Soft Computing (2024) 28:10235–10261
https://doi.org/10.1007/s00500-024-09774-0

APPLICATION OF SOFT COMPUTING

Predicting the potential toxicity of the metal oxide nanoparticles using machine learning algorithms

Gehad Ismail Sayed[1,2] · Heba Alshater[2,3] · Aboul Ella Hassanien[2,4,5]

Accepted: 20 February 2024 / Published online: 10 July 2024

Why model nanoparticle toxicity?

- MexOy NPs have industrial, biomedical, and environmental applications.
- However, there is evidence that they cause **oxidative stress**, **apoptosis**, **lung damage**, and more.
- Experimental methods are expensive and time-consuming.

What did they do in this job?

They developed a predictive toxicity model for metal oxide nanoparticles (MexOy NPs) using machine learning, with an emphasis on three aspects:

They built a dataset with 79 descriptors

They applied intelligent pre processing tools: SMOTE, DBSCAN, EBSTSA

They trained an SVM classifier with k-fold cross-validation, achieving:
F1-score = 99.47%
Sensitivity = 98.87%
Specificity = 100%

Dataset used

- Characterization of the 24 nanoparticles
    - 79 descriptors: physicochemical, structural, and electronic.
    - Toxicity measurement: ATP and LDH assays in cell lines.

Table 1 Attributes definition

Att.	Meaning	Att.	Meaning
A1	Material type	A41	O 4rdN avg core (eV)
A2	Core size (nm)	A42	O 4rdN avg shell (eV)
A3	Hydro size (nm)	A43	NP diameter (A)
A4	Surface charge (mV)	A44	NP surface area(A^2)
A5	Surface area (m2/g)	A45	NP volume(A^3)
A6	Method surface area	A46	Lattice energy of NP (eV)
A7	Hsf (eV)	A47	Relative lattice energy of NP to bulk material (E L bulk-E L NP) (eV)
A8	Ec (eV)	A48	Lattice energy of NP / d NP (eV/A)
A9	Ev (eV)	A49	Lattice energy of NP /S NP (eV/A^2)
A10	MeO (eV)	A50	Lattice energy of NP /V NP (eV/A^3)
A11	Assay	A51	Force vector length avg all (eV)
A12	Cell name	A52	Force vector length avg core (eV)
A13	Cell species	A53	Force vector length avg shell (eV)
A14	Cell origin	A54	Al force vector length 4rdN avg all (eV)
A15	Cell type	A55	Al force vector length 4rdN avg core (eV)
A16	Exposure dose (ug/mL)	A56	Al force vector length 4rdN avg shell (eV)
A17	log(n atoms all)	A57	O force vector length avg all (eV)
A18	log(n atoms core)	A58	O force vector length avg core (eV)
A19	log(n atoms shell)	A59	O force vector length avg shell (eV)
A20	log(n Al atoms all)	A60	Force vector surface 2 component avg all (eV)
A21	log(n Al atoms core)	A61	Force vector surface 2 component avg core (eV)
A22	log(n Al atoms shell)	A62	Force vector surface 2 component avg shell (eV)
A23	log(n O atoms all)	A63	Al force vector surface 2 component 4rdN avg all (eV)
A24	log(n O atoms core)	A64	Al force vector surface 2 component 4rdN avg core (eV)
A25	log(n O atoms shell)	A65	Al force vector surface 2 component 4rdN avg shell (eV)
A26	peng avg all (eV)	A66	O force vector surface 2 component avg all (eV)
A27	peng avg core (eV)	A67	O force vector surface 2 component avg core (eV)
A28	peng avg shell (eV)	A68	O force vector surface 2 component avg shell (eV)
A29	Al peng avg all (eV)	A69	Force vector surface tangent component avg all (eV)
A30	Al peng avg core (eV)	A70	Force vector surface tangent component avg core (eV)
A31	Al peng avg shell (eV)	A71	Force vector surface tangent component avg shell (eV)
A32	O peng avg all (eV)	A72	Al force vector surface tangent component 4rdN avg all (eV)
A33	O peng avg core (eV)	A73	Al force vector surface tangent component 4rdN avg core (eV)
A34	O peng avg shell (eV)	A74	Al force vector surface tangent component 4rdN avg shell (eV)
A35	4rdN avg all (eV)	A75	O force vector surface tangent component avg all (eV)
A36	4rdN avg core (eV)	A76	O force vector surface tangent component avg core (eV)
A37	4rdN avg shell (eV)	A77	O force vector surface tangent component avg shell (eV)
A38	Al 4rdN avg all (eV)	A78	Viability (%)
A39	Al 4rdN avg core (eV)	A79	1ity
A40	Al 4rdN avg shell (eV)		

