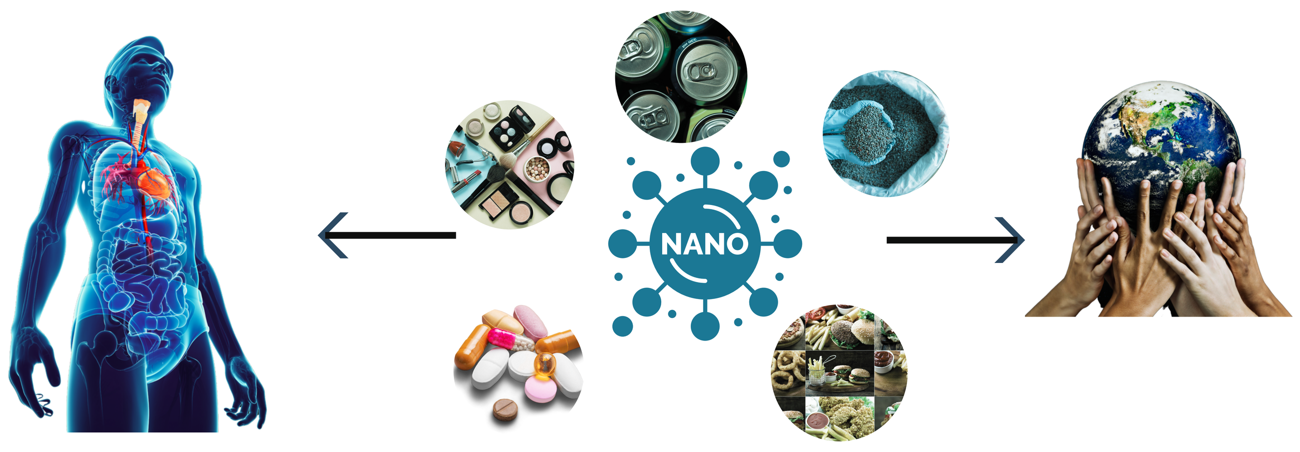
IN SILICO NANO-QSAR TOXICITY CLASSIFICATION BY SUPPORT VECTOR MACHINE

Kryzze Lee T. Serot, John Riz V. Bagnol, Farlley G. Bondaug, Phoebe A. Sombilon, Irelie P. Ebardo, Gabriel C. Engcong, Aries P. Valeriano, Rey Y. Capangpangan

Department of Mathematics and Statistics, University of Southeastern Philippines
REY Laboratories, Research Division, Mindanao State University at Naawan

BACKGROUND

Nanomaterials are materials which consist of particle size ranging between 1-100 nanometer and have been engineered to produce sustainable products on different fields (Klaine et. al, 2008).The increasing usage of ENM's, increases the risk of exposure to nanoparticles as well, which lead to various studies regarding their toxicity and the negative impacts to various organisms (Zoroddu et.al., 2014).An efficient computational method for determining the relationship between a nanoparticle's properties and activity is quantitative structure-activity relationship (QSAR) or Nano-QSAR. Experimental studies are also carried out to assess the toxicity of nanomaterials, specifically *in vivo*, *in vitro and in silico* tests. Furthermore, *In silico* approaches are more time-efficient, less costly, and less labor intensive in their assessment of nanoparticle toxicity compared with the conventional approach. Support Vector Machine, a machine learning approach for Nano-QSAR, modelling had been effective for binary classification for both linear and non-linear data.



Disclaimer : Photo belongs to the rightful owner. No copyright infringement intended. Created@canva.com

OBJECTIVES AND METHODOLOGY

Objectives of the study are:

- Find SVM optimal parameters
- Validate Nano-QSAR SVM model.
- Determine leading nanoparticle toxicity predictors.

