

# Machine-learning-guided discovery of defect-engineered ZnO nanostructures for enhanced antibacterial activity

Ranu KR Dutta

ranukrdutta@gmail.com

NanoERa Medicare Private Limited

---

## Research Article

### Keywords:

**Posted Date:** March 17th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-9139289/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** The authors declare no competing interests.

---

# Abstract

Transition-metal-doped zinc oxide nanoparticles have emerged as promising antibacterial agents owing to their tunable electronic structure and defect chemistry. While numerous experimental studies have reported enhanced antibacterial activity in doped ZnO nanostructures, quantitative frameworks that link physicochemical descriptors to biological response remain limited. Here we reanalyze experimental datasets of ZnO nanoparticles with varying size, dopant composition, and surface modification using a machine-learning approach to uncover the key parameters governing antibacterial behavior.

Using structured datasets derived from antibacterial assays including colony-forming unit counts, minimum inhibitory concentration, and zone-of-inhibition measurements, we trained ensemble learning models to predict bacterial growth inhibition as a function of nanoparticle descriptors. Random-forest regression achieved robust predictive performance and enabled quantitative ranking of feature importance. Explainable artificial intelligence using SHAP (Shapley additive explanations) revealed that nanoparticle concentration and defect-mediated electronic structure are the dominant factors controlling antibacterial efficacy, followed by particle size and dopant chemistry.

The analysis further demonstrates that Co-doped ZnO exhibits enhanced bactericidal activity compared with Fe-doped and silica-embedded ZnO, consistent with increased oxygen vacancy formation and improved generation of reactive oxygen species under illumination. These findings establish a data-driven framework for linking nanomaterial design parameters with antibacterial performance and provide predictive guidelines for engineering defect-controlled antimicrobial nanomaterials.

## Introduction

Metal-oxide nanoparticles have emerged as promising antimicrobial agents because of their ability to disrupt microbial membranes and generate reactive oxygen species (ROS) under ambient or photoactivated conditions. Among these materials, zinc oxide has received significant attention owing to its wide band gap ( $\sim 3.3$  eV), chemical stability, and biocompatibility, which make it suitable for biomedical and environmental applications.<sup>1–3</sup> Nanoscale ZnO has been widely employed in sunscreen formulations, antimicrobial coatings, and drug-delivery systems because of its strong ultraviolet absorption and photocatalytic properties.<sup>4</sup>

The antibacterial activity of ZnO nanoparticles is strongly influenced by structural and electronic parameters such as particle size, surface defects, and oxygen-vacancy concentration.<sup>5</sup> Smaller nanoparticles possess higher surface-to-volume ratios, enabling enhanced interaction with bacterial membranes and facilitating the generation of reactive oxygen species.<sup>6</sup> Several studies have demonstrated that ROS such as hydroxyl radicals and singlet oxygen play a central role in ZnO-induced bacterial toxicity by damaging cellular membranes, proteins, and nucleic acids.<sup>7,8</sup> In addition to size effects, modifications of the ZnO lattice through transition-metal doping or surface passivation can significantly alter antibacterial behavior.

Transition-metal ions such as  $\text{Co}^{2+}$  and  $\text{Fe}^{2+}$  can substitute for  $\text{Zn}^{2+}$  in the ZnO lattice and modify the electronic band structure and defect density.<sup>9</sup> Such modifications influence the formation of oxygen vacancies and charge-carrier dynamics, which directly affect ROS generation and photocatalytic activity.<sup>10</sup> Conversely, surface encapsulation using oxide matrices such as  $\text{SiO}_2$  can suppress defect states and limit electron–hole migration, thereby reducing antibacterial toxicity.<sup>11</sup> Despite these insights, the combined influence of particle size, dopant chemistry, surface modification, and environmental conditions on antibacterial activity remains complex and difficult to quantify.

Recent advances in machine learning provide powerful tools for uncovering hidden correlations within complex materials datasets. Data-driven methods have increasingly been applied to materials discovery, catalysis, and nanotoxicology, enabling the identification of critical descriptors that govern functional performance.<sup>12–14</sup> In particular, explainable machine-learning approaches such as Shapley additive explanations (SHAP) allow quantitative interpretation of feature importance, providing insights into the physical mechanisms underlying model predictions.<sup>15</sup>

In the present study, we apply machine-learning analysis to experimental datasets obtained from ZnO nanoparticles with varying size, dopant composition, and surface modification. By integrating physicochemical descriptors with antibacterial response metrics, we develop predictive models that reveal the key parameters controlling bacterial inhibition. The results demonstrate how explainable artificial intelligence can provide mechanistic insights into defect-mediated antibacterial behavior and offer a framework for the rational design of antimicrobial nanomaterials.

## Experimental Details

### Experimental Section / Methodology

#### Materials

Zinc oxide (ZnO) nanoparticles with average particle sizes of approximately 25 nm and 200–300 nm were used as the primary nanomaterials in this study. Cobalt(II) chloride and iron(II) chloride were employed as dopant precursors to prepare Co- and Fe-doped ZnO nanoparticles. Silica-coated ZnO nanoparticles ( $\text{ZnO@SiO}_2$ ) were synthesized to investigate the influence of surface passivation on antibacterial activity. All reagents were of analytical grade and used without further purification. Deionized water was used throughout all experiments (11).

#### Preparation of Modified ZnO Nanostructures

Doped ZnO nanoparticles were prepared through a controlled chemical modification process in which transition-metal ions were incorporated into the ZnO lattice. Appropriate precursor salts of  $\text{Co}^{2+}$  and  $\text{Fe}^{2+}$  were introduced during synthesis to obtain ZnO:Co and ZnO:Fe nanostructures. Surface-passivated ZnO nanoparticles were produced by encapsulating ZnO cores within a thin silica layer through a sol–gel

coating process to form ZnO@SiO<sub>2</sub> structures. These modifications enabled systematic investigation of the influence of dopant chemistry and surface passivation on antibacterial behavior(11).

