ADVANCED MATERIALS
www.advmat.de

Prediction and Design of Nanozymes using Explainable Machine Learning

*Yonghua Wei, Jin Wu, Yixuan Wu, Hongjiang Liu, Fanqiang Meng, Qiqi Liu, Adam C. Midgley, Xiangyun Zhang, Tianyi Qi, Helong Kang, Rui Chen, Deling Kong, Jie Zhuang, Xiyun Yan, and Xinglu Huang**

An abundant number of nanomaterials have been discovered to possess enzyme-like catalytic activity, termed nanozymes. It is identified that a variety of internal and external factors influence the catalytic activity of nanozymes. However, there is a lack of essential methodologies to uncover the hidden mechanisms between nanozyme features and enzyme-like activity. Here, a data-driven approach is demonstrated that utilizes machine-learning algorithms to understand particle–property relationships, allowing for classification and quantitative predictions of enzyme-like activity exhibited by nanozymes. High consistency between predicted outputs and the observations is confirmed by accuracy (90.6%) and R^2 (up to 0.80). Furthermore, sensitive analysis of the models reveals the central roles of transition metals in determining nanozyme activity. As an example, the models are successfully applied to predict or design desirable nanozymes by uncovering the hidden relationship between different periods of transition metals and their enzyme-like performance. This study offers a promising strategy to develop nanozymes with desirable catalytic activity and demonstrates the potential of machine learning within the field of material science.

1. Introduction

The term, nanozyme, encompasses nanomaterials that possess enzymatic catalytic properties.[1] Different from native enzymes, nanozymes exhibit relatively high stability and tolerance toward variable temperature and pH conditions. In combination with the intrinsic characteristics of nanomaterials, nanozymes have potential widespread applications within the fields of biosensing,[2] antibacterials,[3] environmental pollution,[4] and disease therapy.[5] Since our discovery of ferromagnetic nanoparticles with intrinsic peroxidase-like activity in 2007,[6] there have been thousands of publications that reported on enzyme-mimicking activities of nanomaterials, which involve at least six classes of enzyme-mimicry.[7] According to the literature, different nanomaterials can intrinsically possess the same enzyme-mimicking activities,[8] and certain types of nanomaterials tend to exhibit differential enzyme-like catalytic activities.[9] The heterogeneous results reveal the complexity and diversity of nanozymes in terms of catalytic capacity.[10] Indeed, the particle-property relationship of nanozymes is complicated, with a current lack of fundamental understanding. Furthermore, the synthesis of nanozymes with desired characteristics are generally determined by trial and error, and based on intuition and experience, which are time-consuming, laborious and resource-intensive.

As a branch of artificial intelligence, machine learning aims to develop computational algorithms to infer mathematical

Y. Wei, J. Wu, Y. Wu, H. Liu, Q. Liu, A. C. Midgley, X. Zhang, T. Qi, H. Kang, D. Kong, X. Yan, X. Huang
State Key Laboratory of Medicinal Chemical Biology
Key Laboratory of Bioactive Materials for the Ministry of Education
College of Life Sciences
Frontiers Science Center for Cell Responses
Nankai University
Tianjin 300071, China
E-mail: huangxinglu@nankai.edu.cn

Y. Wei, J. Zhuang
School of Medicine
Nankai University
Tianjin 300071, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adma.202201736.

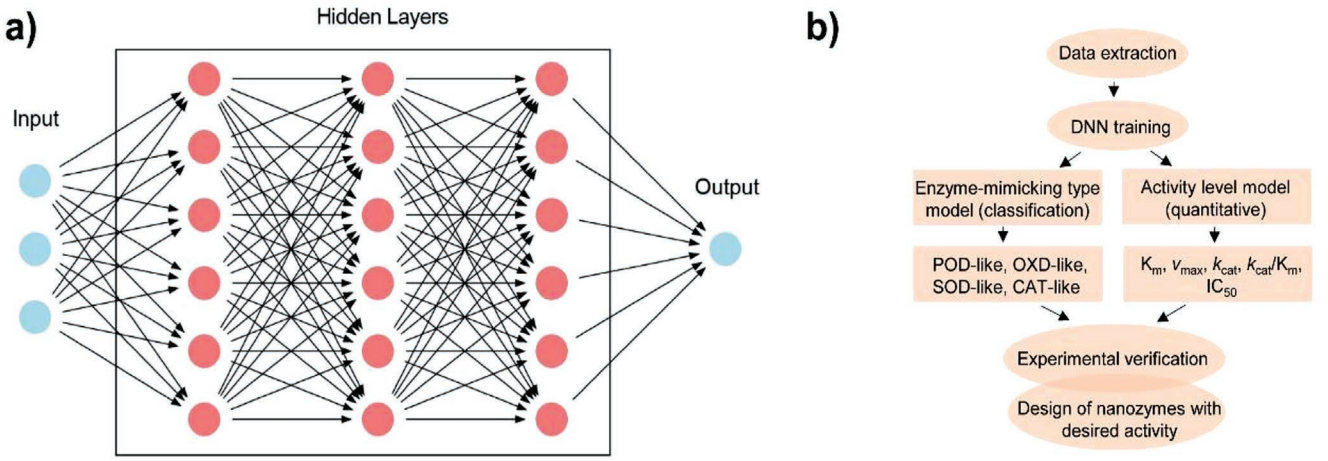
DOI: 10.1002/adma.202201736

F. Meng
College of Information Science and Engineering
China University of Petroleum
Beijing 102249, China

R. Chen
School of Materials Science and Engineering
Nankai University
Tianjin 300350, China

J. Zhuang, X. Yan, X. Huang
Joint Laboratory of Nanozymes
College of Life Sciences
Nankai University
Tianjin 300071, China

X. Yan
CAS Engineering Laboratory for Nanozymes
Institute of Biophysics
Chinese Academy of Sciences
Beijing 100101, China



Scheme 1. Schematic illustration of the construction of machine-learning models. a) Schematic diagram of fully connected DNN-based models. For different models, the neurons and hidden layers were varied to achieve the best fitting effect. b) Scheme of workflow for classification and quantitative prediction of enzyme-like activity of nanomaterials using machine learning.

models based on the existing data, which offers a promising tool for the accelerated development of desired nanoparticles.[11] This approach is particularly useful when applied to the complicated relationships between variables and unknown outcomes, to reveal hidden features within data.[12] For example, fully connected deep neural network (DNN) is an artificial neural network with multiple hidden layers between input and output layers.[13] The "black boxes" of information that comprise the hidden data layers can be uncovered by nonlinear function approximation algorithms. Recently, optimized or new materials were successfully obtained by using data mining by machine-learning models.[14] In particular, machine-learning-assisted materials discovery, using failed experiments, revealed the powerful potential in guiding the development of materials with desirable properties.[15] Given the complexity and diversity of nanozymes, we thus considered if machine learning could uncover the hidden relationship between nanozyme features and enzyme-like performance, to facilitate the prediction and development of nanozymes with preferential functions.

