

Cite this: *Biomater. Sci.*, 2024, **12**, 2229

## Machine learning in nanozymes: from design to application

Yubo Gao,<sup>†a,b</sup> Zhicheng Zhu,<sup>†a</sup> Zhen Chen,<sup>a</sup> Meng Guo,<sup>b</sup> Yiqing Zhang,<sup>a</sup> Lina Wang<sup>\*b</sup> and Zhiling Zhu  <sup>\*a</sup>

Nanozymes, a distinctive class of nanomaterials endowed with enzyme-like activity and kinetics akin to enzyme-catalysed reactions, present several advantages over natural enzymes, including cost-effectiveness, heightened stability, and adjustable activity. However, the conventional trial-and-error methodology for developing novel nanozymes encounters growing challenges as research progresses. The advent of artificial intelligence (AI), particularly machine learning (ML), has ushered in innovative design approaches for researchers in this domain. This review delves into the burgeoning role of ML in nanozyme research, elucidating the advancements achieved through ML applications. The review explores successful instances of ML in nanozyme design and implementation, providing a comprehensive overview of the evolving landscape. A roadmap for ML-assisted nanozyme research is outlined, offering a universal guideline for research in this field. In the end, the review concludes with an analysis of challenges encountered and anticipates future directions for ML in nanozyme research. The synthesis of knowledge in this review aims to foster a cross-disciplinary study, propelling the revolutionary field forward.

Received 31st January 2024,  
Accepted 6th March 2024

DOI: 10.1039/d4bm00169a

rsc.li/biomaterials-science

### 1. Introduction

Nanozymes are a class of nanomaterials that exhibit enzyme-like activity and conform to the enzymatic reaction kinetics.<sup>1</sup> Unlike natural enzymes, nanozymes possess unique physicochemical properties inherent to nanomaterials due to their distinctive nanoscale structures and have the ability to catalyze substrate reactions similar to those of enzymes in nature. The catalytic activity of nanozymes can be optimized by tuning their composition and structure.<sup>2</sup> In addition, the unique multi-enzyme-like activity of nanozymes provides the possibility of designing inexpensive, stable, and a wide variety of new catalytic cascade reactions.<sup>3</sup> This convergence has yielded nanozymes with exceptional stability, cost-effectiveness, and adjustable catalytic activity, positioning them as promising substitutes for natural enzymes.<sup>4</sup> Since the seminal discovery in 2007 by Yan's group, showcasing the peroxidase (POD)-like activity of ferromagnetic Fe<sub>3</sub>O<sub>4</sub> nanoparticles,<sup>5</sup> the scientific community has been captivated by the potential of nanozymes. To date, over 70 countries and 400 research institutions have dedicated their efforts to the development and application of nanozymes.<sup>6</sup> The versatile applications of nanozymes span a multitude of fields, including biomedicine,<sup>7–9</sup> analysis and sensing,<sup>10–12</sup> and environmental management.<sup>13–15</sup>

The diversity in the materials used for nanozymes, ranging from metal oxide nanoparticles<sup>16</sup> and metal nanoparticles<sup>17</sup> to intricate single-atom X–N–C mimetic architectures<sup>18</sup> and biomolecular assemblies,<sup>19,20</sup> underscores the rapid expansion in

<sup>a</sup>College of Materials Science and Engineering, Qingdao University of Science and Technology, Qingdao, Shandong 266042, China. E-mail: zlzhu@qust.edu.cn

<sup>b</sup>College of Environment and Safety Engineering, Qingdao University of Science and Technology, Qingdao, Shandong 266042, China. E-mail: linawang@qust.edu.cn

<sup>†</sup>These authors contributed equally to this work.



Zhiling Zhu

Zhiling Zhu is an associate professor at Qingdao University of Science and Technology. He received his BS degree from Central South University and PhD degree from the University of Houston (advisor: Professor Chengzhi Cai). He started his independent career at Qingdao University of Science and Technology. His research interests are focused on the data-driven rational design of bioactive materials, nanocatalytic biomedical materials, and nanoparticle drug delivery systems and medical devices.

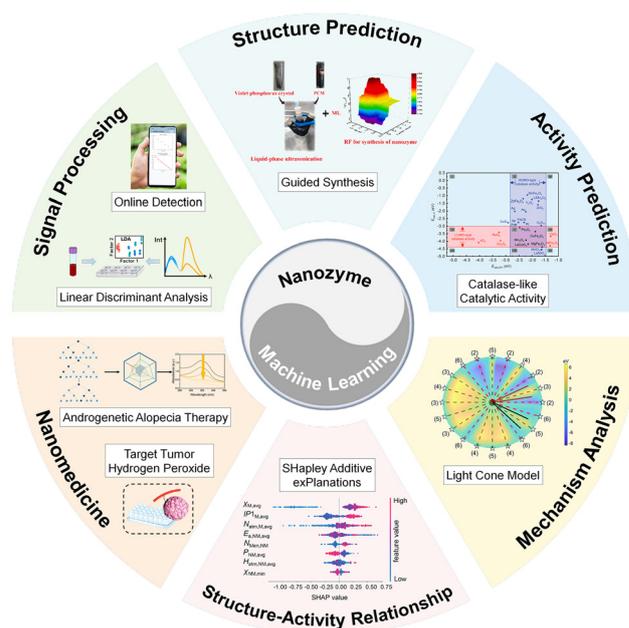
both the quantity and variety of enzyme-like species. Nanozymes, mimicking oxidoreductases,<sup>21</sup> hydrolases,<sup>22</sup> lyases,<sup>23</sup> and isomerases,<sup>24</sup> exhibit a remarkable diversity that transcends traditional enzyme functionalities. This diversity has paved the way for the realization of artificial enzymes that surpass their natural counterparts.<sup>25–27</sup> However, as nanozyme research advances, the conventional trial-and-error approach to discovering novel high-performance nanozymes is becoming increasingly challenging. The intricate design of high-performance nanozymes necessitates consideration not only of the physicochemical properties of the materials themselves, but also of the intricate interplay between these materials and their environments. Nanozymes operating in complex environments face the presence of various chemicals that may compete for binding to the active sites, potentially diminishing the catalytic efficiency. To grapple with the escalating complexity in the design and application of nanozymes, researchers are turning their attention to innovative research methodologies, moving beyond traditional approaches to explore new paradigms.

Machine learning (ML), as a powerful tool for statistical data analysis, utilizes algorithms to analyse and learn from data in order to discover patterns and regularities within the data.<sup>28</sup> Through these processes, ML algorithms can continuously optimize and improve as new data are introduced, enabling computers to learn from experience and make decisions or predictions based on the learned content.<sup>29</sup> In the realm of materials science, ML has proven highly effective, playing a pivotal role in guiding materials synthesis,<sup>30</sup> facilitating materials characterization,<sup>31</sup> predicting materials properties,<sup>32</sup> and elucidating intricate structure–activity relationships.<sup>33</sup> Notably, ML, as a potent tool for statistical data analysis, has recently found its stride in the field of nanozyme research. This has opened up new avenues for tackling challenges related to nanozyme design, performance analysis, and the promotion of applications.<sup>34,35</sup>

This review is dedicated to an in-depth examination of the multifaceted roles that ML plays in the design and application of nanozymes, offering insights into recent findings in related domains (Fig. 1). Initially, the review delves into the ways in which ML contributes to the rational design of nanozymes. Subsequently, it encapsulates diverse application areas where ML is applied to nanozyme design and delineates a structured roadmap for leveraging ML in the design of high-performance nanozymes. The concluding sections of the review address current challenges and illuminate future trends in the application of ML to nanozymes. This comprehensive review not only consolidates the current research landscape in the ML–nanozyme intersection, but also introduces statistical analysis methods utilizing ML. The aim is to catalyse interdisciplinary collaborations, fostering further advancements in this revolutionary field.

## 2. ML-assisted design of nanozymes

To date, the evolution of nanozyme research has traversed four paradigmatic stages: scientific experimentation, theoretical



**Fig. 1** Research achievements of the combination of ML and nanozymes in recent years. Reprinted with permission from ref. 40. Copyright 2023 Elsevier. Reprinted with permission from ref. 42. Copyright 2022 Wiley-VCH GmbH. Reprinted with permission from ref. 46. Copyright 2023 Wiley-VCH GmbH. Reprinted with permission from ref. 44. Copyright 2022 American Chemical Society. Reprinted with permission from ref. 51. Copyright 2021 American Chemical Society.

science, computational science, and the current era of big data science.<sup>36</sup> The scope of nanozyme research has expanded significantly, marked by the increasing scale of material systems. Notably, the exploration of unknown nanozyme activities in existing materials, surface modification and doping of established nanozymes,<sup>37</sup> and the construction of composite-based nanozymes have long constituted the primary focus of nanozyme design.<sup>38</sup> However, efficiently designing nanozymes has emerged as a crucial and formidable challenge amidst the escalating complexity of nanozyme applications.<sup>39</sup> The diminishing cost of computational resources, the enhancement of material databases, and the rapid strides in artificial intelligence (AI) offer a ray of hope for streamlining the design of nanozymes. This section will unfold on four dimensions of ML-guided nanozyme research, encompassing guiding the synthesis of nanozymes, predicting nanozyme activity, deciphering the structure–activity relationships of nanozymes, and navigating the search for optimal nanozyme catalytic reaction pathways.

### 2.1 ML-guided synthesis of nanozymes

In the course of nanozyme synthesis, adjustments to synthesis conditions can profoundly influence the nanozyme structure, consequently impacting its performance. Harnessing the power of ML, optimal experimental conditions can be discerned through the analysis of performance data collected from nanozymes synthesized under varying conditions. For example, Ge *et al.* constructed highly stable violet phosphorene decorated with phosphorus-doped hierarchically porous

carbon microspheres (VP-PCMs) based on ML-guided experimental data.<sup>40</sup> The current values generated by VP-PCMs under different concentrations of violet phosphorene (VP) and porous carbon microspheres (PCMs) were collected for meconium phenolic acid (MPA) detection. Employing this data, a random forest (RF) model was developed to predict the optimal concentration of VP and PCM during VP-PCM synthesis. The model utilized the concentrations of VP and PCM as inputs and current values as outputs, yielding optimal concentrations of 0.2 mg mL<sup>-1</sup> for VP and 1.4 mg mL<sup>-1</sup> for PCM. This optimized VP-PCM nanozyme design demonstrated superior affinity for MPA, achieving Michaelis constant ( $K_m$ ) = 12.4 μM. This ML-driven methodology, utilizing synthesis conditions as input and assay excellence as output, can be extended to diverse synthesis procedures with different material systems. Such an approach holds immense potential for the precise and efficient synthesis of nanozymes, thereby offering valuable guidance in the field.