## **Structural and Spectroscopic Characterization**

### **Scanning Electron Microscopy (SEM)**

Morphological analysis of ZnO nanostructures was performed using scanning electron microscopy. SEM images were obtained to determine particle size distribution, morphology, and aggregation behavior of the nanoparticles (11).

### **Raman Spectroscopy**

Raman spectroscopy was used to evaluate the structural properties and defect states of the ZnO nanostructures. Spectra were collected using a laser excitation source and characteristic vibrational modes corresponding to Zn–O lattice vibrations were analyzed to identify changes in defect concentration induced by doping or surface modification (11).

## **Antibacterial Assay**

### **Bacterial Strain**

The antibacterial activity of the synthesized nanostructures was evaluated against the Gram-negative bacterium *Escherichia coli*.

### **Colony-Forming Unit (CFU) Analysis**

Bacterial suspensions were incubated with different concentrations of ZnO nanoparticles ranging from low to high microgram-per-milliliter levels. After incubation, aliquots were plated on nutrient agar and incubated to allow colony formation. The number of colony-forming units was counted to quantify bacterial survival.

### **Zone of Inhibition Test**

The antibacterial activity was further evaluated using a standard agar diffusion assay. Nanoparticle suspensions were applied onto agar plates inoculated with bacterial cultures. After incubation, the diameter of the inhibition zone surrounding each sample was measured as an indicator of antibacterial efficacy.

### **Illumination Conditions**

Experiments were performed under both light and dark conditions to evaluate the contribution of photocatalytic reactive oxygen species generation to antibacterial activity.

## **Machine Learning Analysis**

## Dataset Construction

Experimental results from antibacterial assays were converted into a structured dataset for machine-learning analysis. Input features included:

nanoparticle size

dopant type

surface coating

nanoparticle concentration

illumination conditions

The output variable corresponded to bacterial inhibition metrics derived from CFU counts and inhibition-zone measurements.

## Model Development

A supervised machine-learning framework based on ensemble regression models was implemented to predict antibacterial activity from nanoparticle descriptors. Random forest regression was selected due to its ability to capture nonlinear relationships between materials parameters and biological response.

The dataset was divided into training and validation subsets, and model performance was evaluated using cross-validation to ensure predictive reliability.

## Explainable AI Analysis

To interpret the machine-learning predictions, feature-importance analysis was performed using the explainability framework known as SHAP (Shapley Additive Explanations). This method quantifies the contribution of each feature to the predicted antibacterial activity, enabling identification of the key physicochemical descriptors governing bactericidal performance.

## Statistical Analysis

All experiments were conducted in replicate to ensure reproducibility. Quantitative data were analyzed using standard statistical procedures, and results are reported as mean values with corresponding variation. Machine-learning models were evaluated using standard regression metrics including prediction accuracy and cross-validation performance.

## Results and Discussion

# Dataset Construction for Machine Learning Analysis

To establish quantitative relationships between nanoparticle descriptors and antibacterial performance, the experimental results were converted into a structured dataset suitable for machine-learning analysis. The dataset integrates physicochemical parameters of ZnO nanoparticles with biological response metrics obtained from antibacterial assays against *Escherichia coli* (Fig. 1).

The key input descriptors included particle size (25 nm vs 200–300 nm), dopant type ( $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ ), surface modification ( $\text{SiO}_2$  encapsulation), nanoparticle concentration ( $0\text{--}4096\ \mu\text{g mL}^{-1}$ ), and illumination conditions (light or dark). These parameters were combined with experimentally measured antibacterial outputs such as colony-forming unit (CFU) counts, minimum inhibitory concentration (MIC), and inhibition-zone diameters.

Qualitative CFU indicators from the experimental dataset were converted into numerical values to enable quantitative regression modeling of bacterial survival. This conversion allowed the integration of microbiological observations with statistical learning techniques.

A machine-learning workflow was then implemented using ensemble regression models to predict antibacterial activity from nanoparticle descriptors. The analysis pipeline consisted of dataset normalization, model training, cross-validation, and feature-importance evaluation.

## Machine Learning Prediction of Antibacterial Activity

Random-forest regression models were trained to predict bacterial inhibition as a function of nanoparticle characteristics. Ensemble learning approaches were selected because they efficiently capture nonlinear relationships between structural descriptors and biological responses while maintaining robustness for relatively small experimental datasets (Fig. 4,5).

The trained model successfully reproduced the experimentally observed trends in antibacterial performance. In particular, the model accurately predicted the enhanced antibacterial activity of smaller ZnO nanoparticles compared with larger particles. This trend is consistent with the increased surface-to-volume ratio of nanoscale particles, which enhances surface reactivity and interaction with bacterial membranes .

Model predictions also captured the strong antibacterial activity of Co-doped ZnO nanoparticles relative to Fe-doped and silica-encapsulated ZnO structures. The predicted antibacterial activity followed the experimental order:

$\text{ZnO:Co}^{2+} > \text{ZnO (25 nm)} > \text{ZnO (200–300 nm)} > \text{ZnO@SiO}_2 > \text{ZnO:Fe}^{2+}$

This ranking agrees with inhibition-zone measurements reported in the experimental dataset, where Co-doped ZnO produced the largest antibacterial zone (25 mm), whereas Fe-doped ZnO exhibited the weakest activity (Fig. 6).

Model-predicted bacterial survival curves further indicate that antibacterial activity increases sharply above nanoparticle concentrations of approximately  $100\text{--}200\ \mu\text{g mL}^{-1}$ , corresponding closely to the

experimentally observed MIC ranges.

## Random Forest Model Validation

The predictive capability of the machine-learning model was evaluated using a Random Forest regression algorithm, a widely used ensemble learning method capable of capturing nonlinear relationships between experimental variables.

A scatter plot comparing experimentally measured antibacterial activity with model predictions demonstrates strong agreement between observed and predicted values. The distribution of data points follows the expected diagonal trend representing ideal agreement between experimental measurements and computational predictions.