Herein, we developed fully connected DNN-based models for classification and quantitative prediction of enzyme-like activity of nanomaterials (**Scheme 1**a), by training on data extracted from published literature. To establish the models, the extracted data were manually classified into internal and external factors after digitalized processing (Scheme 1b). To predict the nanozymes, the factors were first filled into input modules as independent variables. Subsequently, the classification model was utilized to output the enzyme-mimicking types (i.e., dependent variables), including peroxidase (POD), oxidase (OXD), superoxide dismutase (SOD), and catalase (CAT) activities. For the quantitative models, the level of each enzyme-like activity was set as dependent variables by normalizing the coefficients of enzymatic activity. The regularizations of the models and the validation dataset were also introduced to overcome the potential overfitting of the models. Furthermore, to verify the prediction accuracy of two categories of the models, experimental validation was performed using synthesized nanomaterials. Importantly, to uncover the hidden mechanism between nanozyme features and enzyme-like performance, the interpretability of two categories of the models was explored by sensitive analysis.

2. Results

2.1. Construction of Nanozyme Database for Machine Learning

Experimental data for machine-learning-assisted prediction were typically collected from reported results. As shown in **Figure 1**a, at least 5355 published papers were reported by the data retrieval (data from ISI Web of Knowledge, accessed 2021 May 30). To establish a nanozyme database, the data regarding intrinsic enzyme-like activity of nanomaterials were extracted from over 300 papers, as listed in the Excel file in the Supporting Information. According to these reports, we found that the influencing factors on nanozymes were highly heterogeneous, including enzyme-mimicking types and the level of enzymatic activity. As the central point of the reports, the Michaelis–Menten equation was typically utilized to determine enzymatic activity of nanozymes. Thus, 920 pieces of valuable data, as listed in the Excel file in the Supporting Information, were further refined for data mining by screening the data containing at least one variable of the Michaels–Menten equation. Among the data, eight representative internal factors (metal type, metal valence, metal-containing number, size, shape, nonmetal element, surface modification, enzyme-like type) and three representative external factors (buffer pH, temperature, substrate) were classified into independent/dependent variables, and the subfactors accounted for the percentage of each factor are also listed in Figure 1b. These factors were involved in the composition of nanozyme and the reaction condition of catalysis, which are necessary for a catalytic reaction of nanozyme. Additionally, some important characteristic parameters such as zeta potential and catalytic interface of nanoparticles, may also have contributed to the model output. However, these factors were not included within the independent variables due to a large amount of the missing values in the established database.

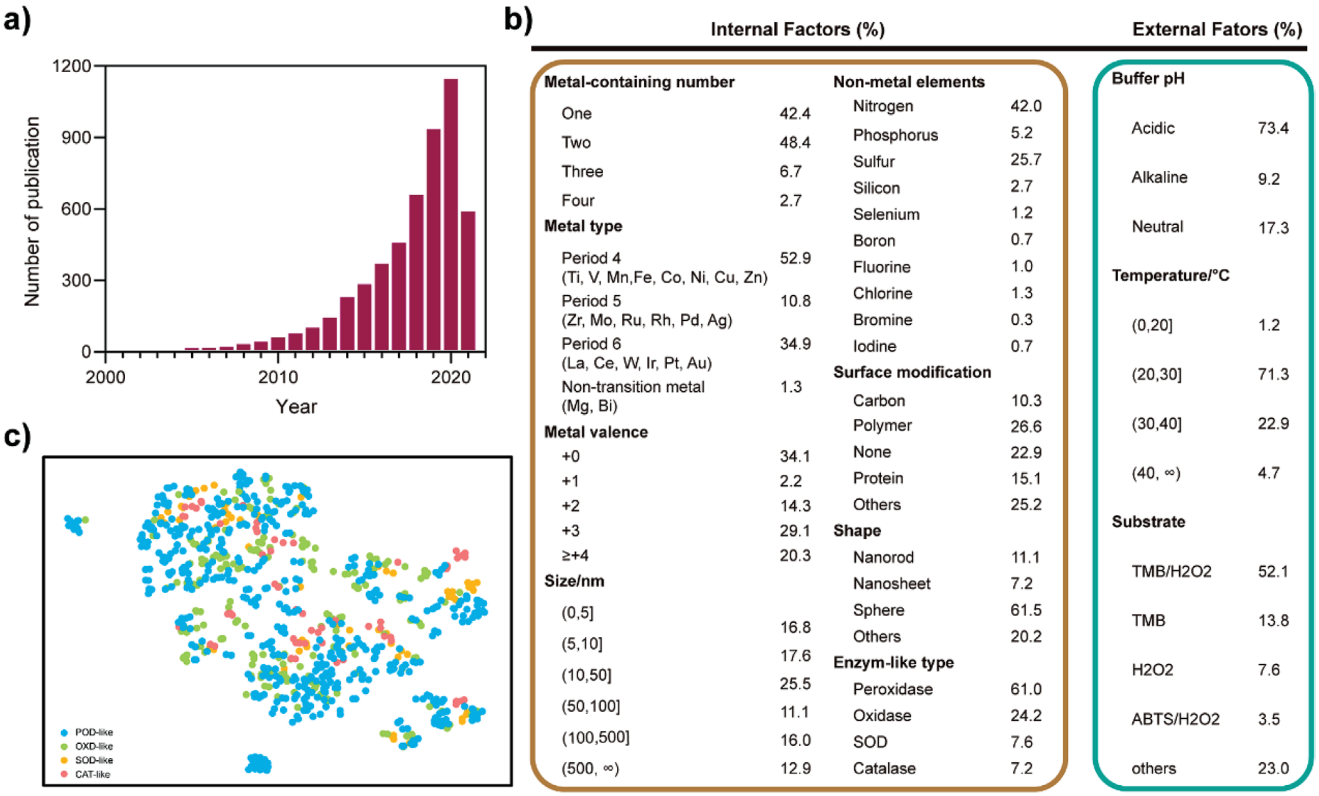


Figure 1. Data collection and dataset establishment. a) The retrieval publication number from ISI web of knowledge by May 30, 2021. Keywords: "nanozyme*" or ("nano*" and ("*ase-like" or "enzyme-like" or "enzyme-mim*" or "*ase-mim*")). b) Distribution of the external and internal factors based on 920 pieces of data. The numbers represent the listed items accounting for the percentage of each factor. TMB: 3,3′,5,5′-Tetramethylbenzidine; ABTS: 2,2′-Azinobis-(3-ethylbenzthiazoline-6-sulphonate. c) The plot showing 2-D distribution of the data for four enzyme-mimicking types. The data were processed by the dimensionality reduction method of UMAP using the "umap" package in R software. Each plot represents a piece of data.