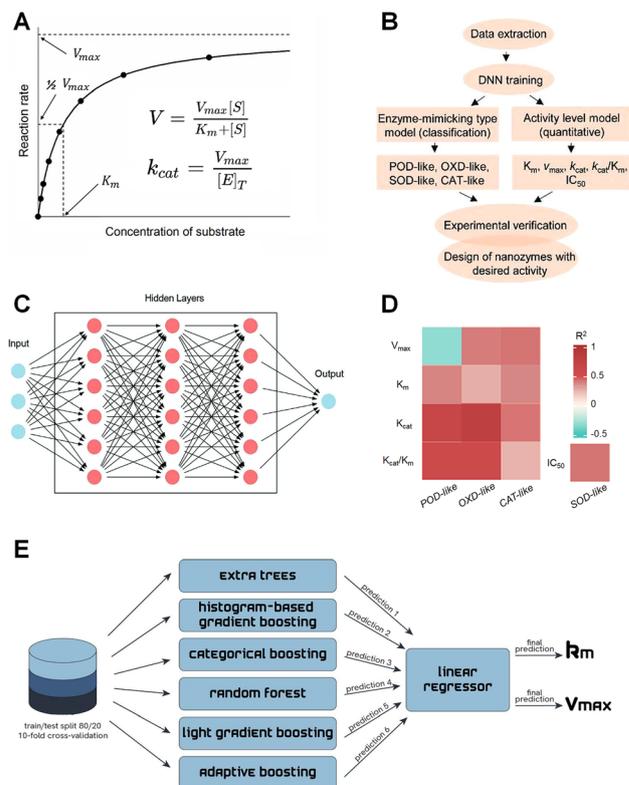
## 2.2 Prediction of nanozyme activity

The prediction of nanozyme activity stands as a pivotal and foremost research focus in the application of ML to nanozyme design. The establishment of an accurate prediction model holds the key to foreseeing potential enzyme-like activities in designed nanozymes, allowing for informed predictions based on the model.

### 2.2.1 Prediction of kinetic parameters of the Michaelis–Menten equation.

The Michaelis–Menten equation, a kinetic equation depicting the initial rate of enzymatic reactions in natural enzymes concerning substrate concentration, is a fundamental tool in enzyme kinetics (Fig. 2A). Remarkably, akin to natural enzymes, the catalytic kinetic curves of nanozymes align with the principles of the Michaelis–Menten equation. Leveraging ML, predictive models can be employed to estimate the kinetic parameters integral to the Michaelis–Menten equation. These parameters include the maximum reaction rate ( $V_{max}$ ), the Michaelis constant ( $K_m$ ), and the turnover number ( $K_{cat}$ ), all of which serve as crucial indicators of the catalytic activity of nanozymes.

Razlivina *et al.* conducted an extensive data collection effort from over 100 published papers, culminating in the establishment of a nanozyme database (Dizyme).<sup>41</sup> This database comprehensively incorporates various parameters, including component properties (*e.g.*, electronegativity, electron affinity, oxidation state, and ionic radius), material characterization (*e.g.*, surface charge, stability, surface adsorption, and surface area), and reaction conditions (*e.g.*, pH, temperature, substrate type, nanozyme concentration, substrate concentration). Employing the random forest regression (RFR) model, they achieved highly accurate predictions of nanozyme kinetic activity. For  $\lg K_{cat}$  and  $\lg K_m$ , the coefficients ( $R^2$ ) of determination reached 0.52 and 0.80, respectively. Wei *et al.* gathered activity data from over 300 papers and identified eight endogenous factors and three exogenous factors influencing nanozyme activity.<sup>42</sup> Utilizing these factors as variables, they constructed fully connected deep neural network (DNN) models to predict various



**Fig. 2** Prediction of kinetic parameters in the Michaelis–Menten equations. (A) Michaelis–Menten equations. (B) The workflow for classification and quantitative prediction of enzyme-like activity of nanomaterials using ML. (C) Schematic diagram of fully connected DNN-based models. (D) Heatmap images of prediction accuracies for 13 DNN-based quantitative model based on  $R^2$ . Each box represents one model built from a certain dataset, estimated by  $R^2$ . Reprinted with permission from ref. 42. Copyright 2022 Wiley-VCH GmbH. (E) Stacking ensemble algorithm scheme for triple-catalytic (POD-, OXD-, and CAT-like) activity prediction. Reprinted with permission from ref. 43. Copyright 2023 Research Square.

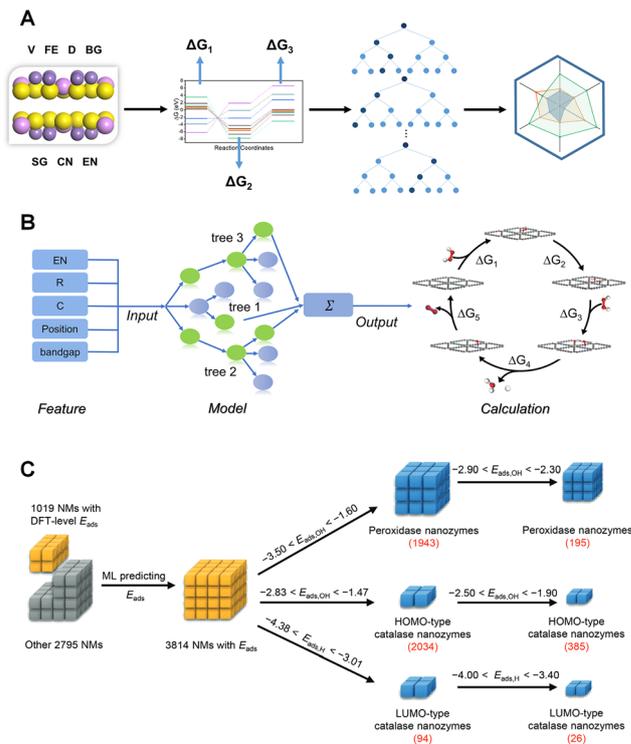
nanozyme types (POD-, OXD-, CAT-, SOD-like activities) and nanozyme activities ( $K_m$ ,  $V_{max}$ ,  $K_{cat}$ ,  $K_{cat}/K_m$ , and  $IC_{50}$ ) (Fig. 2B–D). For predicting the activity levels of POD- and OXD-like nanozymes, the  $R^2$  values reached 0.66 and 0.80, demonstrating robust performance. Furthermore, Vinogradov *et al.* expanded the data in Dizyme to 1210 samples, incorporating additional descriptors such as molecular weight, topological and electronic coating descriptors, synthesis details, and assay conditions.<sup>43</sup> Employing a stacked integrated learning approach, which combines multiple ML models, they ultimately selected a linear regression model as the *meta*-model to predict POD-, OXD-, and CAT-like activities of nanozymes (Fig. 2E).

It is noteworthy that while predicting the kinetic parameters in the Michaelis–Menten equation can enhance the accuracy of determining nanozyme activity, the experimental determination of these kinetic parameters remains a prerequisite. Predicting the activity of numerous yet undiscovered materials with enzyme-like characteristics is challenging given

this experimental dependency. Consequently, the method of characterizing nanozyme activity by predicting the kinetic parameters of the Michaelis–Menten equation, which subsequently informs nanozyme design, is significantly constrained by the size of the available nanozyme databases.

**2.2.2 Prediction of energy changes during the catalytic process.** Various energy changes, including reaction Gibbs free energy, activation energy, and adsorption energy, manifest during the catalytic process of nanozymes. These energy changes serve as descriptors for nanozyme activity, offering insights into the pace of chemical reactions. To address the challenge of numerous nanomaterials with undiscovered enzyme-like activities, researchers can employ first-principles calculations, such as density-functional theory (DFT), and leverage the high computing power of computers to efficiently obtain the energy changes of nanozyme catalytic reactions in high-throughput. By integrating this data with ML techniques and material databases, the prediction of energy changes during the catalytic process for unknown enzyme-like active materials in a broader chemical space becomes feasible and thus facilitates the screening of highly active nanozymes.

Zhang *et al.* identified transition metal thiophosphates ( $M_xP_yS_z$ ;  $x = 1-7$ ,  $y = 1-4$ ,  $z = 1-29$ ), characterized by mixed valence states formed by different bonding states, as potential candidates for SOD-like nanozymes.<sup>44</sup> Employing DFT calculations, they obtained the Gibbs free energies of the first three steps of the catalytic reaction for these materials. Utilizing seven parameters (*e.g.*, electronegativity, dopant atom radius, dopant atom position, dopant atom concentration, and band gap) as input data, and  $\Delta G$  of the first three steps of the SOD-like catalytic reaction as output data, they predicted the changes in the Gibbs free energy for the SOD-like catalytic reaction using an RF model. This approach led to the identification of a new highly active SOD-like nanozyme,  $MnPS_3$  (Fig. 3A). Yu *et al.* explored 14 different nonmetallic atoms doped into graphdiyne (GDY) as potential materials, calculating the Gibbs free energy of the catalytic reaction for different materials using DFT.<sup>45</sup> With input data including electronegativity, dopant atom radius, dopant atom position, dopant atom concentration, and band gap, and output data representing the maximum energy-consuming step and maximum energy barrier of the POD-like catalysed reaction, they employed the extreme gradient boosting (XGB) algorithm to screen highly active POD-like nanozymes. This led to the identification of boron-doped GDY (B-GDY) and nitrogen-doped GDY (N-GDY) as highly active POD-like nanozymes (Fig. 3B). Gao *et al.* utilized the extreme gradient boosting regression (XGBR) algorithm with 11 basic atomic features, including electronegativity and electron affinity energy, as input data for predicting  $E_{ads,OH}$  and  $E_{ads,H}$ .<sup>46</sup> They screened 60 nanozymes with high POD- and CAT-like activities from the computational 2D materials database (C2DB). The predictions were found to be consistent with DFT calculations (Fig. 3C). These studies highlight that ML algorithms for predicting and screening nanozymes can significantly reduce the time required compared to exclusive DFT calculations.



**Fig. 3** Screening high activity nanozymes by predicting energy change. (A) The workflow of ML-assisted screening and prediction of nanozymes. Reprinted with permission from ref. 44. Copyright 2022 American Chemical Society. (B) Prediction of the reaction maximum energy barrier and the energy consuming step using ML models. Reprinted with permission from ref. 45. Copyright 2022 American Chemical Society. (C) Computational screening of POD- and CAT-like nanozymes with the corresponding adsorption energy criteria and the more stringent criteria. Reprinted with permission from ref. 46. Copyright 2023 Wiley-VCH GmbH.