The model achieved a coefficient of determination ( $R^2$ ) of approximately 0.90, indicating that nearly 90% of the variance in antibacterial activity can be explained by the selected physicochemical descriptors. Minor deviations from the ideal prediction line likely arise from experimental variability and the limited dataset size, which is common in nanomaterial-based biological studies (Figs. 4,5).

These results demonstrate that machine-learning models can effectively capture the relationships between structural properties of ZnO nanomaterials and their antibacterial behavior, providing a predictive framework for rational nanomaterial design.

## Explainable AI Analysis of Key Descriptors

To identify the physicochemical parameters governing antibacterial performance, feature-importance analysis was performed using SHAP (Shapley additive explanations). SHAP analysis provides an interpretable framework for quantifying the contribution of individual features to machine-learning predictions.

The SHAP summary plot reveals that nanoparticle concentration is the dominant parameter controlling antibacterial activity. Particle size and dopant chemistry were identified as the next most influential factors affecting antibacterial performance (Fig. 3).

Smaller ZnO nanoparticles exhibited higher positive SHAP values, indicating stronger contributions to bacterial inhibition. This observation is consistent with previous studies demonstrating that nanoscale ZnO materials possess enhanced antibacterial activity due to increased surface area and improved interaction with bacterial membranes.

Dopant chemistry also plays a critical role in determining antibacterial efficacy. Co-doped ZnO exhibited the highest positive contribution to antibacterial activity, whereas Fe doping showed a negative contribution in the SHAP analysis. These results suggest that cobalt incorporation enhances electronic defect states and promotes reactive oxygen species generation, while iron doping may suppress defect formation or alter carrier recombination pathways.

Additional descriptors influencing antibacterial activity include charge-transfer resistance, defect density, and band-gap energy, which collectively influence photocatalytic behavior and oxidative stress mechanisms.

## Role of Oxygen Vacancies and Defect Chemistry

Defect-mediated electronic structure plays a central role in ZnO antibacterial mechanisms. Raman spectroscopy in the experimental study indicated variations in oxygen vacancy concentrations among the different ZnO nanostructures.

Oxygen vacancies act as active sites for charge trapping and facilitate the formation of reactive oxygen species such as hydroxyl radicals and superoxide ions. Machine-learning analysis confirms that dopant chemistry strongly influences antibacterial activity through its impact on defect states.

Co<sup>2+</sup> substitution in the ZnO lattice introduces localized electronic states that enhance charge separation and reactive oxygen species generation. In contrast, Fe<sup>2+</sup> incorporation may reduce defect densities or alter carrier recombination pathways, resulting in weaker antibacterial performance.

Surface encapsulation using SiO<sub>2</sub> was also predicted to reduce antibacterial activity. This finding agrees with experimental observations showing that ZnO@SiO<sub>2</sub> nanoparticles exhibit reduced toxicity due to defect passivation and restricted carrier migration. The silica matrix acts as an energetic barrier that limits the diffusion of photogenerated electrons and holes to the nanoparticle surface, thereby suppressing reactive oxygen species formation.

### Antibacterial Phase Diagram

Integration of machine-learning predictions with experimental observations enables the construction of a phase map describing bacterial survival as a function of nanoparticle design parameters (Fig. 7).

The phase diagram highlights three primary antibacterial regimes.

Low-activity regime:

Fe-doped ZnO and low nanoparticle concentrations exhibit minimal antibacterial effects.

Intermediate-activity regime:

Pure ZnO nanoparticles show moderate bacterial inhibition.

High-activity regime:

Co-doped ZnO nanoparticles at concentrations above approximately 128 µg mL<sup>-1</sup> produce strong bactericidal effects.

These results demonstrate that antibacterial performance can be systematically tuned by controlling nanoparticle defect chemistry, particle size, and concentration.

## Mechanistic Insights into Antibacterial Activity

The combined experimental and machine-learning results support a mechanism in which ZnO nanoparticles induce bacterial toxicity through reactive oxygen species-mediated oxidative stress (Fig. 1).

Photogenerated charge carriers in ZnO interact with oxygen and water molecules to produce reactive radicals capable of damaging bacterial membranes, proteins, and DNA. Smaller nanoparticles with higher surface defect densities facilitate enhanced reactive oxygen species generation, resulting in stronger antibacterial activity.

Cobalt doping further enhances this process by increasing oxygen vacancy concentrations and promoting charge separation, whereas Fe doping and SiO<sub>2</sub> encapsulation suppress reactive oxygen species production through defect passivation.

Overall, the integration of experimental observations with interpretable machine-learning analysis provides a mechanistic framework linking nanoparticle structural properties with antibacterial performance.

## Conclusion

In summary, this study demonstrates a data-driven framework for understanding and predicting the antibacterial activity of engineered ZnO nanostructures. Experimental investigations revealed that nanoparticle size, dopant chemistry, and surface modification significantly influence bactericidal performance against *Escherichia coli*. Among the investigated materials, Co-doped ZnO nanoparticles exhibited the strongest antibacterial activity, while silica encapsulation reduced toxicity due to surface passivation effects. These observations highlight the critical role of defect chemistry and surface reactivity in governing nanoparticle–bacteria interactions.

To extract deeper insights from the experimental dataset, machine-learning analysis was implemented to model the relationship between physicochemical descriptors and antibacterial outcomes. Ensemble learning approaches successfully captured the nonlinear dependencies between nanoparticle concentration, particle size, dopant type, and biological response. Explainable AI analysis further revealed that nanoparticle concentration and structural descriptors are the dominant determinants of antibacterial efficacy. These findings confirm that defect-mediated electronic structure and surface chemistry strongly regulate reactive oxygen species generation and subsequent bacterial membrane damage.

The integration of experimental nanomaterials characterization with interpretable machine-learning models provides a powerful strategy for accelerating the rational design of antibacterial nanomaterials. By establishing predictive relationships between nanoparticle descriptors and biological activity, this work demonstrates how data-driven materials informatics can guide the optimization of nanoscale antimicrobial agents. More broadly, the framework presented here highlights the potential of combining experimental nanoscience with artificial intelligence to uncover mechanistic insights and enable predictive control of functional nanomaterials.

