A Large-Scale Machine Learning Analysis of Inorganic Nanoparticles in Preclinical Cancer Research

Zilu Zhang[1], Bárbara B. Mendes[2], João Conniot[2], Diana P. Sousa[2], João M. J. M. Ravasco[2], Andżelika Lorenc[3,4], Tiago Rodrigues[4], João Conde[2], Daniel Reker[1,*]

BMES BIOMEDICAL ENGINEERING SOCIETY 2023 ANNUAL MEETING
October 11-14, 2023
Seattle Convention Center | Seattle, WA

Duke BIOMEDICAL ENGINEERING

Reker Lab

1 Introduction

Inorganic nanoparticles have become an important tool as **cancer** therapeutics, diagnostics, and theranostics.[1] However, notwithstanding the achievements that have been made, it remains challenging to objectively, comprehensively, and systematically define **key design requirements** to ensure the success of the preclinical development of inorganic nanomedicines.[2] To discover such design rules, we have curated **the world's largest database** of inorganic nanoparticles in preclinical cancer research and use **statistics** and **machine learning** to capture trends in nanoparticle design, correlate nanoparticle properties with in vivo experiment outcomes, and provide predictive tools that can guide the creation of safe and efficacious nanoparticles for cancer drug delivery.

2 Exposition

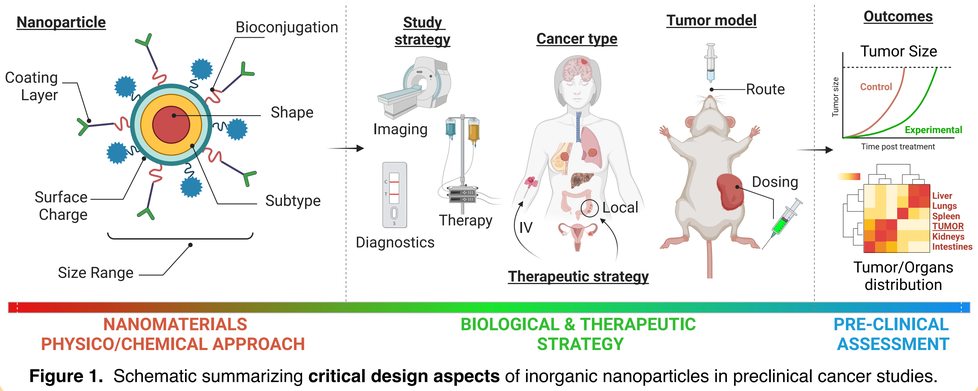

Figure 1. Schematic summarizing **critical design aspects** of inorganic nanoparticles in preclinical cancer studies.