The prediction of energy changes during nanozyme catalysed reactions relies on first-principles calculations. By calculating the energy changes during nanozyme catalysed reactions for a large number of nanomaterials with similar or identical structures and integrating the results obtained from these calculations with ML models, it becomes feasible to predict the enzyme-like activity of a specific material. Experimental determination of the kinetic parameters of the Michaelis–Menten equation is susceptible to errors due to various factors, such as the lack of standardized procedures. In contrast, computational predictions of energy changes during nanozyme catalysis typically utilize uniform theoretical models and computational programs, reducing the variability associated with experimental procedures and enhancing the reproducibility and comparability of the calculated data. However, the accuracy of computational models strongly relies on the realism of the assumptions and parameter settings of the models, which may not fully capture all the complex interactions and environmental conditions under different experimental settings. Therefore, experimental data provide measured data of the actual reaction process, while computational data offer theoretic

tical predictions and support for analysis. More and more scientists combine both approaches to gain a more comprehensive understanding of the kinetic behaviour of nanozymes. Therefore, predicting energy changes during nanozyme catalysed reactions holds greater promise for accurately designing new highly active nanozymes.

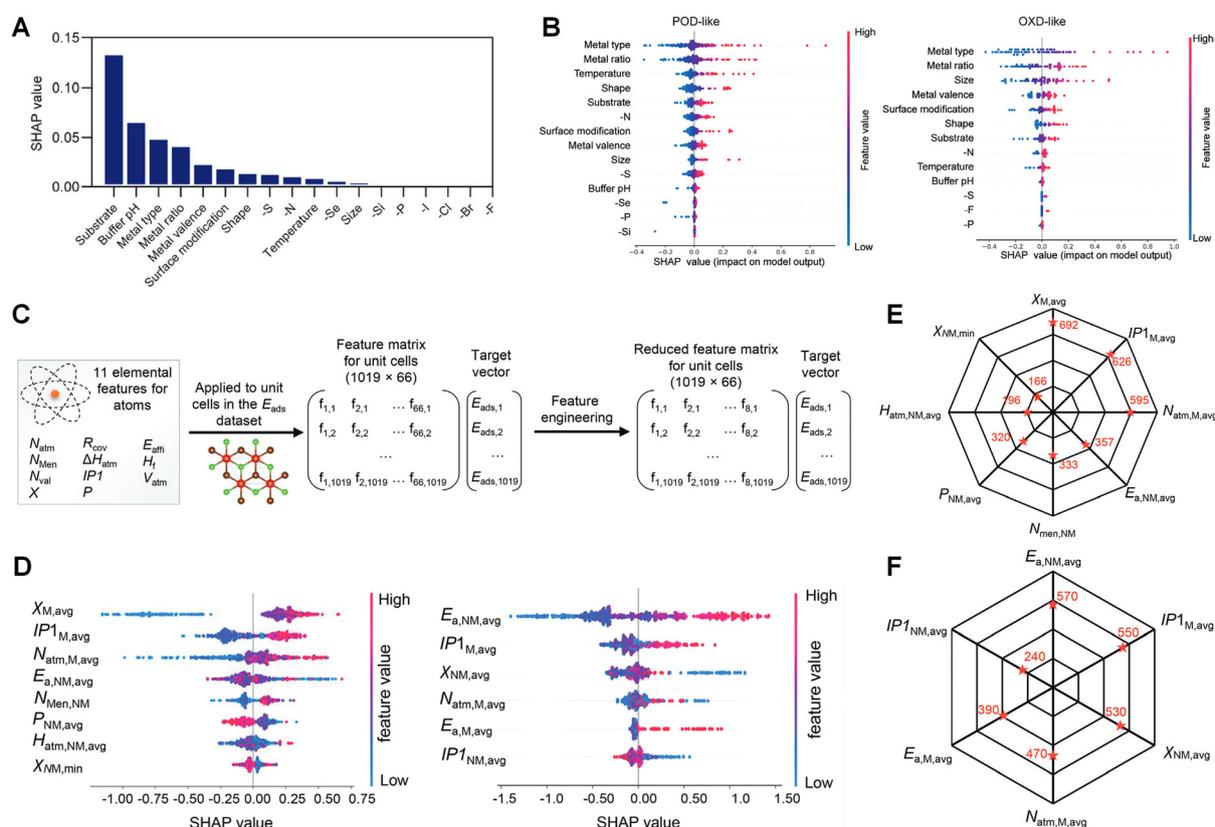
### 2.3 Understanding the structure–activity relationships of nanozymes

ML excels in uncovering implicit relationships between the structure and activity of materials within a multidimensional space, providing insights into the structure–activity relationships of materials. An indispensable element of interpretable ML models is the feature importance analysis, which allows for the measurement of the specific contribution of each input feature to the predicted results of the models, facilitating an understanding and explanation of the ML model.

Wei *et al.* employed SHapley Additive exPlanations (SHAP) to assess the importance of features in the collected data (Fig. 4A) during the development of their ML model for predicting the type and activity of nanozymes.<sup>42</sup> By examining the magnitude of SHAP values, the influence of each feature on the type of nanozyme activity was gauged. Notably, metal type,

metal proportion, and metal valence emerged as the top three internal factors influencing the type of nanozyme activity (Fig. 4B). These findings underscore the pivotal role of adjusting the metal composition in nanozymes for their rational design. Furthermore, through SHAP analysis, the study revealed that altering the type of metallic elements holds greater significance than changing the type of nonmetallic elements in modulating the type of nanozyme activity.

In the development of ML models for predicting 2D materials'  $E_{\text{ads,OH}}$  and  $E_{\text{ads,H}}$ , Gao *et al.* gathered 66 features for 1019 materials, resulting in two  $1019 \times 66$  feature matrices (Fig. 4C).<sup>46</sup> They utilized the XGBoost regression (XGBR) algorithm for feature selection, optimizing the number of features. Subsequently, SHAP was employed for feature importance analysis, revealing the underlying physical laws within the ML model (Fig. 4D). Notably, average electronegativity ( $\chi_{\text{M,avg}}$ ) proved to be the most significant for  $E_{\text{ads,OH}}$ , while the average electron affinity of non-metallic elements ( $E_{\text{a,NM,avg}}$ ) had the greatest impact on  $E_{\text{ads,H}}$  (Fig. 4E and F). The study also unveiled the substantial influence of the electronegativity or first ionization energy of metallic elements on electron transfer capacity, influencing the OH adsorption energy. The non-concentrated distribution of SHAP values indicates a lack of a



**Fig. 4** Understanding the structure–activity relationship of nanozymes through model interpretability. (A) SHAP analysis of the sensitivity of different factors in the classification model. (B) SHAP sensitive analysis of the independent variables in the POD- and OXD-like quantitative models. Reprinted with permission from ref. 42. Copyright 2022 Wiley-VCH GmbH. (C) Feature engineering to select the most important features of adsorption energies. (D) SHAP sensitive analysis of the independent features in the ML models. (E and F) Selected features for hydroxyl ( $E_{\text{ads,OH}}$ ) and hydrogen ( $E_{\text{ads,H}}$ ) adsorption energies with corresponding importance scores. Reprinted with permission from ref. 46. Copyright 2023 Wiley-VCH GmbH.

simple linear relationship between characteristics and adsorption energy. This comprehensive analysis not only enhances the understanding of the relationship between composition and nanozyme activity, but also holds crucial guiding significance for the rational design of nanozymes.

Examining the interplay between atomic and electronic structures and the catalytic properties of nanozymes at the microscopic level is a crucial step in designing high-performance nanozymes. However, elucidating the structure–activity relationships of nanozymes through empirical means poses significant challenges. Interpretable ML modeling not only aids in comprehending the structure–activity relationships of nanozymes but is also anticipated to provide more effective guidance for the design of nanozymes.

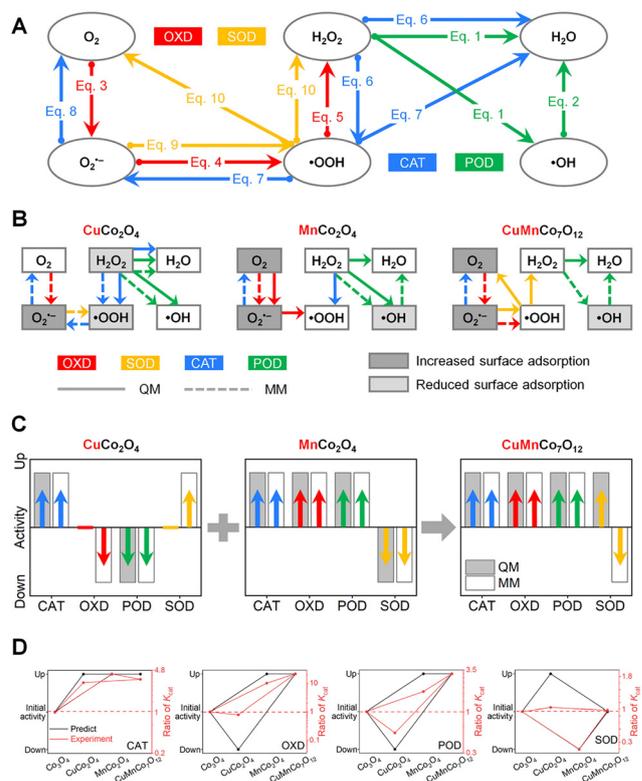
#### 2.4 Searching for optimal catalytic reaction paths of nanozymes

The challenge of uncertainty in reaction paths complicates the study of the catalytic mechanism of nanozymes. Algorithmic determination of potential energy surfaces by calculating energies of possible intermediates with corresponding transition states can address this issue. ML plays a role in searching for potential energy surfaces, with neural network fitting of the potential function being a particularly noteworthy approach. Xu *et al.* employed the random surface walk method combined with a neural network (SSW-NN) to simulate the reaction system of nanozymes.<sup>47</sup> They obtained a series of adsorption configurations after the decomposition of  $\text{H}_2\text{O}_2$  stabilized on a 2D  $\text{MoS}_2$  surface. Utilizing the double end-surface walk (DESW) method for localizing transition states and finding reaction pathways, they identified a critical rate-determining step in the decomposition of  $\text{H}_2\text{O}_2$  on the material surface, specifically the homogenization of  $\text{H}_2\text{O}_2$  on the material surface. To enhance the POD-like activity of 2D  $\text{MoS}_2$ , they conducted Cu single-atom loading, sulfur vacancy engineering, and pH environment modulation based on the reaction mechanism. These interventions successfully reduced the reaction energy barrier of the critical step, resulting in an improved performance of the nanozymes.