Methodological pipeline

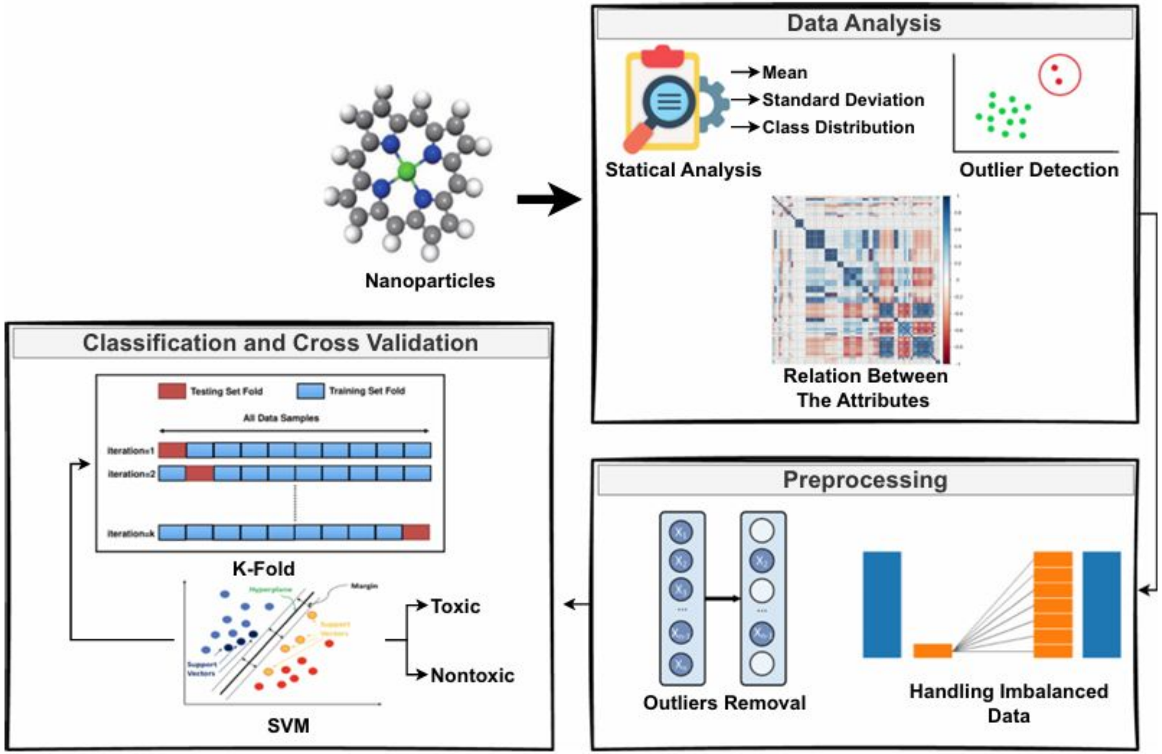

Fig. 1 The proposed ML toxicity detection of metal oxide nanoparticles model block diagram