Following database establishment, we next sought to develop two kinds of machine-learning models with different prediction goals: i) enzyme-mimicking type (classification model); ii) the level of enzyme-like activity (quantitative model). Machine learning could not model the raw factors of the variables, especially for the qualitative factors. To digitize the variables, we manually converted the variables to integer values using LabelEncoder via Python. According to initial assessment, oxidoreductase-like activities possessed by nanozymes were the most common feature. The oxidoreductase-like activity was further defined into four subcategories, including peroxidase (POD)-, oxidase (OXD)-, catalase (CAT)-, and superoxide dismutase (SOD)-like activities. After the classification, the dimensionality reduction method of Uniform Manifold Approximation and Projection (UMAP) was utilized to realize the clustering of the four oxidoreductase-like performances of nanozymes.[16] After dimension-reduced processing, a two-dimensional distribution structure of relevant independent variables was produced, as exhibited in Figure 1c. Poor clustering of four enzyme-like activities was observed, as evidenced by the overlapped topological structure of relevant independent variables via UMAP analysis. The results revealed the necessity of applying machine-learning methods for deep learning and data logic mining. To achieve this, we chose fully connected DNN-based models owing to their ability to handle unstructured data, unlabeled data, and nonlinearity. In a typical DNN, data passes from one layer to another, with the output of the former layer playing the role of the input to the next layer. Hidden layers learn increasingly complex abstract features with more robust discriminative capabilities. Finally, the output layer obtained the results that were derived from the most abstract features of the hidden layers. The training process could be divided into three steps: i) definition of sets of functions; ii) changes of parameters and functions, and iii) results obtained with high accuracy. In this part, the input layer received the nanozyme features extracted from the published reports. These data flowed into the subsequent hidden layers for analysis, and the output layer displayed the predicted results (i.e., enzyme-mimicking types, the level of enzyme-like activity). To better predict the level of enzyme-like activity, the indices representing the level of enzymatic activity were manually digitized. In other words, the values of four frequently used parameters in enzymology (i.e., K_m, v_{max}, k_{cat}, and k_{cat}/K_m) were normalized for evaluation of the level of POD-, OXD- and CAT-like activities, respectively. For quantifying the level of SOD-like activity, the value of IC_{50} was normalized. Thus, a total of 14 DNN-based models were established and provided a two-step route for classification and quantitative prediction of enzyme-mimicking type and the level of enzyme-like activity. Following a multitude of parameter adjustments, the detailed parameters were determined and the finalized model parameters are listed in Table S1 (Supporting Information).

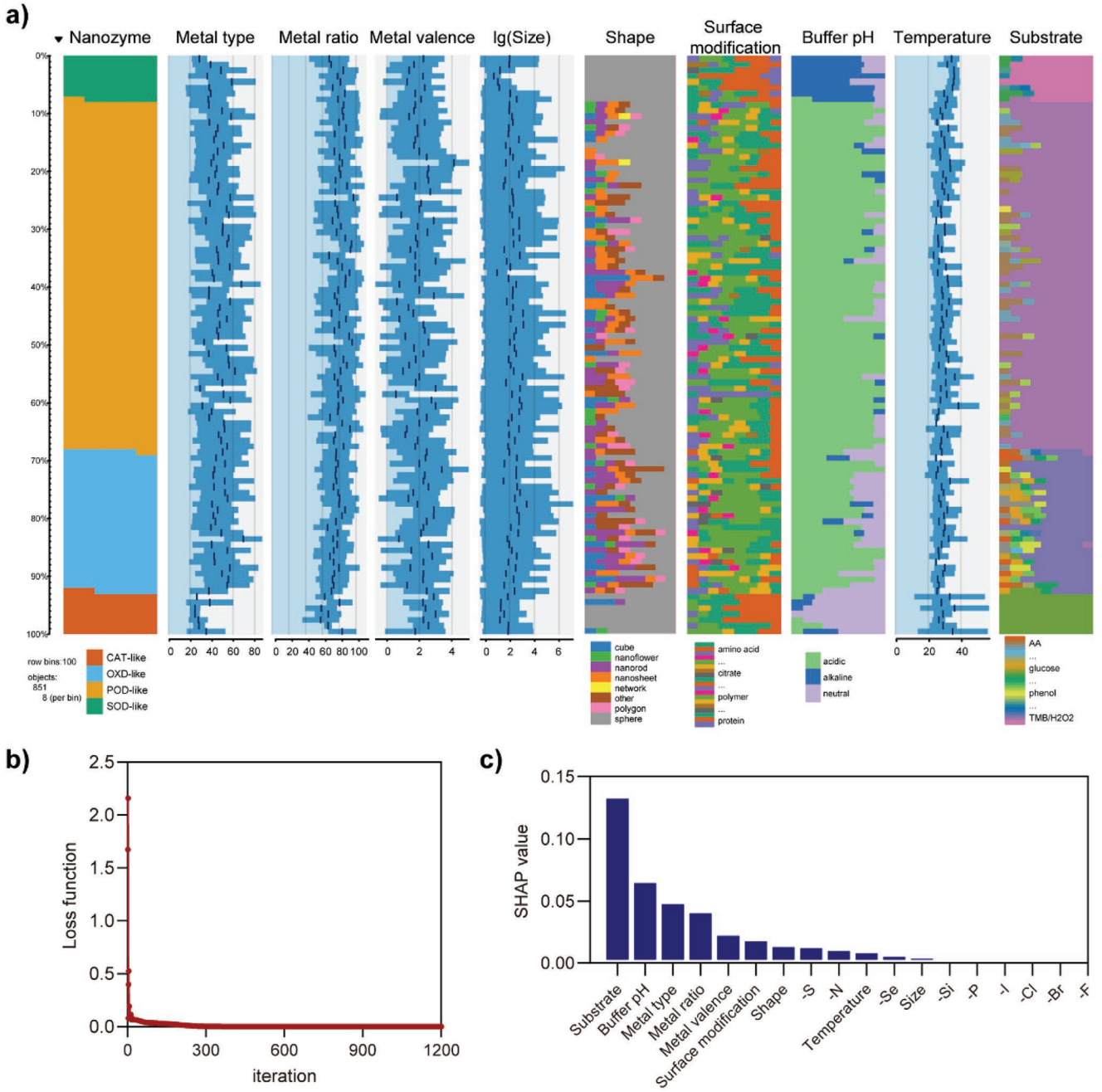


Figure 2. Data analysis of classification model. a) Data distribution of the whole dataset corresponding to four kinds of enzyme types using the "tabplot" package in R software. The total raw data were divided into 100 parts based on data amounts, and the data for each enzyme (i.e., POD, OXD, SOD, or CAT) were observed in the first line, as listed with different colors. Other nine lines displayed the data arrangement of important factors corresponding to the order of each enzyme-mimicking type. The scales of metal type and metal ratio indicate element number and metal-containing percentage of the particles. The size of the particles was logarithmic processing. The data for shape, surface modification, buffer pH and substrate were randomly listed based on the enzyme types, and the categories were indicated with different colors. b) The curve of loss function during iterations in the classification model. c) SHAP analysis of the sensitivity of different factors in the classification model.