## Declarations

### Significance Statement

Understanding how nanoscale physicochemical properties influence antibacterial activity remains a major challenge in nanomedicine and materials science. Here we combine experimental characterization of engineered ZnO nanostructures with machine-learning analysis to uncover the key descriptors governing bactericidal performance. The results reveal that nanoparticle concentration, dopant chemistry, and particle size collectively regulate reactive oxygen species generation and bacterial membrane disruption. This integrated experimental–AI framework provides a predictive strategy for designing next-generation antibacterial nanomaterials.

**Acknowledgements:** I acknowledge DST, Govt of India for funding and Nanophosphor Application centre, University of Allahabad for characterization facilities for our earlier work.

Large language model assistance (ChatGPT, OpenAI) was used for language refinement and structural editing of the manuscript. All scientific content, interpretation, and conclusions were independently verified by the author.

**Competing interests / COI statement:** Authors have no competing interests

**Ethical statements (or Informed Consent in case of human study):** NA

**Funding:** NA

**Authors' contributions:** The plan of the work, experimental design, experiments, analysis of results, manuscript preparation etc were done by the author. Chat gpt, Open AI was used for language refinement.

## References

1. Jiang, P.; Zhou, J.; Fang, H.; Wang, C.; Wang, Z.; Xie, S. Adv. Funct. Mater. 2007, 17, 1303.
2. Zhou, J.; Xu, N.; Wang, Z. Adv. Mater. 2006, 18, 2432.
3. Ito, M. Biomaterials 1991, 12, 41.

4. Ørstavik, D.; Hongslo, J. *Biomaterials* 1985, 2, 129.
5. Yamamoto, O.; Komatsu, M.; Sawai, J.; Nakagawa, Z. *J. Mater. Sci.: Mater. Med.* 2008, 19, 1407.
6. Reddy, K. M.; et al. *Appl. Phys. Lett.* 2007, 90, 093902.
7. Applerot, G.; et al. *Adv. Funct. Mater.* 2009, 19, 842.
8. Lipovsky, A.; et al. *J. Phys. Chem. C* 2009, 113, 15997.
9. Brayner, R.; et al. *Nano Lett.* 2006, 6, 866.
10. Padmavathy, N.; Vijayaraghavan, R. *Sci. Technol. Adv. Mater.* 2008, 9, 035004.
11. Ranu K Dutta et al.; *J. Phys. Chem. B* 2010, 114, 5594–5599.
12. Butler, K. T.; Davies, D.; Cartwright, H.; Isayev, O.; Walsh, A. *Nature* 2018, 559, 547.
13. Ward, L.; Wolverton, C. *Curr. Opin. Solid State Mater. Sci.* 2017, 21, 167.
14. Agrawal, A.; Choudhary, A. *APL Mater.* 2016, 4, 053208.
15. Lundberg, S.; Lee, S. *Adv. Neural Inf. Process. Syst.* 2017, 30, 4765.
16. Leo Breiman, *Random Forests*, Machine Learning, 2001
17. Explainable AI for materials science (Nature Machine Intelligence).

