	QMRF identifier: TOXFENCE-QMRF-005a
	QMRF Title: GCN Model for Developmental/Reproduction Toxicity (DART) TOXFENCE model – version 1.0
	Date of QMRF: 25/03/2026
	Model Developer: RAMC Co., Ltd.

1. QSAR identifier

1.1 QSAR identifier (title)

GCN Model for Developmental/Reproduction Toxicity (DART) TOXFENCE model – version 1.0

1.2 Other related models

No related models identified

1.3 Software coding the model

TOXFENCE v1.0

The model is implemented in TOXFENCE, a web-based SaaS platform developed by Risk Management Consulting Co., Ltd. TOXFENCE is designed to perform QSAR-based toxicity prediction using chemical structure information as input. The software is provided as an online service without local installation and includes functions for model execution, result review, and report output. The model was implemented in a Python-based environment, the backend service is operated using FastAPI, and RDKit was used for molecular structure handling and descriptor generation.

Risk Management Consulting Co., Ltd.
https://www.toxfence.com

2 General information

2.1 Date of QMRF

Feb 2025

2.2 QMRF author(s) and contact details

[1] Organisation: Risk Assessment & Management Consulting 04156 Seoul, Korea

[2] Contact e-mail: ramc0983@naver.com

[3] Corporate website: https://www.ramc0983.com/

[4] TOXFENCE web service: https://www.toxfence.com/

2.3 Date of QMRF update(s)

25/03/2026

2.4 QMRF update(s)

Updated by: Risk Assessment & Management Consulting Co., Ltd.

Contact: ramc0983@naver.com

Modified field: 1.3 Software coding the model

Reason for update: The website information was revised to replace the service address with the company website address.

2.5 Model developer(s) and contact details

The model was developed by Risk Assessment & Management Consulting Co., Ltd., 04156 Seoul, Republic of Korea.

Contact e-mail: ramc0983@naver.com

Website: https://www.ramc0983.com/

TOXFENCE web service: https://www.toxfence.com/

2.6 Date of model development and/or publication

2025

2.7 Reference(s) to main scientific papers and/or software package

Prediction of reproductive and developmental toxicity using an attention and gate augmented graph convolutional network

software package

1) Python 3.10.4

2) PyTorch2.2

3) PyTorch Geometric 2.5

4) RDkit 2023.09

2.8 Availability of information about the model

The model is proprietary. The training dataset and the model are not publicly available. The model is implemented and operated through the TOXFENCE web service.

https://www.toxfence.com/

2.9 Availability of another QMRF for exactly the same model

Another QMRF is not available.

3 Defining the endpoint - OECD Principle 1

3.1 Species

The underlying experimental data included various mammalian species, primarily rats and rabbits, used in studies relevant to GHS reproductive and developmental toxicity classification.

3.2 Endpoint

Reproductive and Developmental Toxicity (GHS classification)

3.3 Comment on endpoint

Reproductive and developmental toxicity refers to the ability of a substance to adversely affect sexual function and fertility in adult males and females and/or the normal development of offspring. For this model, the endpoint is aligned with the GHS classification concept for reproductive toxicity, which includes adverse effects on reproductive performance as well as developmental effects in the conceptus or offspring. This endpoint is generally informed by in vivo mammalian studies, since reproduction and development involve complex whole-organism processes and may be influenced by both the parent substance and its metabolites [1].

3.4 Endpoint units

Dimensionless

3.5 Dependent variable

The dependent variable was defined as positive (1) or negative (0) for reproductive and developmental toxicity according to the GHS classification criteria.

3.6 Experimental protocol

The underlying experimental data were derived primarily from in vivo mammalian reproductive and developmental toxicity studies conducted according to internationally recognised test guidelines, including OECD TG 414, TG 421/422, and TG 443. These studies assess reproductive function, fertility, prenatal development, and postnatal developmental outcomes. In particular, TG 414 addresses prenatal developmental toxicity, typically in rats and rabbits, whereas TG 421/422 provide screening-level information on reproductive and developmental toxicity, and TG 443 provides more comprehensive information on reproductive performance and offspring development, with optional cohorts for developmental neurotoxicity and immunotoxicity [2] [3].