Compared to nanozymes with a single class of enzyme-like activity, the catalytic mechanism of multienzyme-like nanozymes is more intricate, involving different reaction paths and interactions between various enzyme-like activities in cascade, promotion, and antagonism. This complexity poses challenges for theoretically guided design. However, by combining the methods of ML and molecular simulation calculations to explore optimal nanozyme catalytic reaction pathways, the rational design of multienzyme-like nanozymes becomes feasible. Recently, Jiang *et al.* established a comprehensive nanozyme dataset by collating information from 4159 papers, encompassing element types, element ratios, chemical compounds, shapes, and pH values.<sup>48</sup> Based on this dataset, they reorganized the material features of different nanozymes using clustering correlation coefficients of nanozyme features to derive the constituent factors of multienzyme-like nanozymes. Subsequently, they developed a methodology that integrates

quantum mechanics/molecular mechanics (QM/MM) and ML to analyse surface adsorption and desorption energies, as well as binding energies of substrates, transition states, and products in the multienzyme-like reaction pathway. This approach enabled the determination of optimal reaction pathways, leading to the design of a genetically evolved class of multienzyme-like nanozymes (Fig. 5). The outcome was a multienzyme-like nanozyme,  $\text{CuMnCo}_7\text{O}_{12}$ , exhibiting high POD-, CAT-, OXD-, SOD-like activities. Noticeably, the authors used molecular dynamics simulations of MM along with QM calculations, an approach that incorporates environmental factors and allows for a more comprehensive simulation of the effects of environmental factors on the catalytic reactions of nanozymes.

In sum, ML is a reliable method for optimizing the catalytic reaction pathways of nanozymes, offering the potential for the rational design of more high-performance nanozymes in the future.



**Fig. 5** Computational and experimental results of data-driven evolutionary design research on multienzyme-like nanozymes. (A) Illustration of the catalytic reaction paths catalyzed by multienzyme-like nanozymes. (B) Reaction path and adsorption and desorption changes of each generation of nanozyme compared to the previous generations. (C) Comparative assessment of the multienzyme-like activities of the second- and third-generation nanozymes in contrast to the first-generation nanozyme. (D) Comparison of the CAT-, OXD-, POD-, and SOD-like activities of the evolutionarily designed multienzyme-like nanozymes. Reprinted with permission from ref. 48. Copyright 2024 American Chemical Society.

## 2.5 Summary

Moving from the prediction of kinetic performance by ML to the prediction of energy changes during the reaction process, it reflects the research trend of nanozymes evolving from performance analysis to mechanism understanding. As the study of nanozymes becomes more profound and complex data become prevalent, ML, as a potent tool for statistical data analysis, is expected to find even broader applications. ML-assisted optimization of catalytic reaction pathways of nanozymes is essential for handling the statistical analysis of complex data. However, challenges arising from ML in the design studies of nanozymes persist and require further exploration and resolution.

**2.5.1 Dataset issues.** The current datasets related to nanozymes are predominantly derived from published papers, presenting several challenges:

(1) *Small datasets.* Nanozyme research is relatively new, resulting in datasets that are often small with insufficient sample sizes to train complex ML models. This limitation hinders the generalization ability of the model, impacting its capacity to predict unknown data.

(2) *Data imbalance.* Certain types of nanozymes may be studied more frequently than others in practice, leading to a dataset with a much larger number of samples for one type compared to others. This imbalance can bias ML model predictions in favour of the more numerous types, affecting performance on sparser types.

(3) *Data quality and consistency issues.* Data from published papers often depend on specific experimental conditions and methods. The environments in which the data are acquired may be inconsistent, resulting in uneven data quality that can affect the training of ML models.

(4) *Missing data.* Published papers tend to focus on the successful applications of nanozymes, with unsuccessful data often deliberately omitted. In ML, understanding both successful and unsuccessful outcomes is crucial for effective learning. The omission of failed data can impact the training effectiveness of ML models, resulting in overfitting to positive (successful) data points and lacking generalization ability. When training ML models, it is important to fully utilize these failed data points and appropriately handle them during the model construction process. Whether to weight them or exclude them will depend on the characteristics of the dataset and the specific goals of the model. Addressing these challenges is essential for enhancing the reliability and applicability of ML models in nanozyme research.

**2.5.2 Problems with ML algorithms.** As ML algorithms become more sophisticated, the interpretability of ML models is diminishing. Neural network models are often characterized by their opaque and challenging-to-understand “black-box” nature. This opacity hinders the understanding of the decision-making process within the models. In the realm of nanozyme design and development, it is not only crucial to predict results accurately but also imperative to comprehend the catalytic mechanisms of nanozymes. This understanding

is essential for the development of effective design strategies. Addressing the interpretability challenge in ML models, especially in the context of complex systems like nanozymes, remains an important area for improvement.

**2.5.3 The problem of activity descriptors.** Activity descriptors play a crucial role in simplifying the material intrinsic factors influencing complex catalytic reactions and can swiftly determine the catalytic performance of nanozymes. Incorporating catalytic reaction activity with physical quantities to derive meaningful activity descriptors has been a prominent focus in nanozyme research. Presently, nanozyme activity descriptors fall into three main categories:

(1) *Energy descriptors.* Typically, the catalytic activity of nanozymes is associated with the adsorption energy of reaction intermediates. For instance, Gao *et al.* demonstrated that descriptors like  $E_{\text{ads,OH}}$  and  $E_{\text{ads,H}}$  can be utilized for POD- and CAT-like activities of 2D nanozymes.<sup>46</sup> Zheng *et al.* recently extended the use of  $E_{\text{ads,OH}}$  to describe the POD-like activity of metal-based nanozymes.<sup>49</sup>

(2) *Electronic structure descriptors.* Wang *et al.* proposed and demonstrated that  $e_g$  occupancy can serve as a descriptor for the POD-like activity of oxide-based chalcogenides and spinel oxides.<sup>50,51</sup>

(3) *Geometrical structure descriptors.* Wang *et al.* constructed a series of heterogeneous molybdenum nanozymes ( $\text{Mo}_{\text{SA}}\text{-N}_x\text{-C}$ ), establishing a correlation between conformation and POD-like activity.<sup>52</sup> They found that the coordination number of Mo atoms can be a factor in the POD-like activity, serving as a descriptor for POD-like activity. However, the study of activity descriptors for nanozymes faces two key challenges: (1) The reported descriptors are predominantly focused on POD-like activity, with limited exploration of other activities; (2) The reported descriptors are often specific to particular material systems, with weak generalization ability. These limitations underscore the need for broader and more diverse investigations into activity descriptors for various nanozyme activities and material systems.

**2.5.4 Issues in ML simulation of catalytic environments.** In order to more comprehensively simulate the effects under actual reaction conditions and improve the model's generalization ability and prediction accuracy, ML typically treats catalytic environmental condition parameters as input data to predict the catalytic activity of nanozymes under different catalytic environmental conditions. The dataset should include feature variables related to the catalytic environment, such as temperature, pressure, pH, reactant, product concentrations, *etc.* Through this method, the key factors influencing the catalytic activity of nanozymes can be understood, leading to predictions of unknown reaction outcomes and the design of more efficient nanozymes. However, ML models rely heavily on a large amount of data for prediction. If there are insufficient descriptors and diverse data to describe the impact of the catalytic environment, the predictive capabilities of the model may be limited. Additionally, most of these data are obtained through experiments, which may result in human or instrumental data errors, instability, and interference.

### 3. ML in nanozyme applications

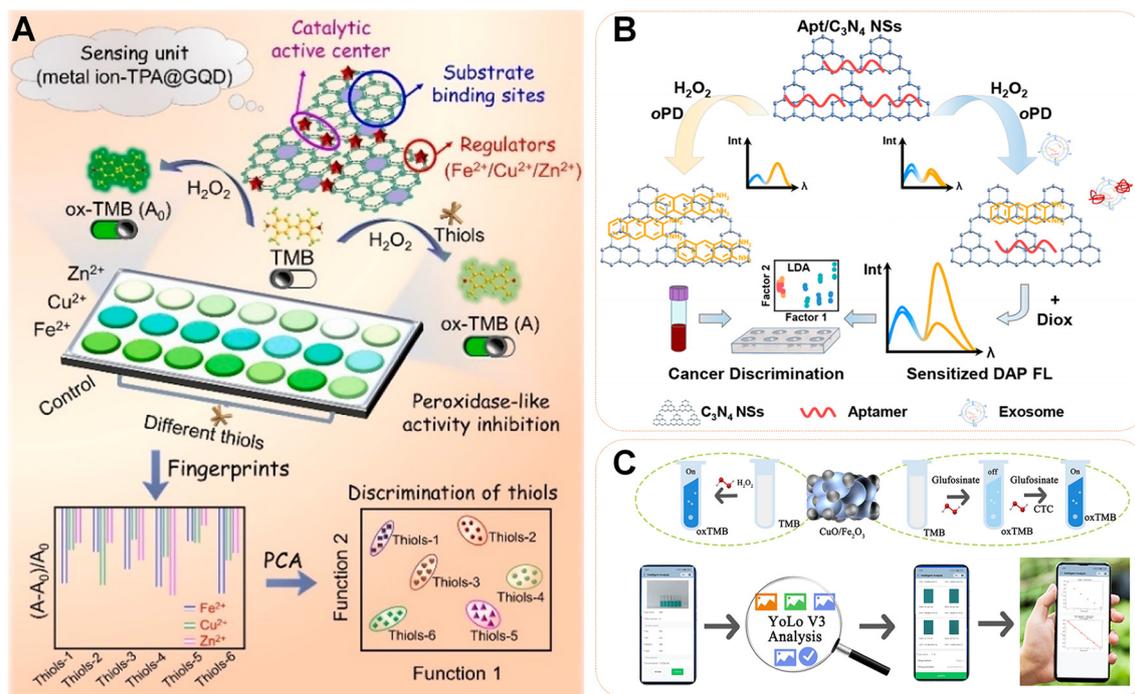
Nanozymes find most of their applications in analysis and sensing, constituting 52% of all nanozyme research. ML analyses various signals generated during the catalytic reaction of nanozymes, which enables efficient output of information, such as the type and concentration of target detectors. This section will focus on optical, electrochemical, and multimodal sensors for nanozymes based on the type of output signals. The application of ML in these types of sensors will be introduced respectively.

#### 3.1 ML in nanozyme-based optical sensing

The application of ML in nanozyme-based optical sensors focuses on analysing changes in optical signals induced by nanozyme catalysed reactions with specific substrates. These changes manifest as alterations in peak absorption, peak position, absorbance, fluorescence emission, fluorescence absorption, *etc.* ML, in comparison with simple optical instruments, can swiftly and accurately identify these optical signal changes. It achieves this by analysing the absorbance detected by the instrument or extracting information such as red, green, blue (RGB) or hue, saturation, value (HSV) in the analysed images. Additionally, optical detection technology can involve sensor arrays generating multi-signal input and output data, making ML analysis crucial for interpreting complex optical signals.

**3.1.1 Analysing substrate concentration based on absorbance.** Nanozymes serve as the foundation for constructing optical sensors by catalysing specific substrates to initiate a chromogenic reaction. The concentration of the target substance can be analysed based on the absorbance of the specific wavelength measured. ML plays a crucial role in processing high-dimensional data when multiple signals are detected. This enables the swift and accurate extraction of detection results, thereby enhancing the reliability and sensitivity of the sensor.