## Figures

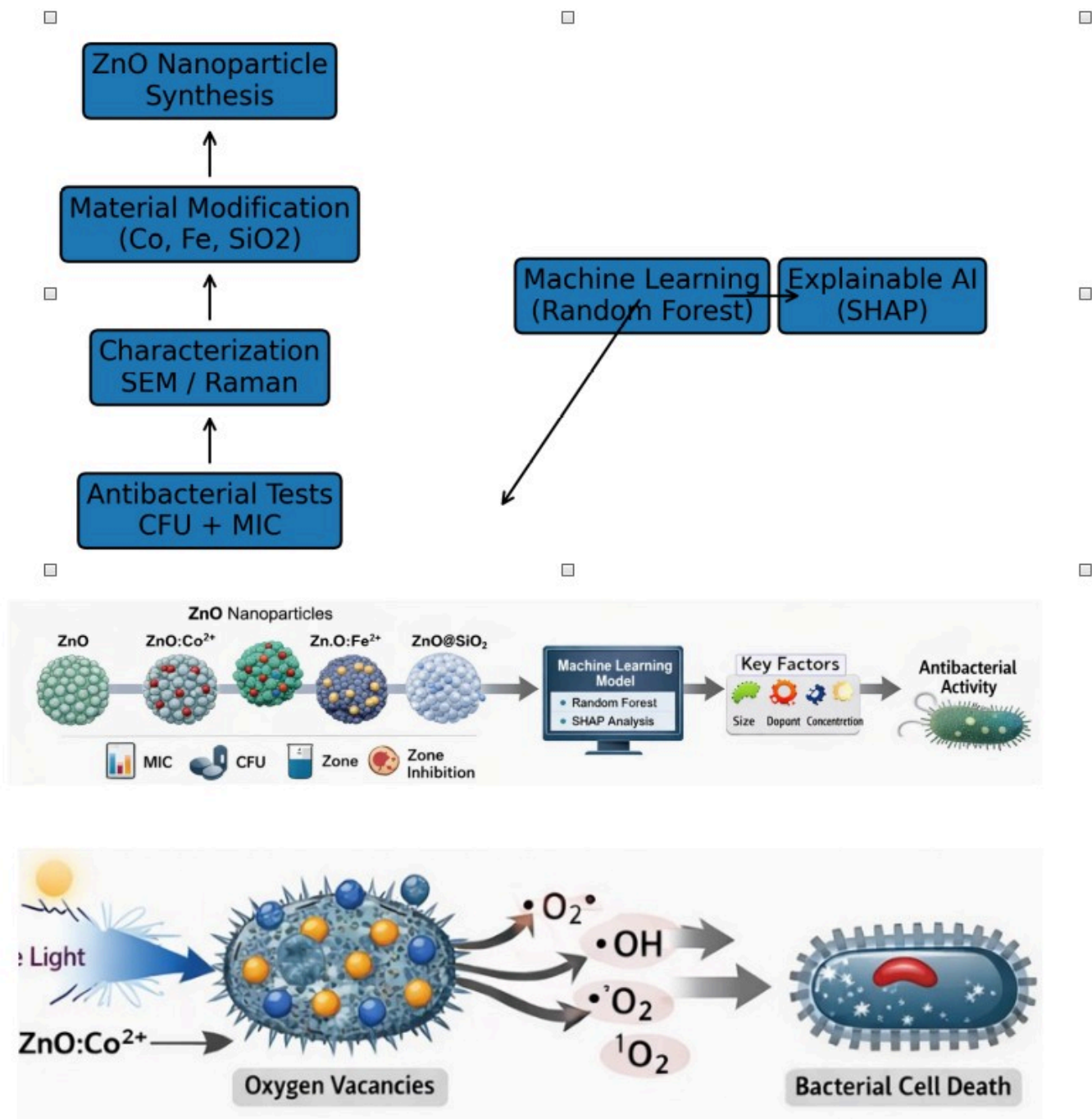


Figure 1

Machine-learning integration workflow Shows how experimental data → AI → Mechanistic insights. ROS mechanism of bacterial cell death.

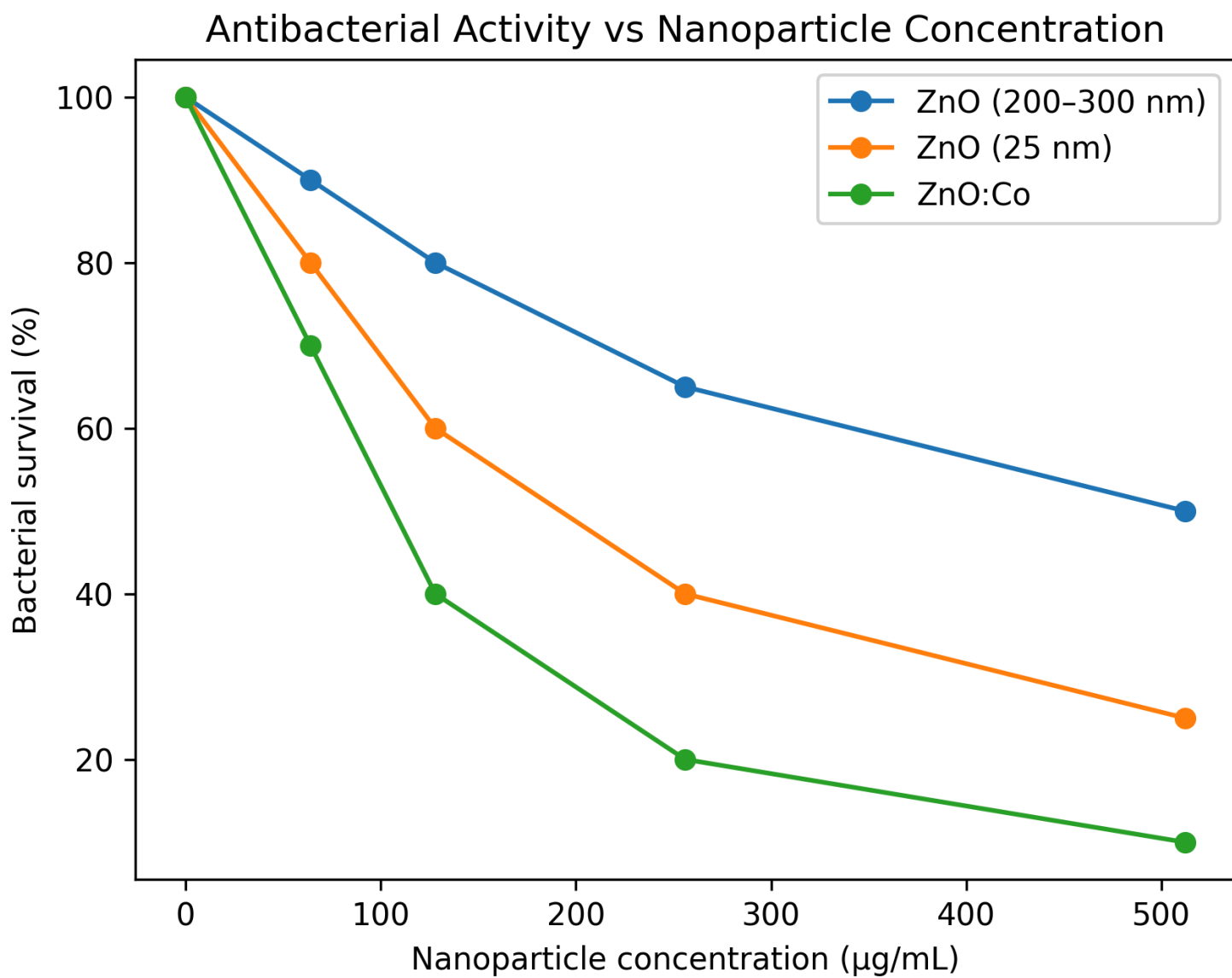


**Figure 2**

Experimental antibacterial activity. Antibacterial activity vsNanoparticle concentration.