2.2. Model of Enzyme-Mimicking Types

The distribution of the extracted data was visualized and sorted by enzyme-like types (**Figure 2**a). Ten important characteristic variables (i.e., enzyme-like type, metal type, metal ratio, metal valence, shape, size, surface modification, buffer pH, temperature, and substrate) were listed according to the reported frequency. It was observed that the data for POD-, OXD-, SOD-, and CAT-like activity accounted for approximately 60%, 20%, 10%, and 10% of all the datasets, respectively. The distribution of various internal/external factors (i.e., independent variables) corresponded to each dataset on enzyme-like

type (i.e., dependent variables). To explore the effect of independent variables on enzyme-like type, a DNN-based model was trained with an optimization process that required a loss function to calculate error. The training process with decreasing loss function was shown following increasing parameter adjustment and iteration (Figure 2b). After approximately 240 iterations, the loss function remained steady, demonstrating the relatively high accuracy of the model. As a data-driven model, overfitting of the training data commonly occurs. To prevent the occurrence of overfitting, we utilized a regularization technique by adding information to the model. Furthermore, the entire database was randomly divided into an 80% training dataset and a 20% validating dataset. Accuracy between the prediction and the observation was applied to determine the performance of the model, as described in the Methods. The results confirmed high accuracy of the model, showing 90.6% accuracy. The multivariate analysis of the confusion matrix was conducted to verify the multiclass classification ability of the model (Figure S1, Supporting Information). The Kappa consistency check (κ) coefficient ($\kappa = 0.803$) revealed the high consistency between predicted outputs and the observations. Together, the prediction results of the model were credible, as evidenced by the close relationship between the independent variables and the enzyme-like performances of nanozymes.

We next sought to understand the effect of independent variables on outputted dependent variables. The sensitivity analysis of the model was conducted by the method of Shapley (SHAP) value.[17] The SHAP values of independent variables are listed in Figure 2c. It was observed that the external factors, especially for substrate and buffer pH, played important roles in dictating enzyme-mimicking types of nanozymes. Given facile experimental verification, the significance of the prediction was limited based on machine learning, and thus, the analysis of the independent variables from the external factors was excluded from this study. For the internal factors, we found that the transition metal-based factors including metal type, metal ratio and metal valence, were ranked the top 3 among all internal factors. The results implied that the metal elements contained within the nanozymes were predicted to be the most influential factor determining nanozyme performance. In comparison to transition metal elements, the effect of the addition of nonmetal elements on the prediction of enzyme-mimicking types was limited, revealing that the model of machine learning provided a guidance with the most significant probability rather than a complete synthesis scheme.

2.3. Prediction of Enzyme-Mimicking Types of Nanozymes

We sought to explore whether the model could be used for the prediction of enzyme-mimicking types of nanozymes. To do this, the prediction accuracy of the classification model for each enzyme-like type was first determined by inputting the validation dataset (i.e., 20% validation data) into the model. The model accuracy toward different enzyme-like performances is shown in **Figure 3**a. The prediction accuracies for POD-, OXD-, SOD- and CAT-like performances of nanozymes were 94.5%, 79.7%, 94.4%, and 70.0%, respectively, showing the feasibility to predict enzyme-mimicking types of nanozymes on the basis of the inputted independent variables. To better understand the prediction ability of the model, we calculated the accuracy value using the observation from the nanozyme database and the predictions outputted by the model. Heatmap images showed the prediction accuracies of the four enzyme-like performances toward various transition metal elements (Figure 3b). The results revealed that the model possessed high accuracy in predicting enzyme-mimicking types when the nanozymes contained transition metal elements. Compared to the other three enzyme-like activities, the model showed better prediction ability for POD-like activity in the covered metal types had the highest degree of accuracy, mainly due to the higher proportion of corresponding data that were included in the dataset.

Next, we checked whether the predicted enzyme-like performances were consistent with the results of experimental validation. We thus randomly prepared four kinds of nanoparticles containing different metal oxides. Four metal oxide nanoparticles including CeO_2, CuO, MnO_2, and Fe_3O_4 were characterized by transmission electron microscopy (TEM) (Figure 3c) and X-ray powder diffraction (XRD) (Figure S2, Supporting Information). TEM images showed the varied sizes and shapes of the four metal-based nanoparticles. The feature data of the nanoparticles were manually input into the classification model. The relationship between the output enzyme-like performances (i.e., predictions) and the laboratory obtained data (i.e., observations) is shown in Figure 3d. All four predictions made by the model were highly consistent with the experimental validation, exhibiting 100% accuracy. Remarkably, the four nanoparticles obtained showed heterogeneity in physicochemical properties, owing to their different synthesis methods, making it difficult to determine which internal factor played essential roles in dictating enzyme-mimicking types. We thus sought to introduce a general strategy to prepare various nanoparticles that contained different metals but were synthesized to keep other aspects similar. We recently utilized recombinant ferritin nanocages (FTn) as nanozyme generators by in situ growth of metal nanoparticles into the cavity of FTn.[18] Such a strategy guarantees the uniform conditions for synthesis, except for the difference of metals. The use of FTn had the capacity to produce various metal-based nanozymes with similar physicochemical properties such as size, shape and surface modification. Five FTn-based nanozymes including Au, Pt, Rh, Ir, and Ru were successfully prepared by taking advantage of FTn as protein scaffolds (Figure 3e). TEM images showed that the properties of five FTn-based nanozymes were similar, with a mean diameter of ≈2 nm and spherical structure. To compare the enzyme-like performances from observed and predicted results, we next input the features (metal-based factors as the major differences) from the five nanozymes into the model. As expected, the observed results were highly consistent with that of the prediction outputted by the classification model (Figure 3f), indicating the robust prediction of enzyme-mimicking types and the rationality of SHAP analysis.