Statistical analysis

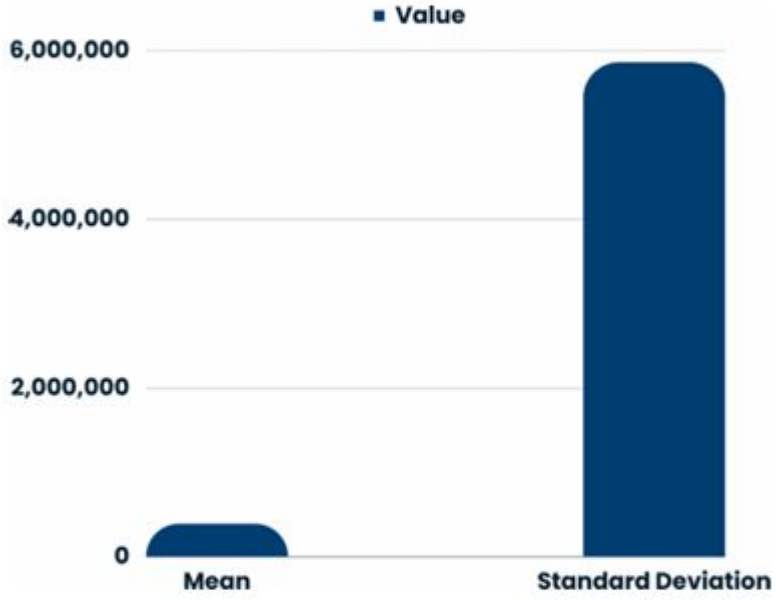

Fig. 3 Mean and standard deviation of the dataset

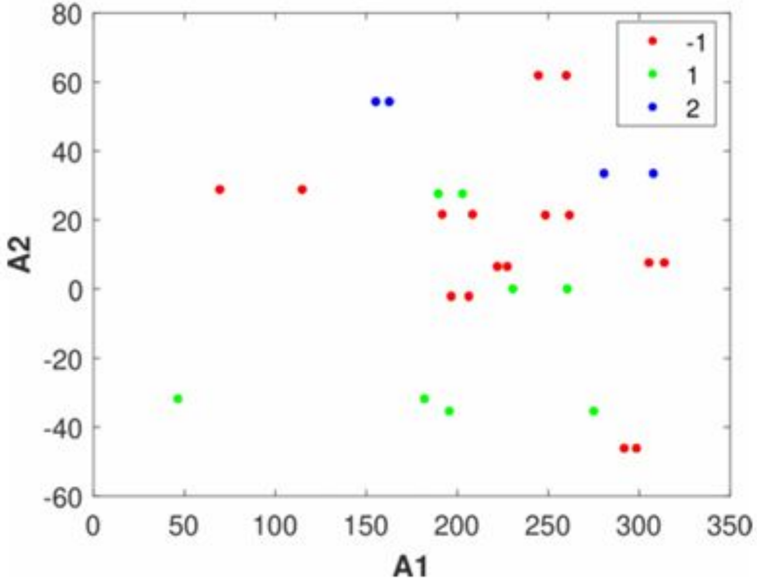

Fig. 4 DBSCAN result using squared Euclidean Distance Metric

Attribute Selection: EBSTSA

Reduction of irrelevant variables → Highlights: 83.36% reduction in variables

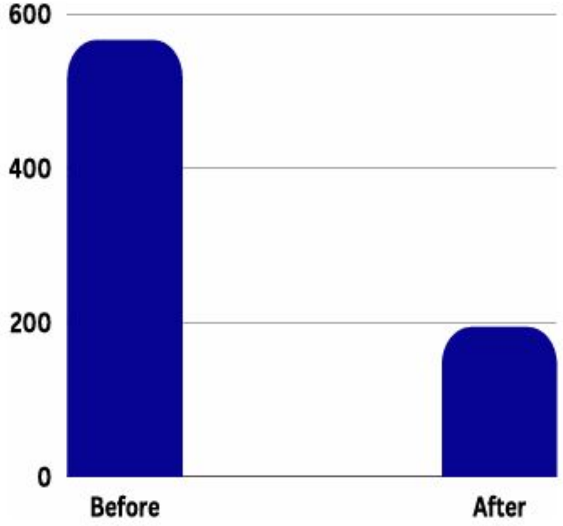

Fig. 5 The number of outliers based on using the Z-score method before and after applying the proposed EBSTSA

- **Why use EBSTSA?**
  - Designed for complex binary problems.
  - Improved local optimum escape capability.
  - High efficiency in terms of convergence and solution quality.
  - Can compete with or outperform other binary algorithms such as BPSO (Binary Particle Swarm Optimization), BBA (Binary Bat Algorithm), BGWO (Binary Grey Wolf Optimizer), etc.

Balancing the dataset

How did they resolve the toxic/non-toxic class imbalance?