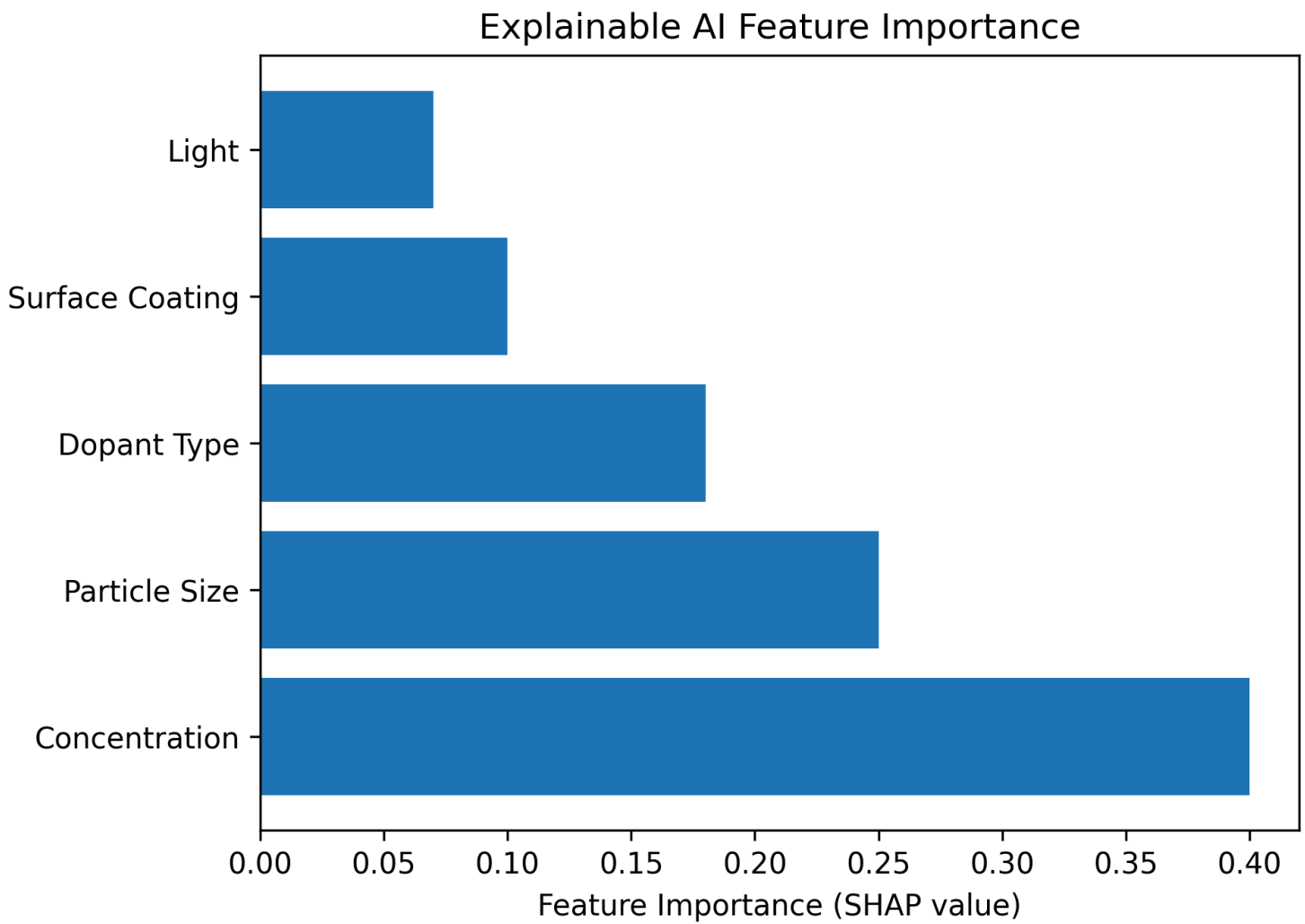


**Figure 3**

SHAP feature importance shows key drivers: Nanoparticle concentration, Particle size, Dopant chemistry, Surface coating, Illumination(light)