Zhu *et al.* utilized benzenedicarboxylic acid-modified graphene quantum dots (TPA@GQDs) in combination with three transition metal ions ( $\text{Fe}^{2+}$ ,  $\text{Cu}^{2+}$ , and  $\text{Zn}^{2+}$ ) as three sensing units to build a nanozyme sensing array.<sup>53</sup> Using 3,3',5,5'-tetramethylbenzidine (TMB) as the chromogenic substrate, each sensing unit generated different colorimetric signals for six thiol analytes in the presence of  $\text{H}_2\text{O}_2$ . Employing the linear discriminant analysis (LDA) model, the absorbance of the samples could be analysed. This classification model utilized the absorbance of TMB at 650 nm as the characteristic peak and thiol class or concentration as the classification label. Through LDA, thiols of the same type or concentration were clustered together, while those of different types or concentrations could be completely separated. This allowed for accurate differentiation between various types of thiol analytes as well as their concentrations (Fig. 6A). Liu *et al.* selected three oligonucleotides specific for tumor exosomal proteins, each



**Fig. 6** Analysing nanozyme optical sensors through ML. (A) Schematics of colorimetric sensor array based on metal ion integrated TPA@GQD nanozyme for thiol discrimination. Reprinted with permission from ref. 54. Copyright 2022 Elsevier. (B) Nanozyme sensor array plus solvent-mediated signal amplification strategy for ultrasensitive detection of exosomal proteins and cancer identification. Reprinted with permission from ref. 55. Copyright 2021 American Chemical Society. (C) Schematic diagram of material catalysis, colorimetric sensing principle, and portable smart sensor. Reprinted with permission from ref. 56. Copyright 2023 American Chemical Society.

modified with  $C_3N_4$  nanosheets, to form a sensing array.<sup>54</sup> Aptamer adsorption enhanced the selectivity of POD-like activity of *o*-phenylenediamine (OPD) oxidation by  $C_3N_4$  nanosheets. In the presence of tumor exosomes, binding occurred between tumor exosomes and oligonucleotides, leading to the separation of oligonucleotides from the  $C_3N_4$  nanosheets and a reduction in the catalytic activity of the nanozymes. This reduction in catalytic products resulted in a weakened fluorescence signal. Calculating the fluorescence intensity ratios as feature values and using exosomal proteins of the five tumors as classification labels, the LDA classification model categorized the five tumor exosomes into five clusters based on the two most significant differentiation factors. This facilitated the detection of tumor exosomes and the identification of different types of cancers (Fig. 6B).

All the aforementioned efforts leverage sensor arrays to enhance sensor reliability. A sensor array has the capability to generate diverse cross-reaction signals (*i.e.*, fingerprint signals) for each analyte, facilitating multiplexed detection and identification of multiple detectors. The nanozyme sensor array incorporates multiple sensing units, each comprising different nanozymes undergoing similar catalytic reactions and producing multiple signals. Therefore, the application of ML algorithms allows for rapid processing of these high-dimensional data through techniques such as feature selection and dimensionality reduction. Data analysis capabilities of ML enable quick prediction and classification, outputting the concentration or class of the target in a shorter timeframe.

**3.1.2 Image colour-based analysis of substrate concentration.** ML not only facilitates the extraction of relevant information from colour development reactions by measuring absorbance, but also enables the direct input of images. By segmenting the images and extracting optical information such as RGB or HSV, ML can infer the concentration of the target detector.

Sun *et al.* utilized the POD-like activity of  $CuO/Fe_2O_3$  for detecting glutathione pesticides and chlortetracycline hydrochloride (CTC).<sup>55</sup> Upon the addition of glufosinate, it adsorbed to the surface of  $CuO/Fe_2O_3$ , inhibiting the active center. The subsequent addition of CTC restored the nanozyme activity due to the interaction between glufosinate and CTC. Changes in nanozyme activity resulted in varying colour development reactions of the TMB colour developer. They established a multifunctional intelligent nanozyme sensor platform capable of automatically recognizing input images and building statistical models through deep learning (Fig. 6C). The fundamental principle involved segmenting the image, extracting mean RGB or HSV values, and fitting the relationship between RGB or HSV and the concentration of the target molecule using a linear support vector machine (SVM), enabling intelligent online detection of glufosinate ammonium and CTC concentration.

Dang *et al.* designed a  $Ni/CoMoO_4$  nanozyme sensor with bienzyme-like activity for detecting glufosinate ammonium and CTC concentrations.<sup>56</sup> The nanozyme exhibited bienzyme-like activity sensitive to organophosphorus (OP) and zirco-

nium. In terms of OXD-like activity, the sulfhydryl molecules produced by acetylcholinesterase (AChE) are easily coordinated with metal atoms, blocking the catalytic site of the nanozyme. OP enhances the catalytic activity by deactivating AChE. For POD-like activity, zirconium complexes with Co, blocking the active site and reducing the POD-like activity. Using deep learning, the nanozyme sensor detected concentrations of AChE, OP, and zirconium. Deep learning models recognized input images, and colour pattern analysis applications analysed the colours of “useful” output photos. Employing the Yolo V3 algorithm, images were segmented, and average RGB or HSV values were extracted. SVM was then used to fit the relationship between RGB or HSV and the concentration of the target molecule, allowing separate detection of multiple target detector concentrations based on the multienzyme-like activities.

Compared to analysing absorbance data, the analysis of input images demands more sophisticated analytical capabilities from ML, making it more suitable for online intelligent detection. This enables real-time detection and intelligent analysis of the data.

### 3.2 ML in nanozyme-based electrochemical sensing

Electrochemical sensors utilize electrochemical principles to detect the presence and concentration of chemicals. These sensors work based on redox reactions on the electrode surfaces and typically consist of a working electrode, a counter electrode, and a reference electrode. Nanozymes can enhance these sensors by acting as catalysts to oxidize specific substrates on the surface of the working electrode, leveraging their enzyme-like activity. Incorporation of nanozymes improves the sensitivity and specificity of the sensors. By adding nanozymes, electrochemical sensors can better detect low concentrations of target analytes, finding applications in agriculture, biomedicine, environmental monitoring, and other fields. However, as detection limits decrease, errors of the same magnitude become more noticeable and influential. To mitigate relative errors, it becomes crucial to employ ML algorithms for the analysis of electrical signals generated during the catalytic process of nanozymes. ML algorithms can learn any continuous function and are well-suited for constructing nonlinear models.

Zhu *et al.* integrated a nanozyme-based electrochemical sensor with ML for measuring electrical signals produced during the OXD-like catalytic reaction of carbazene (CBZ) residues by  $MoS_2$  nano-hybrid nanozymes.<sup>57</sup> Using the artificial neural network (ANN) algorithm, they constructed a neural network model correlating current and various concentrations of CBZ. The electrical signals served as outputs, facilitating the detection of CBZ residue concentrations. This hybrid approach, combining the nanozyme activity with ML analytics, resulted in a sensor with a lower detection limit and heightened sensitivity. The incorporation of ML allowed the electrochemical sensor to handle multiple signal inputs and outputs while maintaining low error, enabling simultaneous detection of concentrations for multiple targets with the nanozyme-

based electrochemical sensor. Likewise, Zhu *et al.* analysed and processed electrical signals generated during the OXD-like catalytic reaction of two substrates, xanthine (XT) and hypoxanthine (HX), by 3D porous graphene nanozymes.<sup>58</sup> Using an ANN algorithm, they determined concentrations for both substrates, demonstrating the capability of the sensor to detect multiple target concentrations. Additionally, Zhu *et al.* employed the OXD-like activity of single-walled carbon nanohorns (SWCNHs) to catalyse 5-hydroxytryptamine (5-HT), producing an electrical signal.<sup>59</sup> Leveraging the derivative technique for signal preprocessing, they improved the signal resolution and sensitivity. The ANN algorithm then effectively modeled the current and various concentrations of 5-HT, enhancing the accuracy of substrate concentration predictions.

Integrating ML with nanozyme-based sensors opens up new possibilities. Given that the intensity of the electrical signal correlates positively with nanozyme activity, ML can enhance the preparation methods for nanozyme-based sensors by analysing electrical signals to identify conditions yielding the highest activity. Xu *et al.* employed an orthogonal experimental design in combination with the BP artificial neural network-genetic algorithm (BP-ANN-GA) to assess the impact of four factors (volume ratio of graphene oxide (GO) and multi-walled carbon nanotubes (MWCNTs), silver nitrate concentration, CV deposition cycle, and pH of phosphate buffer) on the peak current value (I) of benzyl (BN).<sup>60</sup> This approach aimed to optimize the preparation technique for the nanozyme-based sensor and determine the optimal experimental conditions. Under these conditions, the nanozymes catalysed a BN reaction on the working electrode, generating an electrical signal. Subsequently, the support vector machine (SVM) algorithm and least squares support vector machine (LS-SVM) algorithm were employed to achieve intelligent sensing of BN.

Through ML-assisted electroanalysis of chemical reaction signals, nanozyme-based sensors have demonstrated enhanced sensitivity and improved detection capabilities. Furthermore, ML plays a crucial role in refining the preparation and optimizing the performance of nanozyme-based sensors.

### 3.3 ML in multimodal sensing detection of nanozymes

Multimodal sensors for nanozymes involve the use of multiple types of sensors to gather diverse data during the catalytic process of nanozymes, leveraging the various properties of nanozymes to provide more comprehensive information. While arrays of nanozyme-based sensors with multiple signals have been developed for chemical assays, they typically consist of a combination of the same type of sensing units, and data processing is relatively straightforward. In the context of processing multimodal sensing signal data from nanozymes, the challenge lies in establishing multiple regression equations to deduce the concentration of the actual sample. This complexity calls for the computational power and accuracy provided by ML algorithms to effectively address the challenges associated with nanozyme-based multimodal detection.

Yu *et al.* developed a multimodal nanozyme-based sensor where liposomes containing hollow Prussian blue nano-

particles (h-PB) were confined in test cells.<sup>61</sup> The released hollow ferrocyanine blue nanoparticles were then transferred to a TMB-H<sub>2</sub>O<sub>2</sub> system to generate a colorimetric signal through the classical sandwich immunochromatographic analysis. Simultaneously, the temperature signal was detected *via* the photothermal effect under 808 nm near-infrared laser excitation. Both signals were subjected to analysis using an ANN model, resulting in multimodal biosensing for precise targeted protein detection (Fig. 7). This multimodal nanozyme-based sensor demonstrated an ultra-wide dynamic range of 0.02 to 20 ng mL<sup>-1</sup> and a detection limit of 10.8 pg mL<sup>-1</sup>, showcasing improved sensitivity. This study serves as a noteworthy example of a multimodal nanozyme-based sensor, highlighting the considerable potential of multimodal sensors in biosensing applications.