3.7 Endpoint data quality and variability

Experimental data for this model were curated from publicly accessible regulatory and hazard-classification records available through NITE (Japan), HCIS (Australia), and studies accessible through eChemPortal. Only records with reliability scores of 1 or 2 were retained for model development. Where available, source records referred to guideline-compliant in vivo studies relevant to reproductive and developmental toxicity, including OECD TG 414, TG 421/422, and TG 443. During data curation, duplicate records were excluded, and conflicting classifications reported for the same substance were resolved through expert consultation. For each retained substance, the CAS number, SMILES, and a final GHS-based label were recorded, encoded as 1 for substances classified as Repr. 1A, 1B, or 2 and/or effects on or via lactation, and 0 for substances not classified for this endpoint. SMILES were standardised using a custom Python 3.10.4 script based on the RDKit library. After curation, the final dataset consisted of 5,861 organic compounds. As the data were compiled from multiple regulatory and study sources, some inter-source variability may remain due to differences in study design, species, dosing conditions, endpoint interpretation, and reporting practices.

4 Defining the algorithm - OECD Principle 2

4.1 Type of model

A model based on graph neural networks that utilize molecular structure to predict chemical toxicity as a binary outcome.

4.2 Explicit algorithm

The model is an attention-based graph neural network for binary classification. SMILES strings are converted into molecular graphs using RDKit. Atom-level categorical features are embedded and combined with additional atomic and structural alert features to form node representations. The molecular graph is processed through two attention-based graph convolution blocks with multi-head neighbourhood aggregation and gated residual connections. A graph-level representation is obtained by global mean pooling and passed to a multilayer perceptron to generate a single output logit. The output is transformed into a probability using a sigmoid function, and the final prediction is assigned as positive or negative according to the decision threshold.

4.3 Descriptors in the model

The model uses graph-based descriptors derived from molecular structure. Node features include atom type (52 categories), degree (11 categories), number of hydrogens (6 categories), hybridization state (6 categories), formal charge, number of radical electrons, aromaticity, and SMARTS-based structural alert bits. Molecular connectivity is represented as a graph structure and used during message passing.

4.4 Descriptor selection

No explicit feature selection was performed. The model used predefined graph-based structural features derived from molecular structure, including atom-level descriptors and structural alert information. Feature relevance was learned implicitly during model training.

4.5 Algorithm and descriptor generation

The model was implemented in Python using PyTorch and PyTorch Geometric. Chemical structures were provided as SMILES strings and converted into molecular graphs using RDKit. Node descriptors included atom type, degree, number of hydrogens, hybridization, as well as formal charge, radical electrons, aromaticity flag, and SMARTS-based structural alert bits. Bond information was represented through the molecular graph structure during graph construction.

4.6 Software name and version for descriptor generation

Python 3.10.4. (https://www.python.org/)

PyTorch 2.2 (https://pytorch.org/)

PyTorch Geometric 2.5 (https://pyg.org/)

RDKit 2023.09 (https://www.rdkit.org/)

4.7 Chemicals/Descriptors ratio

2,214 chemicals / 55-dimensional input feature vector= 40.3

5 Defining the applicability domain - OECD Principle 3

5.1 Description of the applicability domain of the model

The Applicability Domain (AD) is evaluated in the GCN-derived embedding space. After L2 normalization, AD is determined by combining a global Mahalanobis distance squared (MD2) criterion (Ledoit–Wolf shrinkage covariance) and a local k-nearest neighbor (k-NN) radius criterion. 95th/99th percentile cut-offs from the reference embedding distribution are used to assign IN/BORDER/OUT.

5.2 Method used to assess the applicability domain

The applicability domain is assessed in the L2-normalized GCN embedding space using two complementary criteria. First, a global Mahalanobis distance-squared (MD2) is computed using a Ledoit–Wolf shrinkage covariance model fitted on the reference embedding set, and 95th/99th percentile cut-offs of the MD2 distribution are used. Second, a local k-nearest-neighbor (k-NN) radius is calculated as the k-th nearest Euclidean distance from the query to the reference embeddings, with 95th/99th percentile cut-offs derived from the reference k-NN radius distribution (estimated from a random subsample for efficiency). The final AD classification is reported as IN (both criteria within the 95th percentile), BORDER (one criterion between the 95th and 99th percentile), or OUT (otherwise).

5.3 Software name and version for applicability domain assessment

Applicability domain assessment was implemented internally in TOXFENCE v1.0, using an embedding-based workflow integrated into the model evaluation pipeline.

5.4 Limits of applicability

This model can be applied to both organic and inorganic chemicals; substances containing elements beyond C, O, N, S, P, Cl, Br, F, and I were not excluded. Salts were not neutralized or stripped prior to prediction; outputs were generated on the forms provided. Nevertheless, chemotypes underrepresented in the training data may carry higher uncertainty, so the applicability domain (AD) result should be consulted alongside the prediction.