3 Study Design

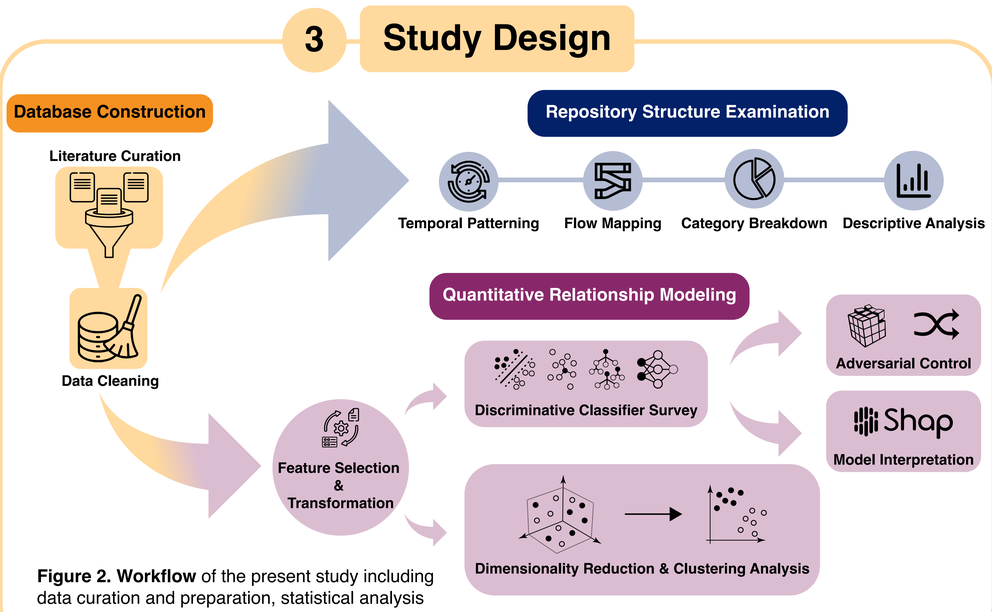

Figure 2. Workflow of the present study including data curation and preparation, statistical analysis and machine learning-driven modeling.

4 Repository Structure Examination

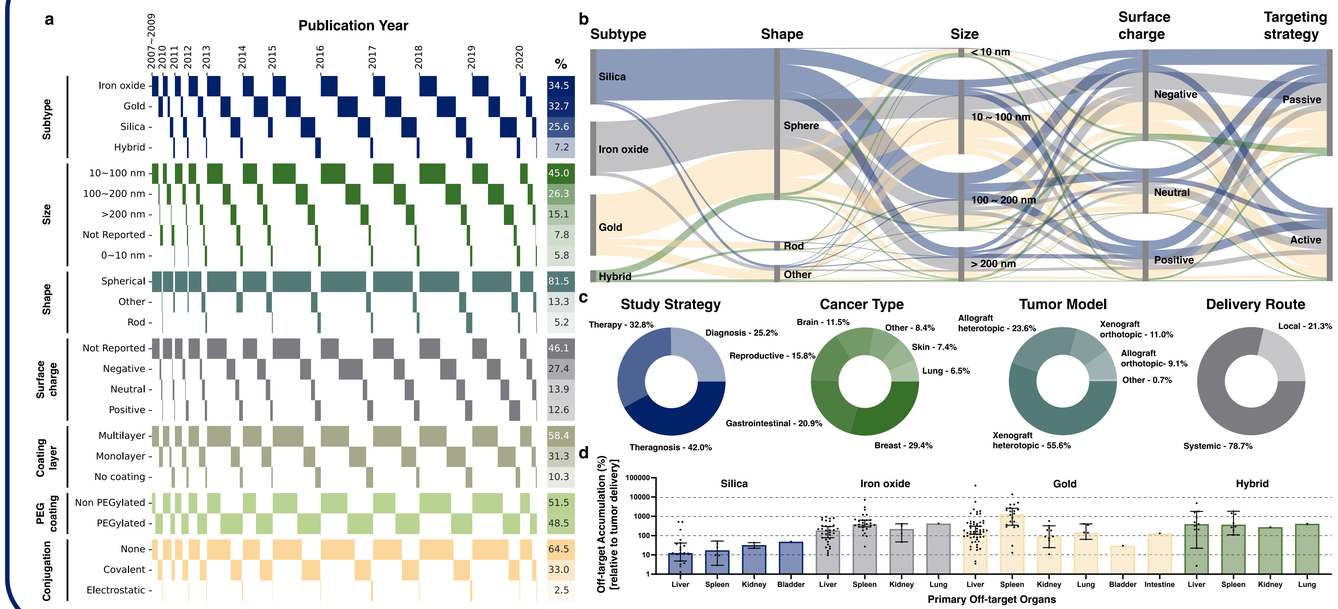

Figure 3. Statistical analysis reveals **trends** in nanoparticle and preclinical study design over the last decades.
(**a**) The relative distribution of key properties in nanoparticle design over time.
(**b**) An alluvial diagram showing the co-occurrences of key properties.
(**c**) Overall proportions of critical settings in preclinical experiment design.
(**d**) Nanoparticle off-target distributions in primary organs relative to tumor accumulation. Error bar, median with 95% CI.