### 3.4 Summary

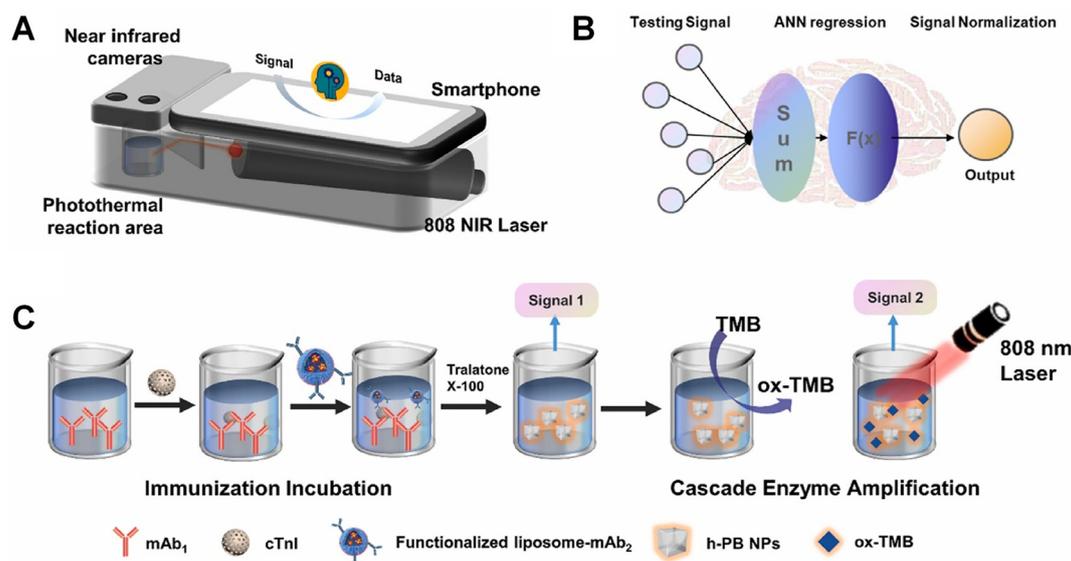
The ML-based analysis method significantly enhances the analytical capabilities of nanozyme-based sensors, enabling them to achieve real-time detection goals with improved accuracy and reduced errors. As ML algorithms continue to develop and improve, it is anticipated that nanozyme-based sensors will demonstrate even greater potential and value in the fields of medicine and environmental monitoring in the future.

## 4. ML roadmap for nanozymes

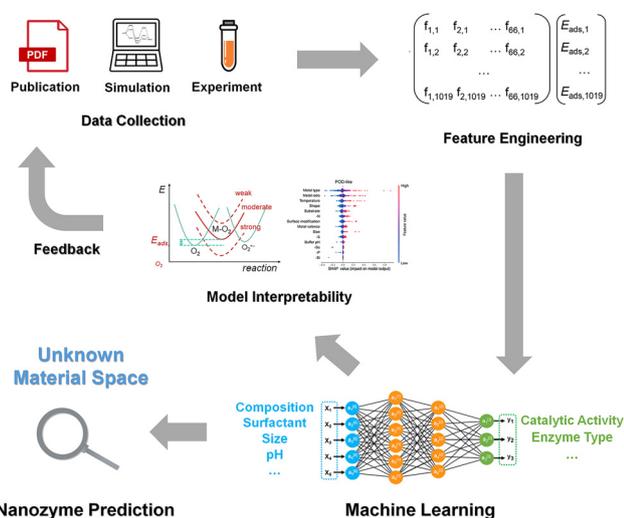
Despite the availability of various open-source ML frameworks like scikit-learn, TensorFlow, PyTorch, *etc.*, selecting and implementing the ML process can be challenging for non-experts. This review presents a generalized workflow for ML in nanozyme research, consisting of four main steps (Fig. 8): (1) Construction of the original dataset; (2) Data preprocessing and feature engineering; (3) Selection, training, and validation of the ML model; and (4) Application of the ML model.

### 4.1 Construction of nanozyme datasets

A high-quality dataset for nanozyme research should encompass clean and comprehensive data, capturing intrinsic material properties (*e.g.*, composition, size, shape, surface modification, *etc.*), external characteristics (*e.g.*, temperature, pH, *etc.*), and corresponding enzyme-like activity descriptors. Researchers typically employ three methods to construct their datasets. The first method involves collecting data from published literature. Razlivina *et al.* curated research data on various nanozymes from published papers, utilizing descriptors from libraries like pubchempy and rdkit to enhance the Dizyme database.<sup>41</sup> They significantly expanded the number of nanozymes and features, offering a valuable resource for the development of high-performance nanozymes. The second method involves extracting material structure and related data from existing material databases, followed by screening material features to construct the dataset. Open access databases such as the Crystallographic database and Materials Project (MP) database provide application programming inter-



**Fig. 7** Nanozyme-based multimodal sensors. (A) Schematic diagram of the constructed portable photothermal colorimetric dual-modality biosensor. (B) Schematic diagram of artificial neural network for multimodal data processing. (C) Schematic representation of the immune strategy for target cTnI-triggered bimodal biosensing, where the obtained colorimetric and photothermal data were passed into the ANN model for further processing. (mAb<sub>1</sub>: monoclonal anti-cTnI capture antibody; mAb<sub>2</sub>: monoclonal anti-cTnI detection antibody; h-PB NPs: hollow Prussian blue nanoparticles; ox-TMB: oxidation-TMB). Reprinted with permission from ref. 62. Copyright 2022 Elsevier.



**Fig. 8** The universal flowchart for ML in the field of nanozymes. Reprinted with permission from ref. 46. Copyright 2023 Wiley-VCH GmbH. Reprinted with permission from ref. 41. Copyright 2022 Wiley-VCH GmbH. Reprinted with permission from ref. 6. Copyright 2023 American Chemical Society.

faces (APIs) and web interfaces for researchers to access material information. Zhang *et al.* constructed a dataset for SOD-like nanozymes using data from the MP database, covering numerous transition metal thiophosphates.<sup>44</sup> Gao *et al.* utilized screening criteria from the C2DB to identify 1019 stable 2D materials, performing DFT calculations to establish an adsorption energy dataset.<sup>46</sup> The third approach entails conducting high-throughput calculations. Yu *et al.* created 168

different GDY-based computational models, calculating Gibbs free energy changes for complete POD-like reaction paths. This resulted in a dataset containing GDY doping strategies and their corresponding POD-like activity magnitudes, using  $\Delta G_1$ ,  $\Delta G_2$ , and  $\Delta G_3$  as descriptors for POD-like activity.<sup>45</sup> These methods showcase diverse strategies for dataset construction, incorporating data from the literature, existing databases, and high-throughput calculations to ensure a comprehensive and reliable foundation for ML-based nanozyme research.

For ML applications in nanozyme detection, researchers must design experiments to gather a substantial amount of data aligned with the specific detection task. The experimental design encompasses selecting the nanozyme type, determining the sensor type, and optimizing detection conditions (*e.g.*, reaction time, temperature, pH, *etc.*). Under these optimal conditions, the collected detection signals serve to construct a dataset, enabling the establishment of a quantitative relationship between detector concentration and the generated signals.

In summary, the quality of raw data significantly influences the performance and reliability of ML models. The approaches to data collection are not limited to the aforementioned methods, and the overarching objective for researchers is to construct a comprehensive database encompassing all nanozymes and their properties.

#### 4.2 Data preprocessing and feature engineering

Data preprocessing, which involves data cleaning, integration, and sampling, along with feature engineering—comprising feature encoding, selection, dimensionality reduction, and normalization—are crucial steps before ML model training.

These processes aim to refine the original dataset, improving the quality of data fed into the ML model. For instance, eliminating redundant features from raw data can enhance the predictive accuracy of ML models. In the context of nanozyme research, data preprocessing and feature engineering are pivotal components due to the intricate and varied nature of the data.

In nanozyme design, researchers often collect data from published publications. However, the quantity and quality of data from different works may vary, leading to datasets with missing values. Effective data preprocessing is crucial, where three methods are commonly used, including deletion, interpolation, and omission. For instance, Razlivina *et al.* encountered a dataset with missing values from over 100 papers on nanozymes.<sup>41</sup> They utilized the K-Nearest Neighbor (K-NN) algorithm to impute missing values, setting a maximum data sparsity threshold of 80%. In cases where features have a high number of missing values, they can be directly deleted. Additionally, Wei *et al.* excluded data with a significant number of missing values, particularly at the  $\zeta$ -potential and the catalytic interface of nanoparticles, as these were not chosen as inputs for the ML model.<sup>62</sup>

In the analytical sensing of nanozymes, experimental data may be affected by noise, outliers, or corrupted data points due to various factors, such as environmental conditions, operational errors, or equipment failures. Data cleaning is a crucial step that involves identifying and removing these inconsistencies. Thresholds can be established to filter out readings that fall outside reasonable ranges, or statistical methods can be applied to identify and correct problematic points. The goal of data cleansing is to enhance the quality of the dataset, ensuring the accuracy and reliability of subsequent analyses. This process enables researchers to analyse data effectively and make accurate predictions.

#### 4.3 Selection, training, and validation of ML Models

Researchers can initially screen the type of model needed based on the study purpose: classification, regression, clustering, or dimensionality reduction models. When it comes to designing nanozymes, the choice between interpretable models for rational design and black-box models for higher accuracy should be considered. The selection of a specific model depends on the dataset size and content. For analytical sensing of nanozymes, different ML models can be used depending on the sensor type and signals. For instance, when analysing absorbance for a small dataset and specific detector species, a classification model is typically chosen. During model training, the  $k$ -fold cross-validation method is commonly used, where the dataset is divided into  $k$  parts for training and testing. For model validation, a portion of the data is reserved as a validation set, and the model's performance is assessed by comparing its output with the original labels. Classification models are evaluated using metrics like accuracy, precision, recall, F1 score, while regression models use metrics such as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and  $R^2$ . Neural

network models are known for their superior prediction accuracy, making them a suitable choice for many studies, including nanozyme detection applications. However, their drawback lies in poor interpretability and the inability to explore conformational relationships in nanozymes. In this context, tree-based ML models, such as RF and XGB, are well-suited for nanozyme research.

To summarize, researchers can initially screen models based on their study purpose (classification or regression). For selecting the specific model, researchers should consider their experience and refer to the dataset's size and characteristics. Using multiple models for learning and subsequently selecting the best-performing model through evaluation is a common practice.

## 5. Challenges and perspectives

An increasing number of researchers are embracing ML methods to expedite the design of high-performance nanozymes or facilitate various applications. Nonetheless, the continuous integration of ML into nanozyme research has brought about certain challenges. In the following section, this review will highlight the prominent issues in utilizing ML for nanozyme research and propose insightful perspectives for addressing these challenges.

