, D.M., Keil, F.J., Warshel, A., Hehre, W.J., Schaefer, H.F., Kong, J., Krylov, A.I., Gill, P.M.W., Head-Gordon, M., 2006. Advances in methods and algorithms in a modern quantum chemistry program package. Phys. Chem. Chem. Phys. 8, 3172–3191. https://doi.org/10.1039/b517914a.

Shukla, S.J., Huang, R., Austin, C.P., Xia, M., 2010. The future of toxicity testing: a focus on *in vitro* methods using a quantitative high-throughput screening platform. Drug Discov. Today 15, 997–1007. https://doi.org/10.1016/j.drudis.2010.07.007.

Snarey, M., Terrett, N.K., Willet, P., Wilton, D.J., 1997. Comparison of algorithms for dissimilarity-based compound selection. J. Mol. Graph. Model. 15, 372–385. https://doi.org/10.1016/S1093-3263(98)00008-4.

Sperandei, S., 2014. Understanding logistic regression analysis. Biochem. Med. 24, 12–18. https://doi.org/10.11613/BM.2014.003.

Stoltzfus, J.C., 2011. Logistic regression: A brief primer. Acad. Emerg. Med. 18, 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x.

Takeuchi, S., Shiraishi, F., Kitamura, S., Kuroki, H., Jin, K., Kojima, H., 2011. Characterization of steroid hormone receptor activities in 100 hydroxylated polychlorinated biphenyls, including congeners identified in humans. Toxicology 289, 112–121. https://doi.org/10.1016/j.tox.2011.08.001.

Tam, N., Lai, K.P., Kong, R.Y.C., 2022. Comparative transcriptomic analysis reveals reproductive impairments caused by PCBs and OH-PCBs through the dysregulation of ER and AR signaling. Sci. Total Environ. 802, 149913 https://doi.org/10.1016/j.scitotenv.2021.149913.

Tehrani, R., Aken, B.V., 2014. Hydroxylated polychlorinated biphenyls in the environment: Sources, fate, and toxicity. Environ. Sci. Pollut. Res. Int. 21, 6334–6345. https://doi.org/10.1007/s11356-013-1742-6.

Toporova, L., Balaguer, P., 2020. Nuclear receptors are the major targets of endocrine disrupting chemicals. Mol. Cell. Endocrinol. 502, 110665 https://doi.org/10.1016/j.mce.2019.110665.

Tukker, A.M., de Groot, M.W.G.D.M., Wijnolts, F.M.J., Kasteel, E.E.J., Hondebrink, L., Westerink, R.H.S., 2016. Is the time right for *in vitro* neurotoxicity testing using human iPSC-derived neurons? ALTEX—Alternat. Anim. Exp. 33 (3), 261–271. https://doi.org/10.14573/altex.1510091.

Wang, J., Huang, Y., Wang, S., Yang, Y., He, J., Li, C., Zhao, Y.H., Martyniuk, C.J., 2021. Identification of active and inactive agonists/antagonists of estrogen receptor based on Tox21 10K compounds library: binomial analysis and structural alert. Ecotoxicol. Environ. Saf. 214, 112114 https://doi.org/10.1016/j.ecoenv.2021.112114.

Wang, F., Xing, J., 2019. Classification of thyroid hormone agonists and antagonists using statistical learning approaches. Mol. Divers. 23, 85–92. https://doi.org/10.1007/s11030-018-9857-9.

Warmuth, A., Ohno, K., 2013. The PCBs elimination network: the information exchange platform created for the risk reduction of polychlorinated biphenyls (PCBs). J. Epidemiol. Community Health 67 (1), 4–5. https://doi.org/10.1136/jech-2012-201025.

Yadav, S., Shukla, S. (2016) Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. *IEEE 6th International Conference on Advanced Computing (IACC)*, Bhimavaram, India, pp. 78–83. https://doi.org/10.1109/IACC.2016.25.

Yap, C.W., 2011. PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 32, 466–1474. https://doi.org/10.1002/jcc.21707.

Zefirov, N.S., Palyulin, V.A., 2001. QSAR for boiling points of "small" sulfides. Are the "high-quality structure-property-activity regressions" the real high quality QSAR models? J. Chem. Inf. Comput. Sci. 41, 1022–1027. https://doi.org/10.1021/ci0001637.

Zhu, C., Wang, P., Li, Y., Chen, Z., Li, W., Ssebugere, P., Zhang, Q., Jiang, G., 2015. Bioconcentration and trophic transfer of polychlorinated biphenyls and polychlorinated dibenzo-*p*-dioxins and dibenzofurans in aquatic animals from an e-waste dismantling area in East China. Environ. Sci. Processes Impacts 17, 693–699. https://doi.org/10.1039/c5em00028a.

Zou, Q., Xie, S., Lin, Z., Wu, M., Ju, Y., 2016. Finding the best classification threshold in imbalanced classification. Big Data Res. 5, 2–8. https://doi.org/10.1016/j.bdr.2015.12.001.