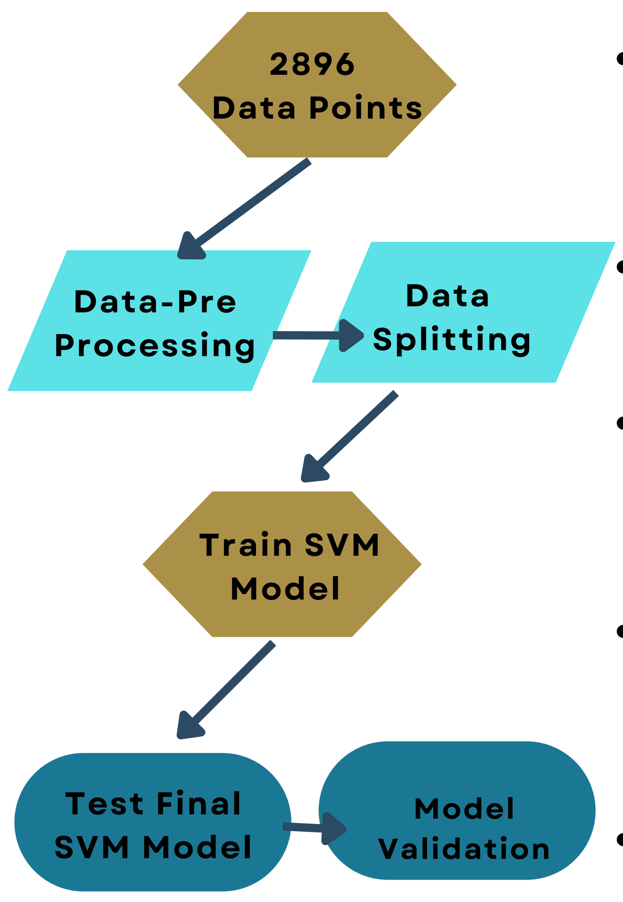


Figure 1 : SVM Model Building

- 2896 nanoparticles are utilized with 8 features and cell viability as endpoint.
- DATA-PREPROCESSING. Setting Endpoint (cell viability to binary (0,1) toxic (≤50%) and non-toxic (>50%) nanoparticles.
- DATA SPLITTING. 80:20 proportion for training and external validation Set.
- TRAIN SVM MODEL. Use of grid search for hyperparameter tuning with 10 fold cross validation.
- TESTING FINAL SVM MODEL. Testing of SVM model with optimal parameter on external validation set.
- MODEL VALIDATION. Assessing Nano-QSAR SVM model with grid search with its classification performance.