2.4. Models of the Level of Enzyme-Like Activity

In comparison to the classification model of enzyme-mimicking type, the quantitative model of the level of

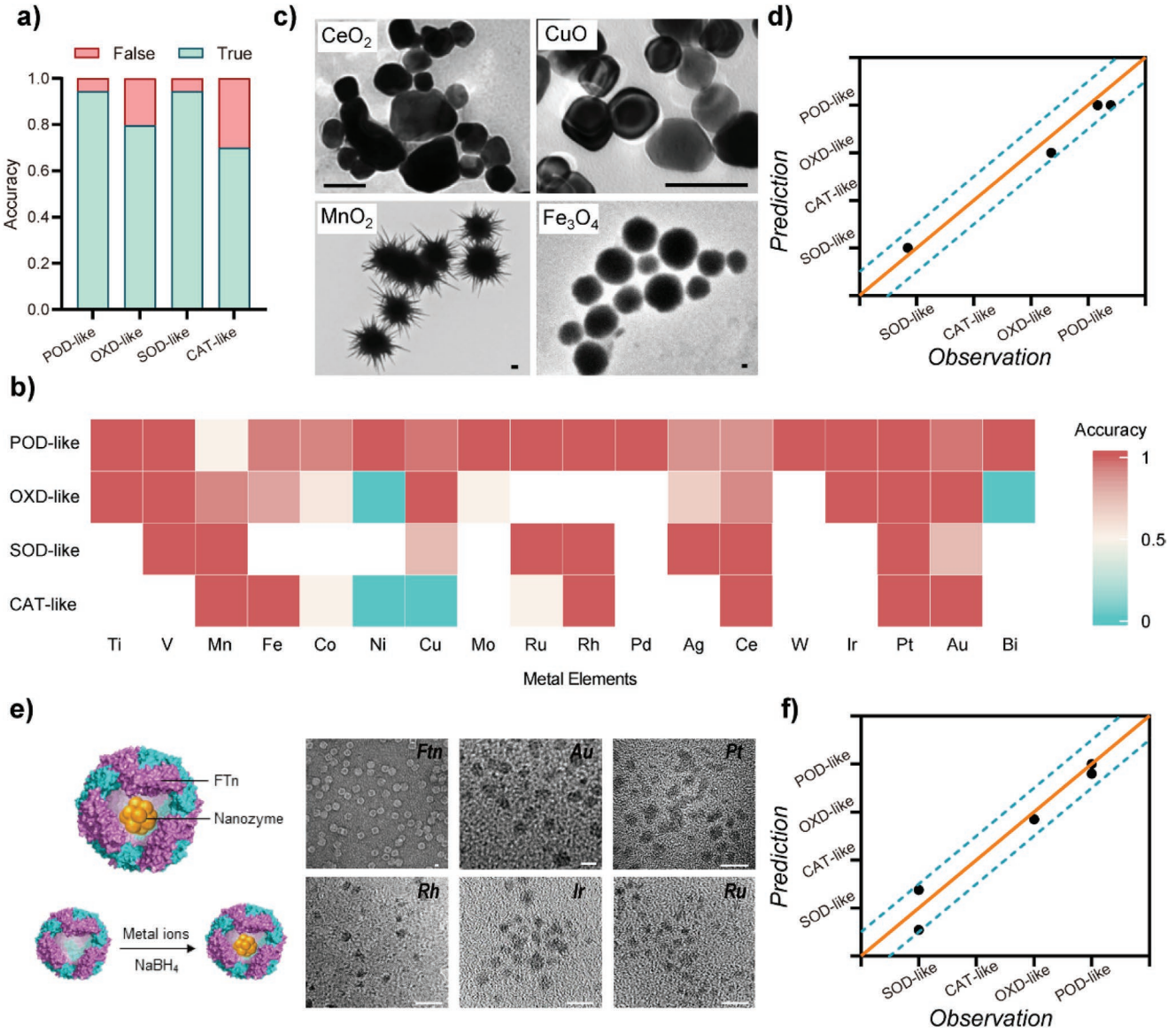


Figure 3. Prediction and experimental verification of enzyme-mimicking types. a) Prediction accuracy of the enzyme-mimicking types by inputting independent variables of the validation dataset into the model. b) Heatmap images of the prediction accuracy of the model by analyzing the enzyme-like activities of various transition metal elements. c) TEM images of different transition metal-based nanoparticles. Scale bar = 100 nm. d) The relationship between the prediction and the observation for enzyme-like activity by obtained nanoparticles from (c). The observed enzyme-like activity was tested by standard method shown in Methods. The slope of the lines is 1 and the dotted green lines represent the intercept of ±0.5. The prediction value existing between the dotted green line is considered as the consistence of enzyme-like activity. e) Schematic illustration (left) of the integration of metal nanoparticles into FTn cavity and TEM images of FTn-based metal nanoparticles (right). Scale bar = 5 nm. f) The prediction and the verification of enzyme-like activity using FTn-based metal nanoparticles. All experiments were performed in triplicate and data are representative of three individual experiments.

enzyme-like activity holds more importance for nanozyme prediction, facilitating the design or discovery of nanozymes with heightened enzyme-like activities. Therefore, we next sought to establish DNN-based quantitative models to predict the level of enzyme-like activity by taking advantage of the internal and external factors as independent variables and the level of enzyme-like activity as dependent variables. Through initial evaluation of the level of the four oxidoreductase-like activities, we found a high degree of heterogeneity in the quantitative value of the different mimicked enzymes. As such, the dataset was divided into four parts on the basis of enzyme types (i.e., POD, OXD, SOD and CAT). The levels of POD-, OXD-, and CAT-like activity were typically evaluated by four coefficients in the Michaels–Menten equation, K_m, v_{max}, k_{cat}, and k_{cat}/K_m. For SOD-like nanozyme, activity was evaluated according to the value of IC_{50}, which corresponds to the median inhibition concentration of SOD-like activity. Then, data on K_m, v_{max}, k_{cat}, and k_{cat}/K_m values of nanozymes were processed by logarithm transformation, followed by a min–max normalization. Each set of data regarding the level of POD-/OXD-/SOD-/CAT-like