5 Quantitative Relationship Modeling

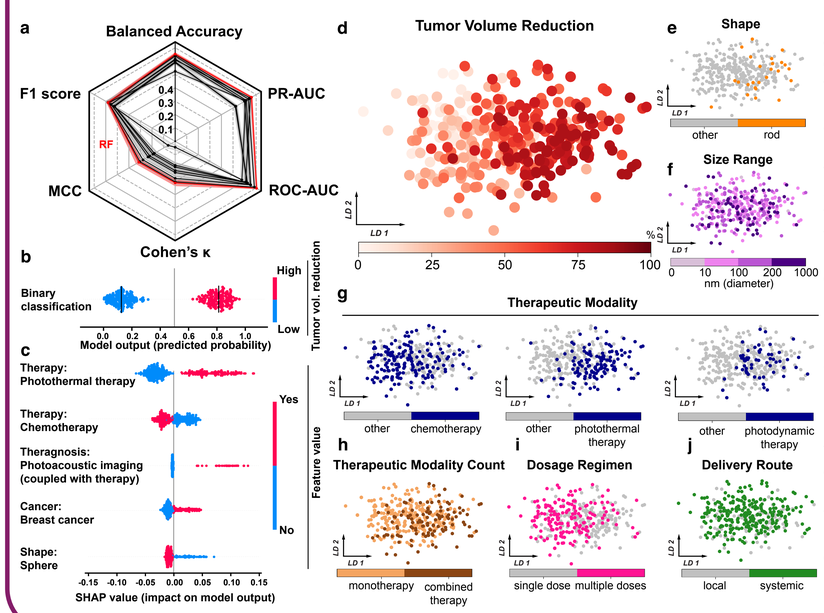

Figure 4. Machine learning demonstrates predictive capability in **anticipating treatment outcomes** and **identifying significant contributing factors**.
(**a**) A spider plot summarizing the performance of the surveyed models. Random forest classifier (RF), red. Other tested models, gray.
(**b**) Prediction outputs of the complete dataset given by the random forest model.
(**c**) Five important features are selected based on importance and displayed to represent a broad range of different nanoparticle properties.
(**d**) Dimensionality reduction of the original dataset using linear discriminant analysis. Data points are colored according to the reported tumor volume reduction. The same distribution is also colored based on other features (**e** - **j**).

6 Conclusions

- We created the **world's largest database** covering more than 700 publications to capture the advancements and underscore the remaining gaps in applying inorganic nanoparticles for preclinical cancer research.
- A tree-based **supervised machine learning model** demonstrated superior performance in classifying nanodrugs using the provided descriptors for nanoparticle design and preclinical study setup.
- **Explainable AI** (XAI) proves invaluable in identifying key contributing features and clarifying the decision-making process.
- **A standardized framework** is urgently required for comprehensive reporting of nanoparticle design and outcome evaluation.

7 References

1. Mendes, B. B. *et al.*, Nanomedicine-based strategies to target and modulate the tumor microenvironment. *Trends Cancer* **7**, 847–862 (2021).
2. Faria, M. *et al.* Minimum information reporting in bio–nano experimental literature. *Nat Nanotechnol* **13**, 777–785 (2018).

Author Affiliations
[1]Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA.
[2]ToxOmics, NOVA Medical School, Faculdade de Ciências Médicas, NMSIFCM, Universidade NOVA de Lisboa; Lisboa, Portugal.
[3]Instituto de Investigação do Medicamento (iMed), Faculdade de Farmácia, Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisboa, Portugal.
[4]Department of Biopharmacy, Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, Jurasza 2, 85-089 Bydgoszcz, Poland.
*Correspondence: daniel.reker@duke.edu

Research Supported by
Duke University
Universidade Nova de Lisboa
U Lisboa | Universidade de Lisboa
Nicolaus Copernicus University in Toruń
National Institutes of Health
European Research Council