ACKNOWLEDGEMENTS

Funding Agency : DOST-PCIEERD

Lead Implementing Agency : MSU-Naawan

Co-Implementing Agency : USeP-Obrero

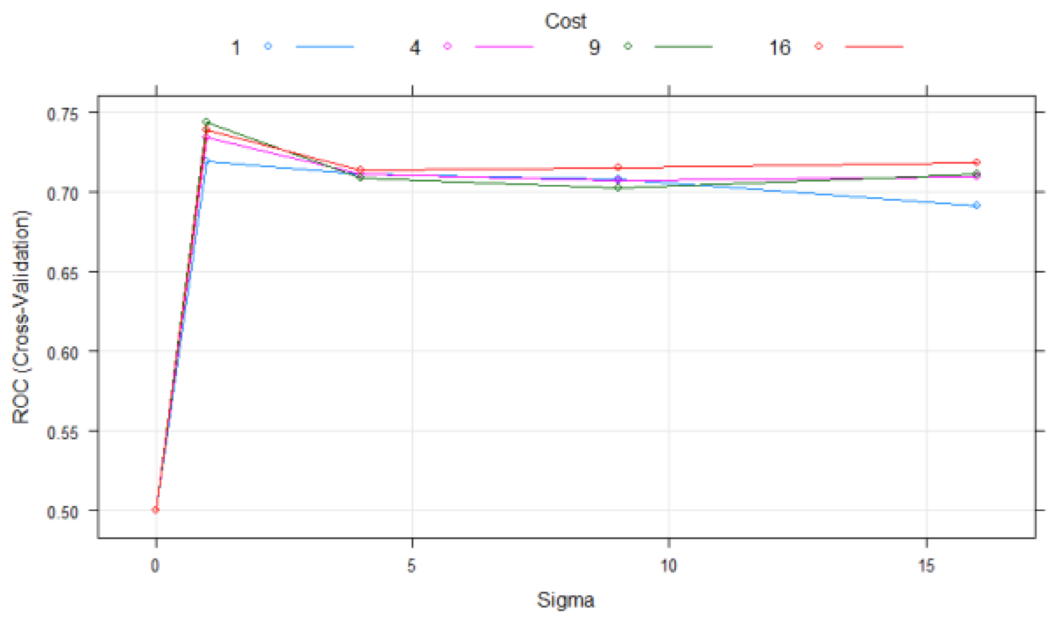
RESULTS

SVM OPTIMAL PARAMETERS

Table 1 : SVM Parameter Tuning By Grid Search

sigma	C	ROC	Sens	Spec
0	1	0.5000000	1.0000000	0.0000000
0	4	0.5000000	1.0000000	0.0000000
0	9	0.5000000	1.0000000	0.0000000
0	16	0.5000000	1.0000000	0.0000000
1	1	0.7162208	0.9666121	0.2581818
1	4	0.7087571	0.9507672	0.2745455
1	9	0.7265564	0.9468220	0.2981818
1	16	0.7223991	0.9411627	0.3000000
4	1	0.7047536	0.9524718	0.2709091
4	4	0.7024045	0.9411595	0.2981818
4	9	0.7066148	0.9377664	0.3200000
4	16	0.7106178	0.9355065	0.3200000
9	1	0.6796504	0.9530303	0.2872727
9	4	0.6814451	0.9377696	0.3236364
9	9	0.7079296	0.9309739	0.3381818
9	16	0.7033024	0.9292790	0.3509091
16	1	0.7059340	0.9519036	0.2945455
16	4	0.7074357	0.9321103	0.3327273
16	9	0.6962141	0.9275809	0.3618182
16	16	0.7112065	0.9253082	0.3763636

Figure 1 : Plot of SVM Parameter Tuning By Grid Search

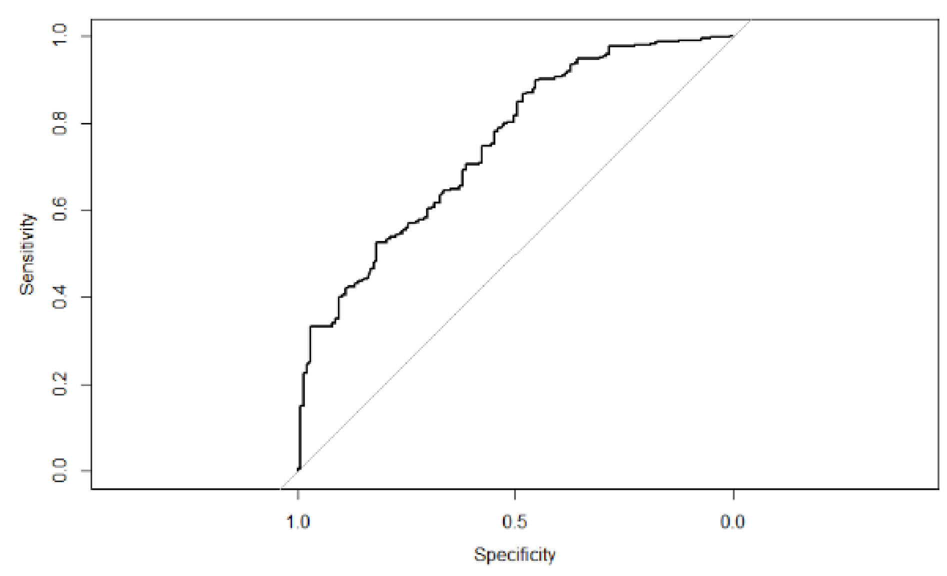


SVM CLASSIFICATION MODEL VALIDATION

Table 2 : SVM Classification Performance

SVM Metrics	
Accuracy	0.801
Sensitivity	0.9388
Specificity	0.3577
PPV	0.8247
NPV	0.7630
Recall	0.8247
Precision	0.9388
F1-Score	0.8780
AUC	0.7484