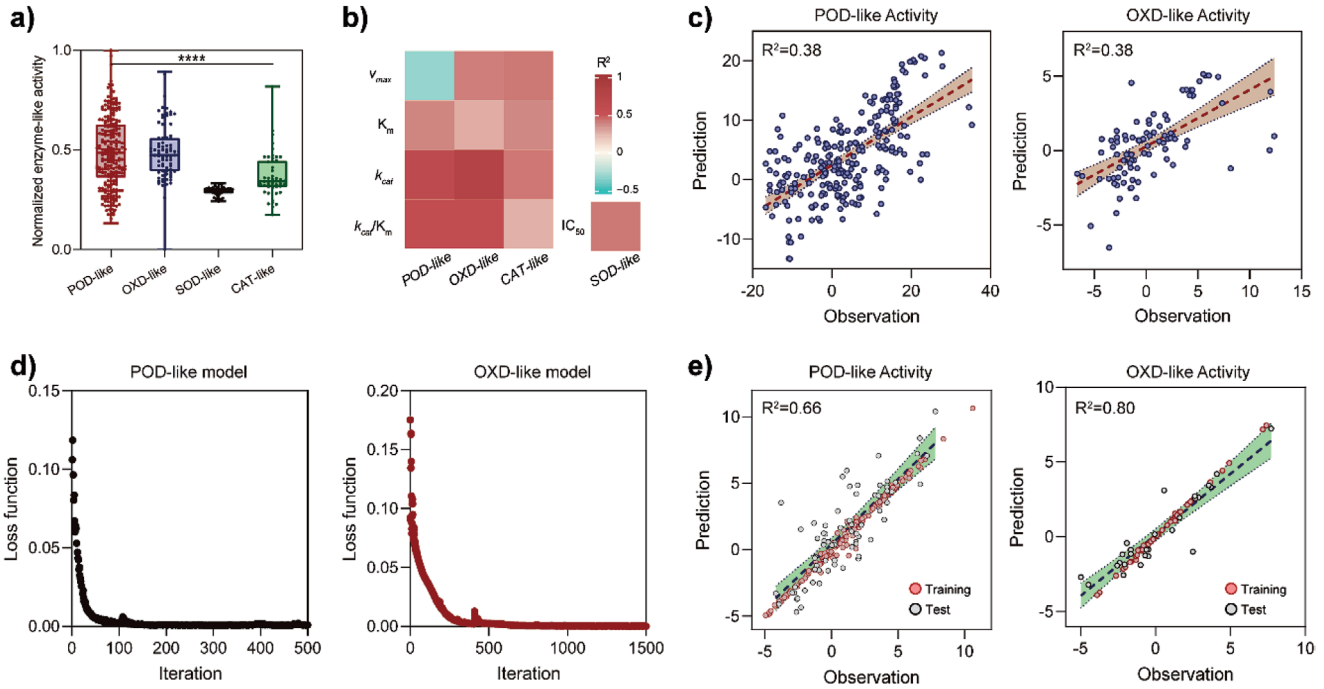


Figure 4. Construction and analysis of the quantification models. a) Distribution of different enzyme activity coefficients for by min–max normalization and Kruskal–Wallis test of k_{cat}/IC_{50}. Kruskal–Wallis test, $^{****}p < 0.0001$. b) Heatmap images of prediction accuracies for 13 DNN-based quantitative model based on R^2 values. Each box represents one model built from a certain dataset, estimated by the correlation coefficient R^2. c) Prediction performances of POD-, OXD-like activities by analyzing the extracted data via multiple linear regression models. Sample size = 243 and 94, respectively. The shading represents 95% confidence intervals. R^2 values were obtained by determining the linear relationship between prediction and observation. d) Loss function of machine-learning models toward iteration for quantitative prediction of the lg k_{cat} of POD-, OXD-like activity, respectively. e) Prediction performances of the POD-, OXD-like activities using the quantitative models of machine learning (sample size = 170 and 66 for Training set, 73 and 28 for Test set, respectively). Red circles: the data after model training; gray circles: the manually inputted values for testing prediction ability. Shaded area: the 95% confidence intervals.

activity was visualized and presented in **Figure 4**a and Figure S3 (Supporting Information). The data distributions of the coefficients for the four enzymes were verified by Kruskal–Wallis rank sum tests in R software. The results illustrated that there were obvious distribution discrepancies for different enzymatic activities, revealing that the separation of the models of different enzymes was necessary for improving the model prediction accuracy. To determine an optimum coefficient for dependent variables of different enzymes, the four coefficients (K_m, v_{max}, k_{cat}, and k_{cat}/K_m) for POD, OXD and CAT and single coefficient (IC_{50}) for SOD were trained to build a total of 13 DNN-based models. The prediction accuracy of the 13 DNN-based models was assessed in the 20% validation dataset according to R^2 values. We found that the accuracy of k_{cat} and IC_{50} models was valuable for the prediction of POD/OXD/CAT and SOD, respectively (Figure 4b). As such, we constructed DNN-based quantitative models based on $lg(k_{cat}/IC_{50})$ of the four enzyme-like activities. According to $lg(k_{cat}/IC_{50})$ values of enzymatic activity, the distribution features of important independent variables for the four enzyme-like activities are correspondingly shown in Figure S4 (Supporting Information). We observed that the data regarding SOD- and CAT-like activities were limited compared to POD- and OXD-like activities. To avoid overfitting, we reduced the number of the independent variables in these two constructed models (SOD and CAT) by involving in one or two kinds of transition metal elements.

To demonstrate the advantage of DNN-based models, we first tested the prediction effect of the extracted data on the level of enzyme-like activity using traditional multiple linear regression (MLR) models. The R^2 values, as determined by the prediction values and the observed values of POD, OXD, SOD, and CAT, were 0.38, 0.38, 0.14, and 0.32, respectively (Figure 4c and Figure S5a, Supporting Information). The detailed parameters of the MLR are listed in Table S2 (Supporting Information). The results revealed poor relationships between independent variables and dependent variables as predicted by the analysis of the MLR model, making it difficult to predict the level of enzyme-like activity. Next, we sought to construct and analyze quantitative models to determine the level of enzyme-like activity. The loss functions of four DNN-based machine-learning models were shown by the corresponding learning processes (Figure 4d and Figure S5b, Supporting Information). It was observed that the loss function rapidly declined with the increase of iteration, demonstrating that the accuracy of the models consistently improved. Loss functions of the other nine models are shown in Figure S6 (Supporting Information). The predicted and observed lg k_{cat}/IC_{50} were further analyzed by the established models. As shown in Figure 4e and Figure S5c (Supporting Information), after model training, the linear distribution of the data was determined by the observed and predicted data (training, red circles), revealing the powerful ability of machine learning in data mining. The model performances

for prediction ability (test, black circles) were determined by R^2, showing R^2 = 0.66, 0.80, 0.43, and 0.42 for POD, OXD, SOD, and CAT, respectively. These results revealed that the DNN-based quantitative models of POD and OXD may serve as suitable predictions of the level of enzyme-like activity. In contrast, the models for SOD and CAT failed achieve similar performances, due to their low R^2 values, mainly a result of the limited data that were extracted from published papers.