### 5.1 Expansion of nanozyme databases

In the design of nanozymes, it is necessary to understand the effects of component properties, material characteristics, and reaction conditions on enzyme-like activity, which directly relate to the catalytic efficiency and affinity of nanozymes. By thoroughly understanding and controlling these variables, it is possible to design and optimize nanozymes, improving their application effectiveness. Therefore, considering cost savings and nanozyme design, it is essential to establish the databases for ML. The activity descriptors in the database mainly include experimental descriptors obtained through experiments, as well as theoretical descriptors obtained through theoretical calculations, both of which need to be enriched to provide more possibilities for ML prediction.

Currently, there are two publicly accessible nanozyme databases at <https://dizyme.net/> and <https://nanozymes.net/>. The former database is divided into three tiers, including a basic version (chemical formula), an progressive version (crystal system, particle size, shape, surfactant), and an advanced version (pH, temperature, hydrogen peroxide concentration, substrate concentration, catalyst concentration, *etc.*). These tiers are designed to predict kinetic parameters such as  $V_{\max}$ ,  $K_m$ , and  $K_{\text{cat}}$  in the Michaelis–Menten equation. Users have the option to contribute to the content of database, which currently contains information on over 300 nanozymes. The latter database is an aggregation of more than 1000 publications, encompassing thousands of materials with details about their kinetic parameters, applications, and references. It is important to note that the former database provides a prediction

function for nanozyme activity, whereas the latter database focuses on collecting and organizing information about materials and the applications of nanozymes.

However, the lack of uniform research and testing conditions and methods has resulted in many nanozyme performance data lacking comparability, hindering the effective expansion of nanozyme databases. Therefore, developing a set of standardized operating procedures (SOPs), including reaction conditions (such as temperature, pH, substrate concentration, *etc.*), dosages, detection methods, *etc.*, is an effective approach to ensure the comparability of data generated in different laboratories. Additionally, establishing nanozyme characterization standards, including catalytic efficiency, stability, toxicity, biocompatibility, *etc.*, is crucial. These evaluation standards help define which performance data are considered important and necessary. Standardizing data and research through the implementation of protocols contributes to the further expansion of databases, enabling the design of more efficient and specific nanozymes.

### 5.2 Improving interpretability of ML models for nanozyme research

The poor interpretability of ML models, including tree model-based ML, remains a significant challenge for nanozyme researchers. Enhancing the interpretability of ML models in the context of nanozyme research is a critical objective. There are two main categories of interpretable ML methods: those with self-interpretable models and those with external co-interpretation methods.<sup>63</sup> For self-interpretable models, achieving interpretability can involve the direct adoption of interpretable ML models. Examples of such models include decision trees and linear regression, which inherently provide insights into the decision-making process. Another approach is to externally co-interpret ML models. This involves using methods and tools designed to provide additional insights into the model's decision rationale. Some examples include SHAP, knowledge graphs, and feature visualization with cluster analysis.

The current understanding of the catalytic mechanism of nanozymes relies heavily on molecule simulations and empirical judgments by researchers. However, the catalytic mechanisms may vary across different material systems. The development of interpretable ML models holds the potential to unveil unified descriptors of nanozyme activity, leading to the discovery of activity laws governing nanozymes. While the Michaelis–Menten equation has been utilized in nanozyme research, the differences between nanozymes and natural enzymes may necessitate the development of more tailored kinetic equations. Gao *et al.* proposed a microkinetic equation for POD-like activity on the material surface, which outperformed the Michaelis–Menten equation.<sup>46</sup> Interpretable ML models could aid in deriving microkinetic equations for various nanozyme activities, offering a more accurate representation of catalytic mechanisms. Interpretable ML models have the potential to provide microkinetic equations specific to different nanozyme activities, potentially replacing or updating the tra-

ditional Michaelis–Menten equation. This shift could result in a more nuanced and accurate representation of the catalytic mechanisms of nanozymes.

### 5.3 ML for complex nanozyme systems

The exploration of nanozymes has introduced increasing complexity into research systems. This complexity arises from both the intricate material systems of nanozymes and their diverse application environments. On the one hand, complex material systems like heterojunctions and high-entropy alloys may feature numerous defects and intricate electronic structures. They may even harbour a multitude of multiple catalytic active sites, giving rise to phenomena such as active site migration during catalytic reactions. On the other hand, nanozymes find applications in varied environments, ranging from extreme pH conditions to the complex biological milieu within the human body. Environmental factors and substance interferences can impact the catalytic reactions of nanozymes. This interference includes substrate adsorption competition and alterations to the reaction pathway, leading to changes in the catalytic activity of nanozymes.

Fortunately, the integration of ML potentials and ML force fields stands out as a promising avenue for overcoming the challenges posed by the intricate material systems and complex application environments of nanozymes. The utilization of ML potentials facilitates the swift calculation of intricate material systems. ML-driven potentials enhance the simulation of complex electronic structures and multi-catalytic active sites, offering a more efficient approach to understanding nanozyme behaviour. The incorporation of ML force fields contributes to maintaining calculation accuracy while significantly expediting large-scale MD studies. ML-driven force fields reduce computational costs, enabling a more comprehensive exploration of the nanozyme reaction mechanisms.

### 5.4 ML helps to expand the enzyme types of nanozymes

To date, the utilization of ML has predominantly centered around the investigation of oxidoreductase-like nanozymes. However, there is a notable absence of reports concerning ML applications in the exploration of other types of nanozymes. Among these, hydrolase-like nanozymes represent a well-studied category distinct from the oxidoreductase-like counterparts. Although there has been a data-driven discovery of novel hydrolase-like nanozymes,<sup>64</sup> the limited dataset available for this class has impeded the application of ML for predicting material activities. In the future, we believe that with datasets expand and research efforts intensify, ML is likely to be applied in the field of hydrolase-like nanozymes to undertake some tasks of predicting enzyme-like activities.

### 5.5 Large language model aided design of nanozymes

Large language models, such as the transformer-based GPT-4, have emerged as powerful tools in materials science, demonstrating capabilities in understanding and generating human language. Boiko *et al.* recently employed an AI system driven by GPT-4 to autonomously reproduce an optimized palladium-

catalysed cross-coupling reaction.<sup>65</sup> This AI program exhibited the ability to independently search for relevant papers, collect data from the internet, and formulate experimental protocols, effectively assisting researchers in their work. The incorporation of AI has the potential to broaden the horizons of nanozymes research. Non-ML experts could leverage AI to design novel nanozymes based on extensive data, obtaining synthesis methods by interacting with the AI.

## 6. Conclusions

In summary, the integration of ML into the design and application of nanozymes is in its early stages, and the full potential of ML, especially its capacity to handle extensive datasets, is yet to be fully harnessed. The research potential of ML in the nanozyme field is considerable, and as theoretical advancements in nanozyme research align with the evolution of ML algorithms, there exists the possibility that ML could eventually supplant traditional research methodologies entirely. The ongoing development in both nanozymes and ML is anticipated to pave the way for transformative advancements, shaping the future of research in this interdisciplinary domain.

## Author contributions

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 32371465).

## References

- 1 Y. Y. Huang, J. S. Ren and X. G. Qu, *Chem. Rev.*, 2019, **119**, 4357–4412.
- 2 J. Lee, H. Liao, Q. Wang, J. Han, J. H. Han, H. E. Shin, M. Ge, W. Park and F. Li, *Exploration*, 2022, **2**, 20210086.
- 3 L. Z. Gao, H. Wei, B. Jiang, D. J. Wang, R. F. Zhang, J. Y. He, X. Q. Meng, Z. R. Wang, H. Z. Fan, T. Wen, D. M. Duan, L. Chen, W. Jiang, Y. Lu and K. L. Fan, *Prog. Chem.*, 2023, **35**, 1–87.
- 4 J. J. X. Wu, X. Y. Wang, Q. Wang, Z. P. Lou, S. R. Li, Y. Y. Zhu, L. Qin and H. Wei, *Chem. Soc. Rev.*, 2019, **48**, 1004–1076.
- 5 L. Gao, J. Zhuang, L. Nie, J. Zhang, Y. Zhang, N. Gu, T. Wang, J. Feng, D. Yang, S. Perrett and X. Yan, *Nat. Nanotechnol.*, 2007, **2**, 577–583.
- 6 Z. Chen, Y. X. Yu, Y. H. Gao and Z. L. Zhu, *ACS Nano*, 2023, **17**, 13062–13080.
- 7 L. Z. Gao, H. Wei, B. Jiang, D. J. Wang, R. F. Zhang, J. Y. He, X. Q. Meng, Z. R. Wang, H. Z. Fan, T. Wen, D. M. Duan, L. Chen, W. Jiang, Y. Lu and K. L. Fan, *Prog. Chem.*, 2023, **35**, 1–87.
- 8 Y. H. Zhang, W. L. Liu, X. Y. Wang, Y. F. Liu and H. Wei, *Small*, 2023, **19**, 2204809.
- 9 T. Kang, Y. G. Kim, D. Kim and T. Hyeon, *Coord. Chem. Rev.*, 2020, **403**, 213092.
- 10 X. Y. Ren, D. X. Chen, Y. Wang, H. F. Li, Y. B. Zhang, H. Y. Chen, X. Li and M. F. Huo, *J. Nanobiotechnol.*, 2022, **20**, 92.
- 11 X. H. Niu, B. X. Liu, P. W. Hu, H. J. Zhu and M. Z. Wang, *Biosensors*, 2022, **12**, 92.
- 12 J. J. Liu and X. H. Niu, *Chemosensors*, 2022, **10**, 386.
- 13 E. Sánchez-Tirado, P. Yáñez-Sedeño and J. M. Pingarrón, *Micromachines*, 2023, **14**, 1746.
- 14 C. A. S. Ballesteros, L. A. Mercante, A. D. Alvarenga, M. H. M. Facure, R. Schneider and D. S. Correa, *Mater. Chem. Front.*, 2021, **5**, 7419–7451.
- 15 X. Li, L. J. Wang, D. Du, L. Ni, J. M. Pan and X. H. Niu, *Trends Anal. Chem.*, 2019, **120**, 115653.
- 16 Y. T. Meng, W. F. Li, X. L. Pan and G. M. G. Gadd, *Environ. Sci. Nano*, 2020, **7**, 1305–1318.
- 17 Q. Liu, A. Zhang, R. Wang, Q. Zhang and D. Cui, *Nano Lett.*, 2021, **13**, 154.
- 18 M. Pietrzak and P. Ivanova, *Sens. Actuators, B*, 2021, **336**, 129736.
- 19 S. F. Ji, B. Jiang, H. G. Hao, Y. J. Chen, J. C. Dong, Y. Mao, Z. D. Zhang, R. Gao, W. X. Chen, R. F. Zhang, Q. Liang, H. J. Li, S. H. Liu, Y. Wang, Q. H. Zhang, L. Gu, D. M. Duan, M. M. Liang, D. S. Wang, X. Y. Yan and Y. D. Li, *Nat. Catal.*, 2021, **4**, 407–417.
- 20 Y. Yuan, L. Chen, L. Kong, L. Qiu, Z. Fu, M. Sun, Y. Liu, M. Cheng, S. Ma, X. Wang, C. Zhao, J. Jiang, X. Zhang, L. Wang and L. Gao, *Nat. Commun.*, 2023, **14**, 5808.
- 21 J. Chen, K. Shi, R. Chen, Z. Zhai, P. Song, L. W. Chow, R. Chandrawati, E. T. Pashuck, F. Jiao and Y. Lin, *Angew. Chem., Int. Ed.*, 2024, **63**, e202317887.
- 22 P. Li, X. J. Gao and X. Gao, *ACS Symp. Ser.*, 2022, **1422**, 67–89.
- 23 T. Chen, Y. Lu, X. Xiong, M. Qiu, Y. Peng and Z. Xu, *Adv. Colloid Interface Sci.*, 2024, **323**, 103072.
- 24 Z. Tian, T. Yao, C. Qu, S. Zhang, X. Li and Y. Qu, *Nano Lett.*, 2019, **19**, 8270–8277.
- 25 F. Li, S. Li, X. C. Guo, Y. H. Dong, C. Yao, Y. P. Liu, Y. G. Song, X. L. Tan, L. Z. Gao and D. Y. Yang, *Angew. Chem., Int. Ed.*, 2020, **59**, 11087–11092.
- 26 W. H. Gao, J. Y. He, L. Chen, X. Q. Meng, Y. N. Ma, L. L. Cheng, K. S. Tu, X. F. Gao, C. Liu, M. Z. Zhang, K. L. Fan, D. W. Pang and X. Y. Yan, *Nat. Commun.*, 2023, **14**, 160.