6 Internal validation - OECD Principle 4

6.1 Availability of the training set

No

6.2 Available information for the training set

CAS RN: Yes

Chemical Name: Yes

Smiles: Yes

Formula:No

INChI: No

MOL file: Yes

6.3 Data for each descriptor variable for the training set

All

6.4 Data for the dependent variable for the training set

All

6.5 Other information about the training set

After exclusion of invalid structures during graph construction, a total of 4,719 substances were available for model development, including 2,214 positive and 2,505 negative substances. These data were subsequently divided into training and test sets using a stratified 80:20 split. The compounds used for model development consisted of curated organic substances with reliable binary classifications suitable for graph-based modelling.

6.6 Pre-processing of data before modelling

Before modelling, duplicate records were removed during data curation. In addition, SMILES entries with missing values were excluded, chemical structures were verified using RDKit, and invalid structures that could not be converted into molecular graphs were discarded. The remaining valid structures were then transformed into graph-based representations for model development.

6.7 Statistics for goodness-of-fit

Accuracy = 80%, Precision = 82%

Recall = 75%, Specificity = 86%

F1-score = 78%, MCC = 61%

TP 1294, TN 2505, FP 438, FN 482

6.8 Robustness - Statistics obtained by leave-one-out cross-validation

6.9 Robustness - Statistics obtained by leave-many-out cross-validation

6.10 Robustness - Statistics obtained by Y-scrambling

6.11 Robustness - Statistics obtained by bootstrap

6.12 Robustness - Statistics obtained by other methods

7 External validation - OECD Principle 4

7.1 Availability of the external validation set

No

7.2 Available information for the external validation set

CAS RN: Yes

Chemical Name: Yes

Smiles: Yes

Formula: No

INChI: No

MOL file: Yes

7.3 Data for each descriptor variable for the external validation set

All

7.4 Data for the dependent variable for the external validation set

All

7.5 Other information about the external validation set

The test set is composed of 1,180 substances (553 positive, 627negative)

7.6 Experimental design of test set

Prior to model development, the full dataset was randomly divided into a training set and a test set at a ratio of 8:2.

7.7 Predictivity - Statistics obtained by external validation

Accuracy = 74%, Precision = 68%

Recall = 83%, Specificity = 66%

F1-score = 75%, MCC = 51%

TP 459, TN 414, FP 94, FN 213

7.8 Predictivity - Assessment of the external validation set

The external validation set was considered appropriate for assessing model predictivity because it was sufficiently large and was generated by random splitting prior to model development. The test set was assumed to be broadly representative of the descriptor space and response distribution of the training set. Therefore, the external validation results were considered suitable for evaluating the predictive performance of the model within its applicability domain.

7.9 Comments on the external validation of the model

The external validation was performed using a hold-out test set obtained by random splitting of the full dataset prior to model development. Therefore, the validation can be regarded as an independent test of the final model, although the test compounds originated from the same overall dataset rather than from a completely separate external source. The reported performance should thus be interpreted as evidence of predictive ability within the chemical space represented by the dataset, while predictions for compounds outside the applicability domain or for underrepresented chemotypes should be interpreted with caution.

8 Providing a mechanistic interpretation - OECD Principle 5

8.1 Mechanistic basis of the model

The model determines genotoxicity by learning directly from molecular graph representations. Canonical SMILES are converted into molecular graphs with RDKit, where atoms serve as nodes and bonds as edges. Node-level descriptors include atom type, degree, number of hydrogens, hybridization state, formal charge, radical electrons, aromaticity flag, and toxicophore alert bits. Edge-level descriptors such as bond type, conjugation, and aromaticity are also encoded. These graph-structured features are processed through a graph neural network implemented in PyTorch Geometric, which applies message passing to capture both local atomic environments and higher-order structural relationships. By integrating these chemically meaningful descriptors in the GNN framework, the model captures mechanistic patterns relevant to genotoxic potential.

8.2 Other information about the mechanistic interpretation

The mechanistic interpretation of the model was established a posteriori, based on the interpretation of the final set of descriptors and model outputs after model development.

8.3 Other information about the mechanistic interpretation

No additional information is available regarding the mechanistic interpretation.

9 Miscellaneous information

9.1 Comments

No additional comments.

9.2 Bibliography

[1] ECHA/REACH Regulation – Substances of Very High Concern (SVHC): Defines SVHC as chemicals with hazards such as carcinogenicity, mutagenicity, or reproductive toxicity, which are subject to strict regulatory control.

[2][3] OECD Test Guideline 443, Extended One-Generation Reproductive Toxicity Study (adopted 2011, updated 2025): Outline of the study design (cohorts for reproductive/developmental endpoints, including optional second generation, developmental neurotoxicity and immunotoxicity assessments).

9.3 Supporting information

Not available.