2.5. Prediction of the Level of Enzyme-Like Activity

To obtain intrinsic information of the quantitative models, we conducted sensitivity analysis of different factors by data mining. The effect of different factors on the level of POD- or OXD-like activity was determined by SHAP values. **Figure 5**a,b indicate potential increased (red points)/decreased (blue points) effects of the data on the activities. The scattered distribution for the metal-based factors was observed and exhibited that most datapoints for other factors were gathered near SHAP value 0. The results revealed that metal-based factors were sensitive for the model output, which was consistent with the classification model. In other words, the level of enzyme-like activity was mainly determined by the metal composition of nanomaterials. To verify the SHAP analysis, we synthesized several kinds of FTn-based nanozymes that contained certain metals or hybridizations of two metals. The metal concentration of each FTn was quantified by measuring protein and metal content, as shown in Table S3 (Supporting Information). The prediction and observation of the level of enzyme-like activity for the FTn-based nanozymes were performed by inputting the parameters into the models and measuring the enzymatic activity, respectively. Both the observations and predictions for POD-like and OXD-like activities are shown in Figure 5c,d. The results demonstrated that the observed activities were mostly consistent with the predicted activities, implying the rationality of the SHAP analysis and the robustness of the models.

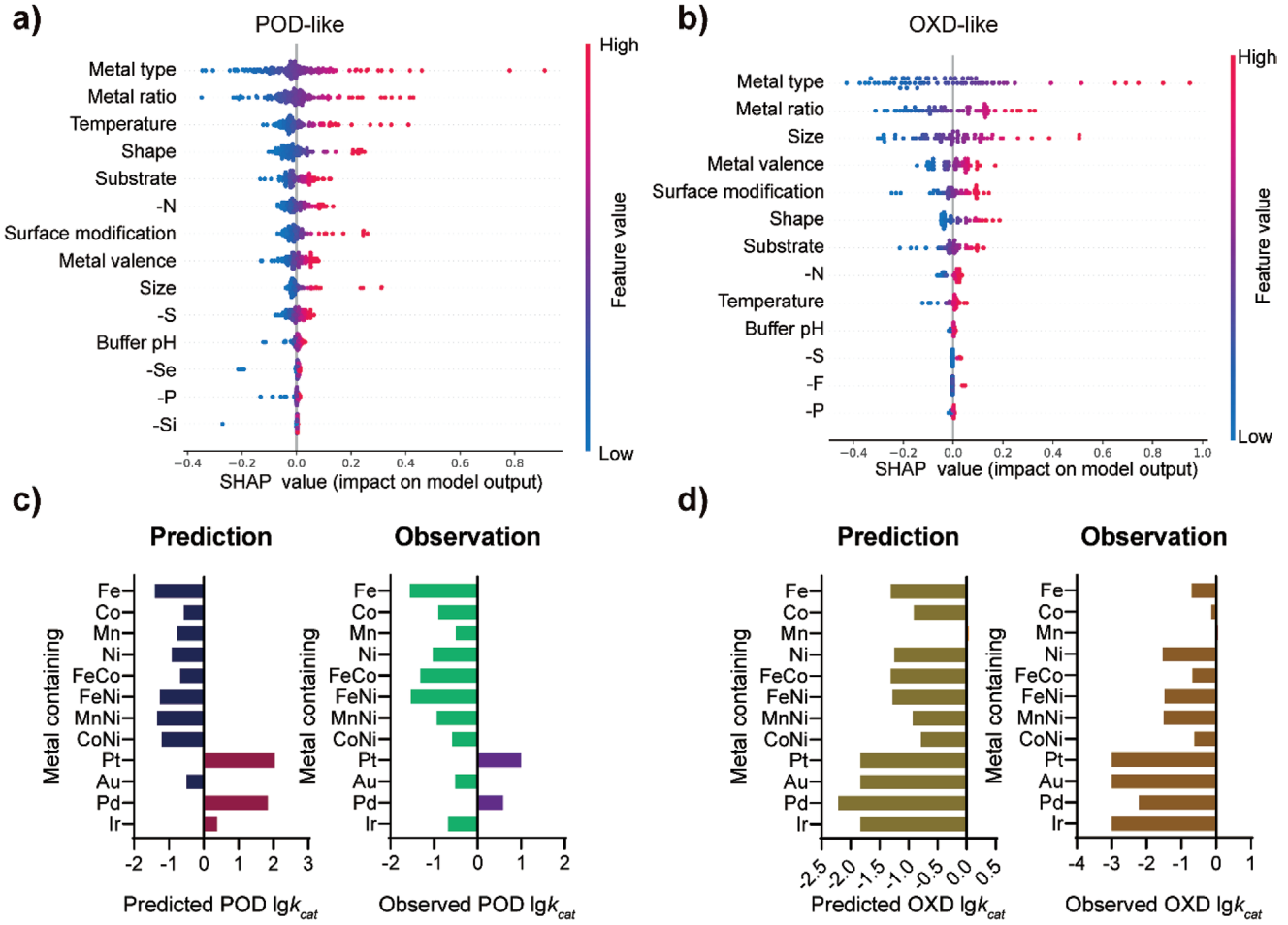


Figure 5. Importance of metal-based factors for the quantification models. a,b) SHAP sensitive analysis of the independent variables in the POD-like quantitative model (a) and the OXD-like quantitative model (b) (n = 170 and n = 66, respectively). The colored dots on the graph represent the SHAP values of the corresponding data. The wider distribution of the SHAP values is more sensitive for influencing enzymatic activity. c,d) Experimental validation of the prediction ability of the POD-like quantitative model (c) and the OXD-like quantitative model (d). Nanomaterials with different metal types and metal doping were synthesized to verify the prediction ability of the models. The observed k_{cat} for POD- and OXD-like activity were obtained by measuring enzymatic activity and determining Michaelis–Menten dynamic curves. Nanomaterials with lg$k_{cat} \leq -3$ of enzyme-like activity were considered to be without enzyme-like activity.

