Nanoparticle Toxicity Data Card

Dataset Documentation and Provenance

Data Card: Nanoparticle Toxicity Dataset

Generated: 2025-11-01 09:12

Dataset: KONA2025_final.csv

Samples: 308 nanoparticles

Features: 25 variables

1. Dataset Overview

DATASET PURPOSE: Collection of nanoparticle characterization and toxicity data for machine learning model development in nanotoxicology.

DATA COLLECTION: Compiled from published literature and experimental studies on nanoparticle biological interactions. Includes metal oxides, carbon-based, and polymeric nanoparticles.

DATA COMPLETENESS: 90 complete samples with full characterization and toxicity assessment. No missing values in final curated dataset.

ETHICAL APPROVAL: All data from published studies with appropriate experimental ethics approvals where applicable.

2. Target Variable: Toxicity Classification

VARIABLE: Toxicity Class

TYPE: Categorical (4 classes)

DESCRIPTION: Classification of nanoparticle toxicity based on biological response metrics

Nanoparticle Toxicity Data Card

Dataset Documentation and Provenance

CLASS DEFINITIONS:

- Class 0: Non-toxic (Cell viability >70%, minimal adverse effects)

- Class 1: Low toxicity (Cell viability 50-70%, mild effects)

- Class 2: Moderate toxicity (Cell viability 30-50%, significant effects)

- Class 3: High toxicity (Cell viability <30%, severe adverse effects)

DISTRIBUTION: Class 0: 56 samples, Class 1: 15 samples, Class 2: 14 samples, Class 3: 5 samples

3. Feature Descriptions

Complete feature descriptions with categories and measurement units:

Feature Name	Category	Description
Core Size (nm)	Physical Property	Primary particle diameter in nanometers
Shape	Physical Property	Nanoparticle morphology (spherical, rod, etc.)
Zeta Potential (mV)	Physical Property	Surface charge measurement indicating stability
Surface Area (m2/g)	Physical Property	Specific surface area per mass unit
Aggregation State	Physical Property	Degree of nanoparticle aggregation in solution
Dissolution Rate (mg/L)	Chemical Property	Rate of dissolution in biological media
Zn2+ Ion Release (mg/L)	Chemical Property	Zinc ion release concentration
Impurity Content (%)	Chemical Property	Percentage of impurities in nanoparticle synthesis
Surface Chemistry	Chemical Property	Chemical functional groups on surface
Coating Type	Chemical Property	Type of surface coating or functionalization
Oxidative Potential	Chemical Property	Ability to generate reactive oxygen species

Nanoparticle Toxicity Data Card

Dataset Documentation and Provenance

Exposure Dosage (ug/mL)	Exposure Condition	Concentration administered to cells
Exposure Time (hrs)	Exposure Condition	Duration of nanoparticle exposure
Cell Type	Exposure Condition	Type of cells used in toxicity assay
ROS Production	Biological Response	Reactive oxygen species generation level
Membrane Damage (%)	Biological Response	Percentage of cells with membrane damage
Apoptosis (%)	Biological Response	Percentage of cells undergoing programmed death
Necrosis (%)	Biological Response	Percentage of cells undergoing unplanned death
IC50 Value (ug/mL)	Biological Response	Concentration causing 50% inhibition
Cell Viability (%)	Biological Response	Percentage of living cells after exposure

4. Data Quality and Preprocessing

DATA COLLECTION STANDARDS:

- All measurements follow standardized protocols

- Biological assays performed in triplicate

- Physicochemical characterization using validated methods

- Consistent units and reporting formats

PREPROCESSING STEPS:

1. Missing value imputation (median for numeric, mode for categorical)

2. Feature encoding (one-hot for categorical variables)

3. Outlier detection using IQR method

4. Feature engineering (surface-to-volume ratios, dosage-time interactions)

5. Train-test split with stratification (80-20 split)

QUALITY ASSURANCE:

- Cross-validation used for model evaluation

- Group-based validation to prevent data leakage

- Multiple performance metrics reported

- Feature importance analysis for interpretability

5. Usage Guidelines and Limitations

RECOMMENDED USE:

- Machine learning model development for nanotoxicology

- Feature importance analysis in nanoparticle safety

- Educational purposes in materials science

- Hypothesis generation for experimental studies

Nanoparticle Toxicity Data Card

Dataset Documentation and Provenance

DATA LIMITATIONS:

- Limited sample size for some nanoparticle types

- Class imbalance in toxicity labels

- Primarily in-vitro data (limited in-vivo correlation)

- Focus on acute toxicity (chronic effects not captured)

CITATION REQUIREMENTS:

When using this dataset, please cite the original publications and this data card.

DATA ACCESS:

Available for academic and research purposes with proper attribution.

Commercial use requires additional permissions.