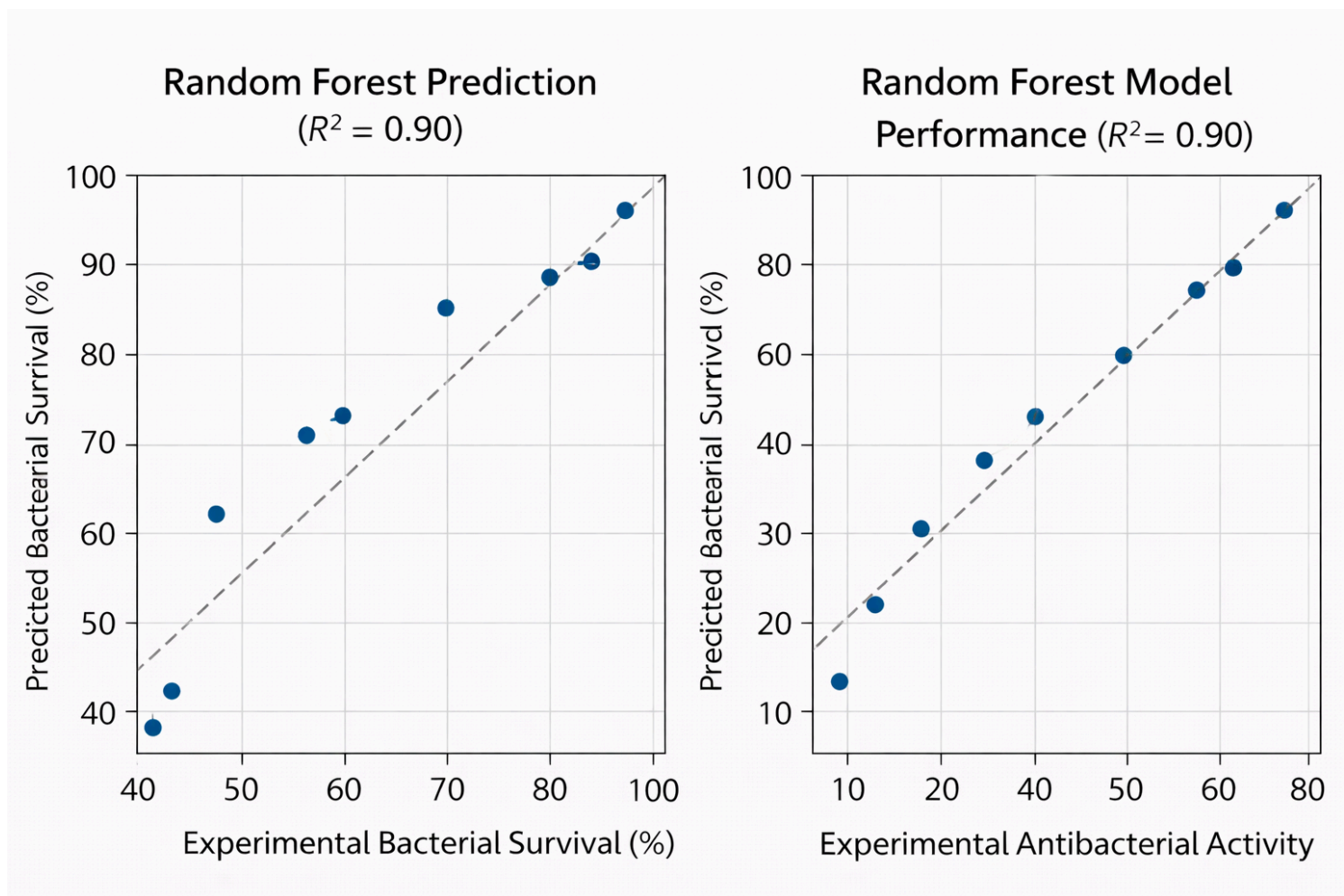


**Figure 4**

Random forest Prediction ( $R^2=0.90$ ). Random Forest predicted vs experimental antibacterial activity.

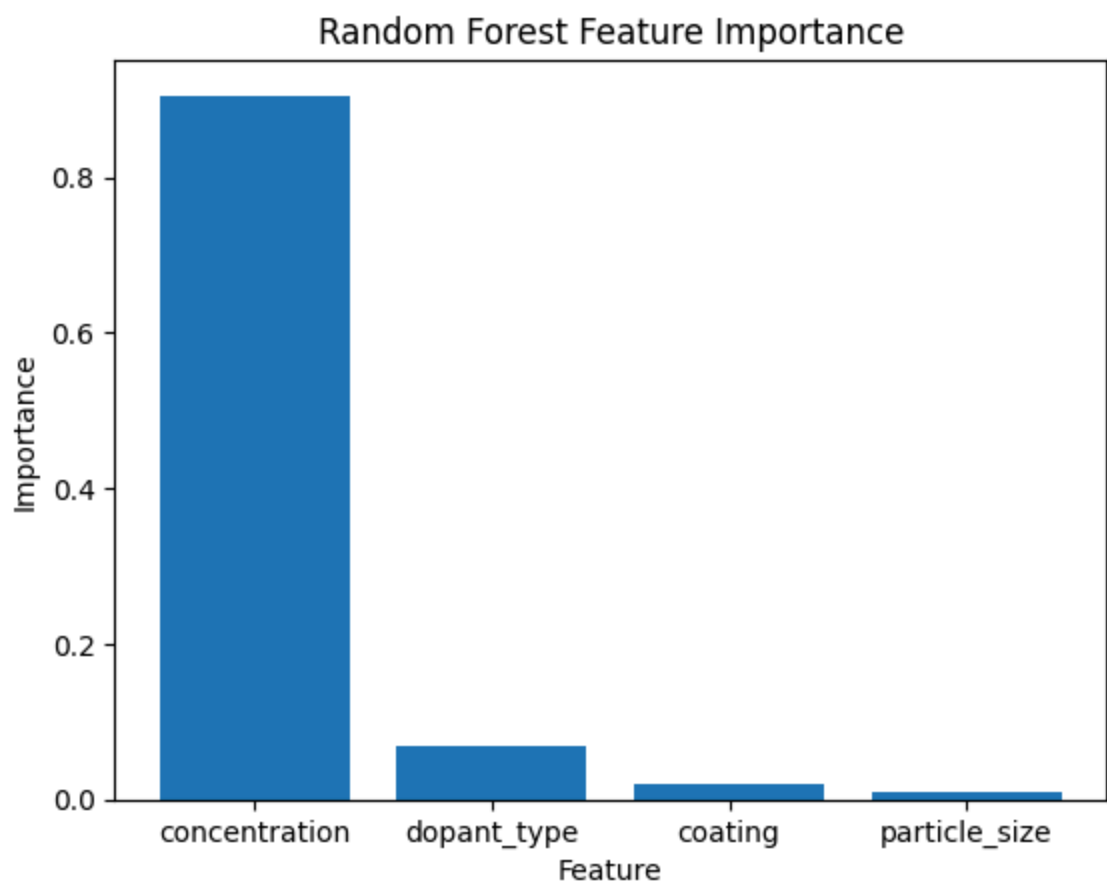


**Figure 5**

Random forest Feature importance ranking.