- 27 H. Fan, J. Zheng, J. Xie, J. Liu, X. Gao, X. Yan, K. Fan and L. Gao, *Adv. Mater.*, 2024, **36**, 2300387.
- 28 S. Zhang, Y. Li, S. Sun, L. Liu, X. Mu, S. Liu, M. Jiao, X. Chen, K. Chen, H. Ma, T. Li, X. Liu, H. Wang, J. Zhang, J. Yang and X.-D. Zhang, *Nat. Commun.*, 2022, **13**, 4744.
- 29 P. Xu, X. Ji, M. Li and W. Lu, *npj Comput. Mater.*, 2023, **9**, 42.
- 30 T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa and K.-i. Shimizu, *ACS Catal.*, 2019, **10**, 2260–2297.
- 31 F. Pellegrino, R. Isopescu, L. Pellutiè, F. Sordello, A. M. Rossi, E. Ortel, G. Martra, V. D. Hodoroaba and V. Maurino, *Sci. Rep.*, 2020, **10**, 160.
- 32 Y. Mao, L. Wang, C. D. Chen, Z. Yang and J. Wang, *Laser Photonics Rev.*, 2023, **17**, e2200357.
- 33 T. Xie and J. C. Grossman, *Phys. Rev. Lett.*, 2018, **120**, 145301.
- 34 R. Ding, Y. W. Chen, P. Chen, R. Wang, J. K. Wang, Y. Q. Ding, W. J. Yin, Y. D. Liu, J. Li and J. G. Liu, *ACS Catal.*, 2021, **11**, 9798–9808.
- 35 Y. C. Li, R. F. Zhang, X. Y. Yan and K. L. Fan, *J. Mater. Chem. B*, 2023, **11**, 6466–6477.
- 36 J. Zhuang, A. C. Midgley, Y. H. Wei, Q. Q. Liu, D. L. Kong and X. L. Huang, *Adv. Mater.*, 2024, **36**, 2210848.
- 37 J. Yang, R. Zhang, H. Zhao, H. Qi, J. Li, J. F. Li, X. Zhou, A. Wang, K. Fan, X. Yan and T. Zhang, *Exploration*, 2022, **2**, 20210267.
- 38 G. Tang, J. He, J. Liu, X. Yan and K. Fan, *Exploration*, 2021, **1**, 75–89.
- 39 C. Li, T. Hang and Y. Jin, *Exploration*, 2023, **3**, 20220151.
- 40 Y. Ge, P. Liu, Q. Chen, M. Qu, L. Xu, H. Liang, X. Zhang, Z. Huang, Y. Wen and L. Wang, *Biosens. Bioelectron.*, 2023, **237**, 115454.
- 41 J. Razlivina, N. Serov, O. Shapovalova and V. Vinogradov, *Small*, 2022, **18**, 2105673.
- 42 Y. Wei, J. Wu, Y. Wu, H. Liu, F. Meng, Q. Liu, A. C. Midgley, X. Zhang, T. Qi, H. Kang, R. Chen, D. Kong, J. Zhuang, X. Yan and X. Huang, *Adv. Mater.*, 2022, **34**, 2201736.
- 43 V. Vinogradov, J. Razlivina and A. Dmitrenko, *et al.* DiZyme: The Ultimate Resource for Nanozyme Multiple Catalytic Activity Prediction, 08 November 2023, PREPRINT (Version 1) available at Research Square [DOI: [10.21203/rs.3.rs-3540876/v1](https://doi.org/10.21203/rs.3.rs-3540876/v1)].
- 44 C. Zhang, Y. Yu, S. Shi, M. Liang, D. Yang, N. Sui, W. W. Yu, L. Wang and Z. Zhu, *Nano Lett.*, 2022, **22**, 8592–8600.
- 45 Y. Yu, Y. Jiang, C. Zhang, Q. Bai, F. Fu, S. Li, L. Wang, W. W. Yu, N. Sui and Z. Zhu, *ACS Mater. Lett.*, 2022, **4**, 2134–2142.
- 46 X. J. J. Gao, J. Yan, J. J. Zheng, S. L. Zhong and X. F. Gao, *Adv. Healthcare Mater.*, 2023, **12**, 2202925.
- 47 D. Xu, W. Yin, J. Zhou, L. Wu, H. Yao, M. Sun, P. Chen, X. Deng and L. Zhao, *Nanoscale*, 2023, **15**, 6686–6695.
- 48 Y. Jiang, Z. Chen, N. Sui and Z. Zhu, *J. Am. Chem. Soc.*, 2024, DOI: [10.1021/jacs.3c13588](https://doi.org/10.1021/jacs.3c13588).
- 49 J.-J. Zheng, X. Wang, Z. Li, X. Shen, G. Wei, P. Xia, Y.-G. Zhou, H. Wei and X. Gao, *ACS Nano*, 2024, **18**, 1531–1542.
- 50 X. Wang, X. J. J. Gao, L. Qin, C. Wang, L. Song, Y. Zhou, G. Zhu, W. Cao, S. Lin, L. Zhou, K. Wang, H. Zhang, Z. Jin, P. Wang, X. Gao and H. Wei, *Nat. Commun.*, 2019, **10**, 704.
- 51 Q. Wang, C. Y. Li, X. Y. Wang, J. Pu, S. Zhang, L. K. Liang, L. N. Chen, R. H. Liu, W. B. Zuo, H. G. Zhang, Y. H. Tao, X. F. Gao and H. Wei, *Nano Lett.*, 2022, **22**, 10003–10009.
- 52 Y. Wang, G. Jia, X. Cui, X. Zhao, Q. Zhang, L. Gu, L. Zheng, L. H. Li, Q. Wu, D. J. Singh, D. Matsumura, T. Tsuji, Y.-T. Cui, J. Zhao and W. Zheng, *Chem*, 2021, **7**, 436–449.
- 53 X. Y. Zhu, T. Li, X. Hai and S. Bi, *Biosens. Bioelectron.*, 2022, **213**, 114438.
- 54 M.-X. Liu, H. Zhang, X.-W. Zhang, S. Chen, Y.-L. Yu and J.-H. Wang, *Anal. Chem.*, 2021, **93**, 9002–9010.
- 55 M. Sun, L. Zhao, T. Liu, Z. Lu, G. Su, C. Wu, C. Song, R. Deng, M. He, H. Rao and Y. Wang, *ACS Appl. Mater. Interfaces*, 2023, **15**, 54466–54477.
- 56 Y. Dang, G. Wang, G. Su, Z. Lu, Y. Wang, T. Liu, X. Pu, X. Wang, C. Wu, C. Song, Q. Zhao, H. Rao and M. Sun, *ACS Nano*, 2022, **16**, 4536–4550.
- 57 X. Zhu, P. Liu, Y. Ge, R. Wu, T. Xue, Y. Sheng, S. Ai, K. Tang and Y. Wen, *J. Electroanal. Chem.*, 2020, **862**, 113940.
- 58 X. Y. Zhu, L. Lin, R. M. Wu, Y. F. Zhu, Y. Y. Sheng, P. C. Nie, P. Liu, L. L. Xu and Y. P. Wen, *Biosens. Bioelectron.*, 2021, **179**, 113062.
- 59 Y. Zhu, T. Xue, Y. Sheng, J. Xu, X. Zhu, W. Li, X. Lu, L. Rao and Y. Wen, *Microchem. J.*, 2021, **170**, 106697.
- 60 L. Xu, Y. Xiong, R. Wu, X. Geng, M. Li, H. Yao, X. Wang, Y. Wen and S. Ai, *J. Electrochem. Soc.*, 2022, **169**, 047506.
- 61 Z. C. Yu, H. X. Gong, M. J. Li and D. P. Tang, *Biosens. Bioelectron.*, 2022, **218**, 114751.
- 62 Y. H. Wei, J. Wu, Y. X. Wu, H. J. Liu, F. Q. Meng, Q. Q. Liu, A. C. Midgley, X. Y. Zhang, T. Y. Qi, H. L. Kang, R. Chen, D. L. Kong, J. Zhuang, X. Y. Yan and X. L. Huang, *Adv. Mater.*, 2022, **34**, 2201736.
- 63 J. X. Mi, A. D. Li and L. F. Zhou, *IEEE Access*, 2020, **8**, 191969–191985.
- 64 S. R. Li, Z. J. Zhou, Z. X. Tie, B. Wang, M. Ye, L. Du, R. Cui, W. Liu, C. H. Wan, Q. Y. Liu, S. Zhao, Q. Wang, Y. H. Zhang, S. Zhang, H. G. Zhang, Y. Du and H. Wei, *Nat. Commun.*, 2022, **13**, 827.
- 65 D. A. Boiko, R. MacKnight, B. Kline and G. Gomes, *Nature*, 2023, **624**, 570–578.