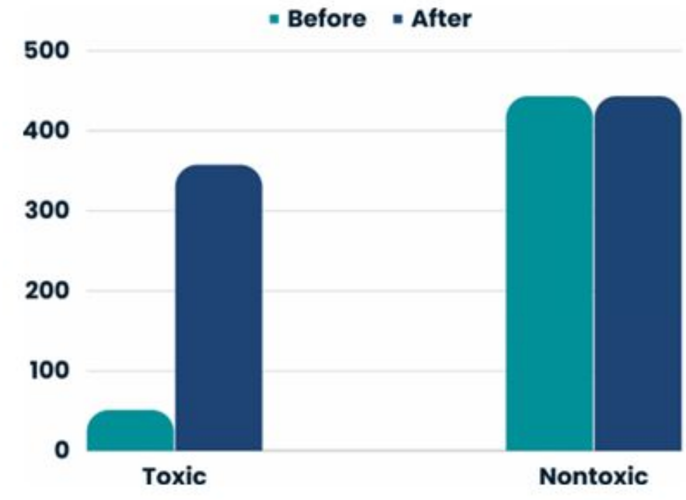

Fig. 7 Classes distribution of the data before and after using SMOTE

This was key because there were only 51 toxic cases compared to 443 non-toxic ones.

Classification results

- Effectiveness of the model
  - Figure 9: Error rate (1.02%)
  - Figure 10: F1 score 99.47%, Sensitivity 98.87%, Specificity 100%
  - Figure 12: ROC curve, AUC = 0.986

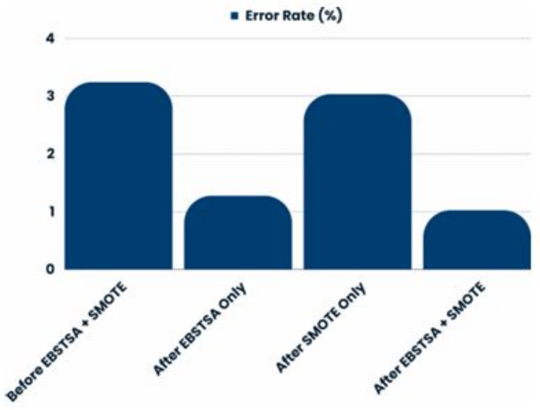

Fig. 9 The average error rate before and after applying SMOTE and EBSTSTA

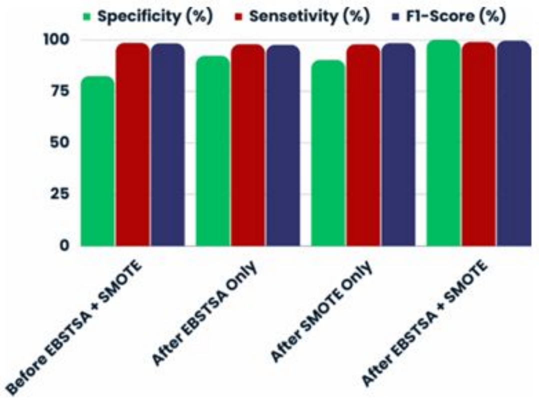

Fig. 10 The average sensitivity, specificity, and F1-score before and after applying SMOTE and EBSTSTA

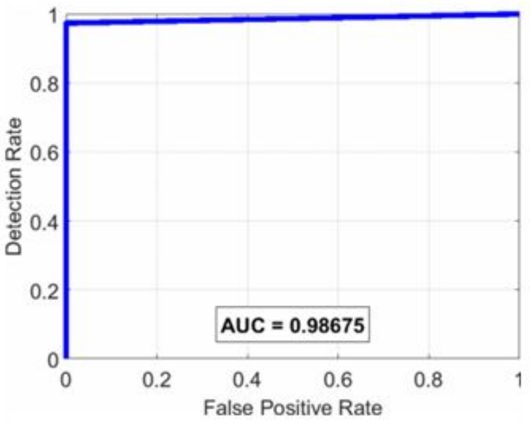

Fig. 12 The ROC curve of the proposed model

Conclusion of the article

- It proposes a reproducible pipeline for predicting NP toxicity.
- It is adaptable to other materials such as Sm_2O_3 or Eu_2O_3.
- It highlights the role of attribute selection and class balancing.
- It provides evidence for moving toward "safe by design" nanotechnology.

How would I apply this approach to my thesis?

- Construct a dataset based on toxicity studies of Sm and Eu NPs
- Use similar attributes for NP characterization: size, charge, crystal structure, bandgap, etc.
- Apply a pipeline based on the one in this paper (EBSTSA or SHAP + SMOTE + SVM)
- Objective: Develop a robust and interpretable model that predicts toxicity with experimental data