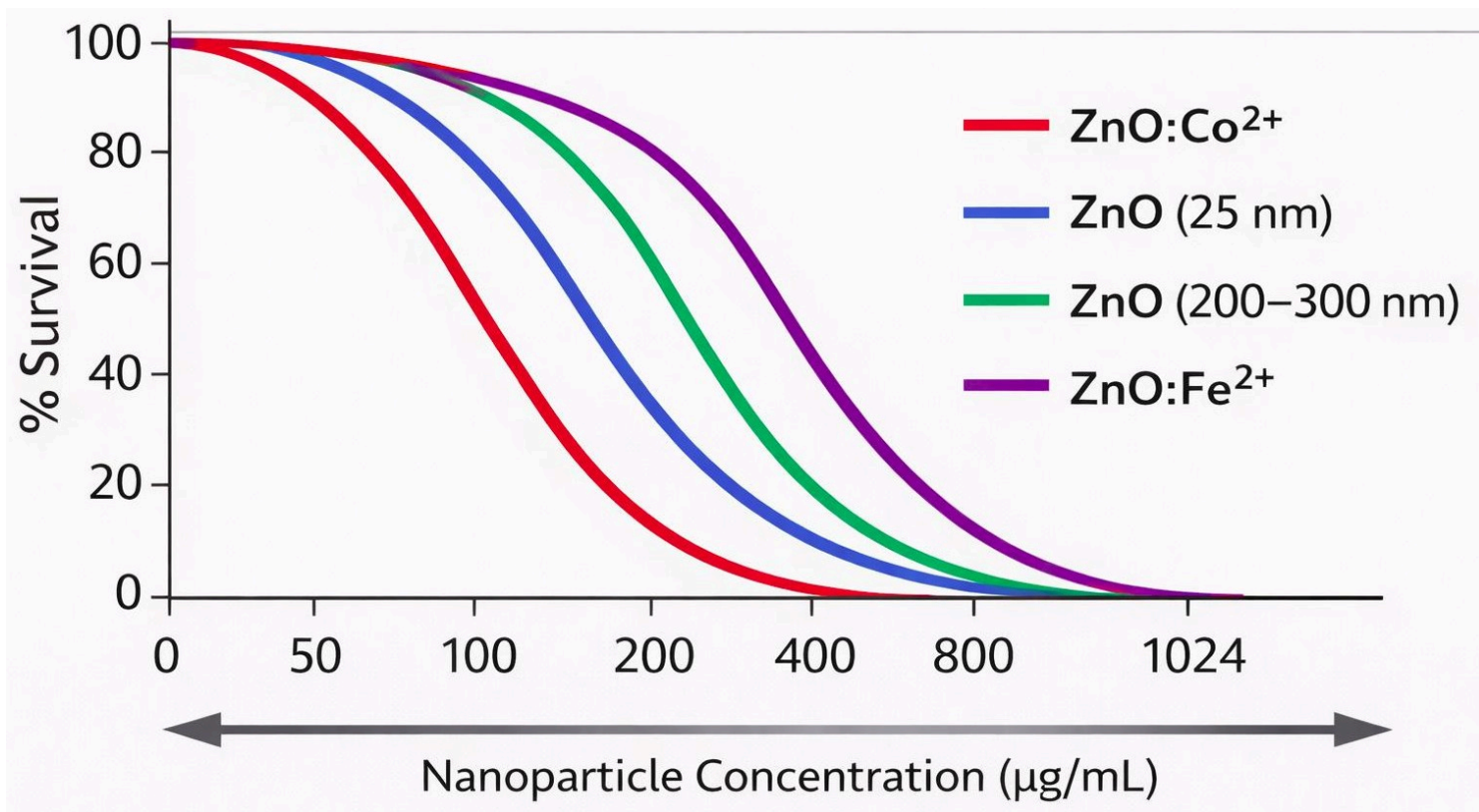


Figure 6

Predicted Bacterial Survival.

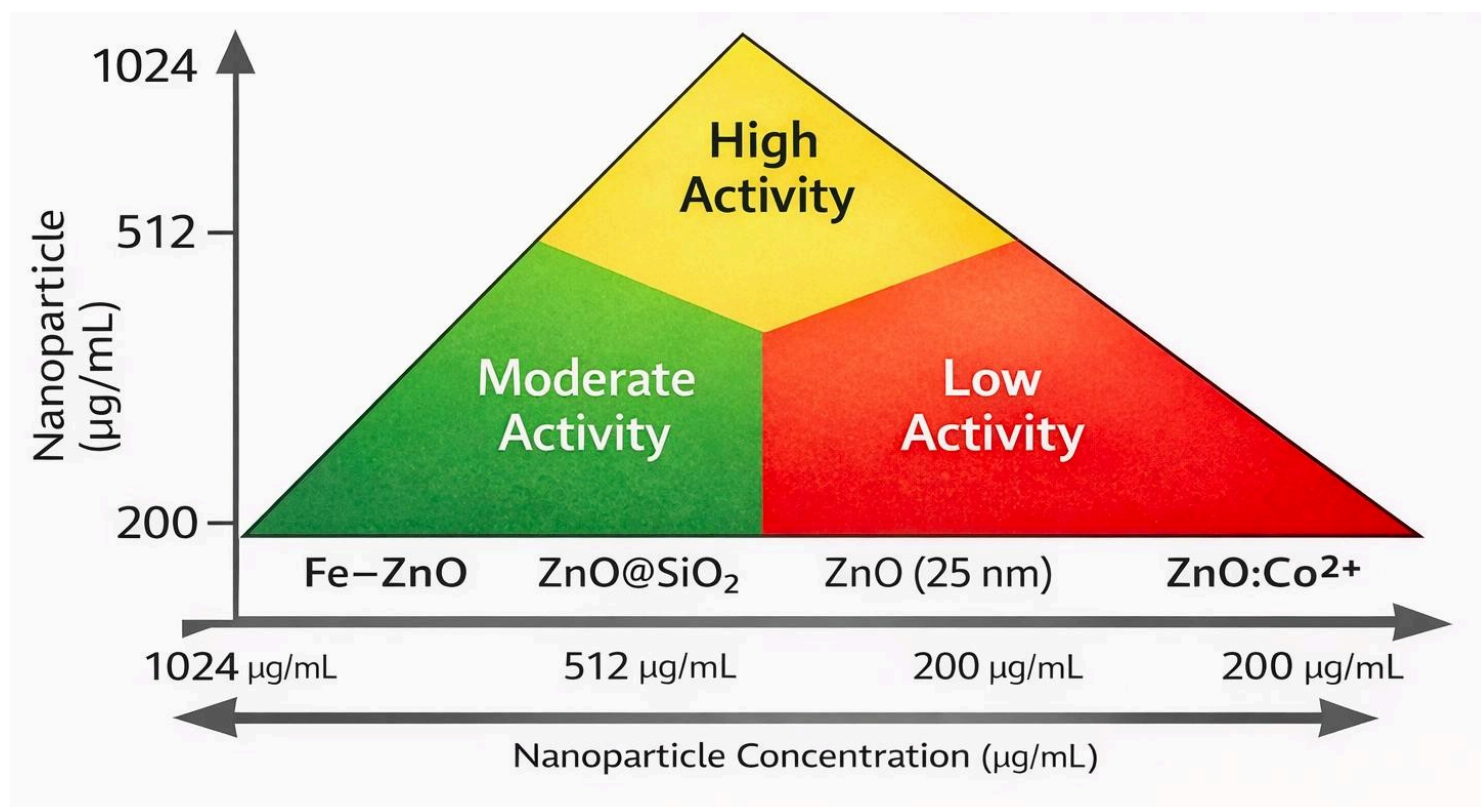


Figure 7

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GraphicalAbstract.docx](#)