Figure 2 : SVM ROC Curve



SVM CLASSIFICATION MODEL VALIDATION & LEADING TOXICITY DESCRIPTORS

Figure 3 : Plot for Classification of NP's

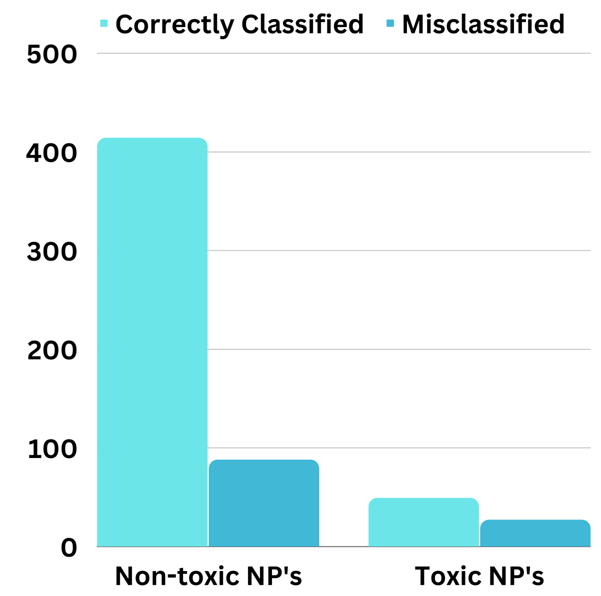
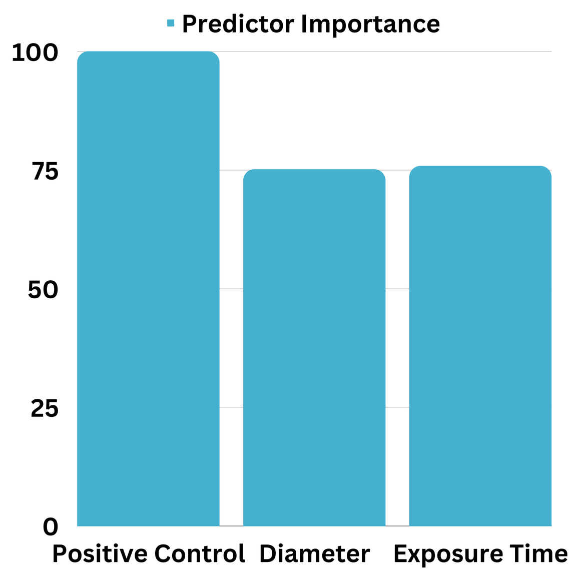


Table 3 : Confusion Matrix

	Non-Toxic	Toxic
Non-toxic	414	88
Toxic	27	49

Figure 4 : Leading Toxicity Predictors



CONCLUSION

The final Nano-QSAR model via Support Vector Machine with Radial Basis Function (RBF) as kernel and using Grid Search for setting optimal parameters *C* = 9, *sigma*=1 generated a classification model that is efficient in classifying nanoparticles as toxic or non-toxic using 9 features (type, size, diameter, concentration, cell line, exposure time, colloidal stability, positive control and cell type).An overall accuracy of 80.10% implies a good classification model. However, considering the imbalances of dataset F1-Score metric can provide the general model classification performance which yielded a value of 87.80% with leading toxicity predictors-positive control, diameter, exposure time. Hence the developed Nano-QSAR SVM model with grid search shows a good performance on classifying nanoparticles as either toxic or non-toxic.

REFERRENCES

[1] Klaine SJ,et.al.Nanomaterials in the environment: behavior, fate, bioavailability, and effects. Environ Toxicol Chem. 2008 Sep;27(9):1825-51. doi: 10.1897/08-090.1. Erratum in: Environ Toxicol Chem. 2012 Dec;31(12):2893. PMID: 19086204.

[2] Zoroddu MA,et.al. Toxicity of nanoparticles. Curr Med Chem. 2014;21(33):3837-53. doi: 10.2174/0929867321666140601162314. PMID: 25306903.