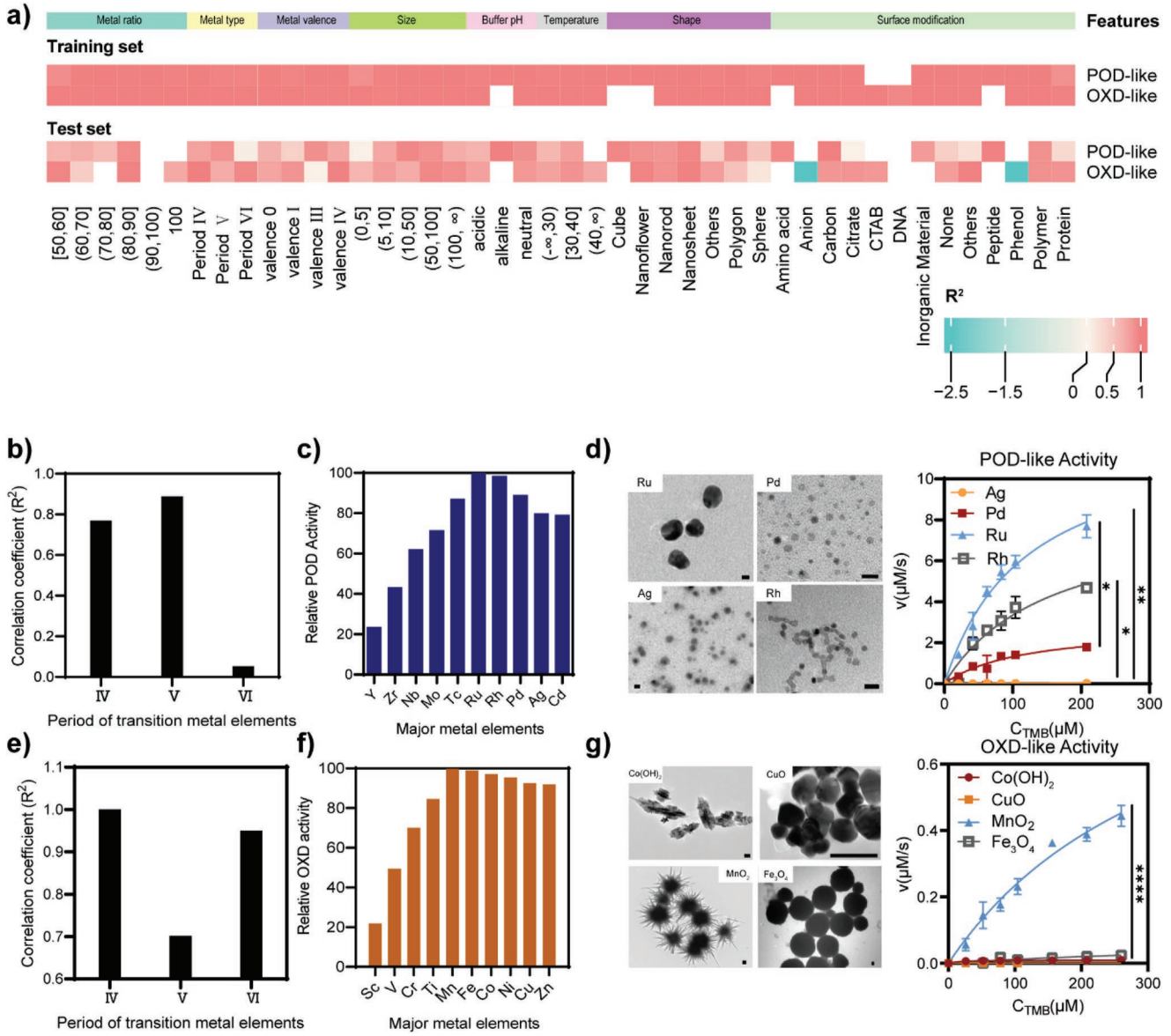


Figure 6. Design of metal-based nanozymes based on the model prediction. a) The correlation coefficients (R^2) for the segmentations of various independent variables, as calculated by the outputs of training dataset and validation dataset. b) R^2 analysis of nanomaterials containing the period IV–VI metals for POD-like activity, as calculated by the prediction value and the observation value of the validation dataset. c) The relative POD-like activity of the period V transition metals after model prediction. d) TEM images of the synthesized nanomaterials containing period V transition metals, such as Ag, Pd, Rh, and Ru (left); experimental validation of POD-like activity of Ag, Pd, Rh, and Ru nanomaterials (right). Scale bar = 10 nm. Data represented as mean ± s.d. ($n = 3$ samples for each data point); $^*p < 0.05$, $^{**}p < 0.01$. e) R^2 analysis based on the prediction and the observation of the validation dataset. The OXD-like activity of nanomaterials containing the period IV–VI metals was analyzed. f) The relative OXD-like activity of the period IV transition metals was evaluated following model prediction. g) TEM images of the synthesized $Co(OH)_2$, CuO, Fe_3O_4, and MnO_2 nanoparticles (left); the dynamic curves of the enzymatic activity of the four metal-based nanoparticle types (right). Scale bar = 100 nm. Data represented as mean ± s.d. ($n = 3$ samples for each data point); $^{****}p < 0.0001$.

We next sought to design nanozymes with desired enzyme-like activities according to the established models. To achieve this, further data mining was performed by manually dividing each independent variable into several segments. The R^2 for each segment was calculated and shown in **Figure 6**a and Figure S7 (Supporting Information). The high R^2 value of each segment involved in the training dataset confirmed the accuracy of the models after machine learning. For the test dataset, different periods of transition metals were studied on the basis of metal type as the most important factor according to SHAP analysis (Figure 5). It was observed that the R^2 values of the period V metals were highest for POD-like activity, compared to period IV and VI metals (Figure 6b). By fixing other independent variables such as pH, temperature, size and shape, the only independent variables input into the models were the ten kinds of period V transition metals. Compared to other

transition metals in this period, Ru elements showed the greatest POD-like activity (Figure 6c). To verify the results, Ru-, Pd-, and Ag-nanoparticles were successfully synthesized and characterized by TEM (Figure 6d, left). The POD-like activity of the three nanoparticles was measured, and the results exhibited typical Michaelis–Menten kinetics (Figure 6d, right). Importantly, the k_{cat} values of POD-like activity for Ru-, Rh-, Pd-, and Ag-nanoparticles were 6.25, 4.29, 1.28, and 0.03 s^{-1}, respectively. Thus, the trend for the activity of different transition metals was fully consistent with the prediction outputs. Likewise, for OXD-like activity, period IV metals were selected for investigation, based on their high R^2 values (Figure 6e). The potential activity of different period IV transition metals was also predicted (Figure 6f). After preparation and characterization of nanoparticles containing Mn, Co, Fe, and Cu (Figure 6g, left), dynamic activity curves were calculated and the enzymatic activities were ranked accordingly: $MnO_2 > Fe_3O_4 > Co(OH)_2 > CuO$ (Figure 6g, right). Together, the quantitative models not only predicted the level of POD- and OXD-like activities but also facilitated the design of desired nanozymes.

3. Conclusion

This study shows that the construction of two categories of DNN-based machine-learning models afforded the qualitative and quantitative prediction of enzyme-like catalytic activity of nanomaterials. The DNN-based models we established in the present study supported that transition metal-based factors played essential roles in enzyme mimicry by nanomaterials, based on the sensitive SHAP analysis, which was also verified by laboratory experiments via the synthesis of different metal-based nanomaterials. A true innovation of machine-learning-assisted prediction would be the synthesis of nanomaterials with unexpected composition-, structure-, size-, and/or shape-property relationships,[14c] which are properties rarely reported on. Additionally, poor explicability is a challenge in machine learning.[19] Sensitive analysis could be used to calculate the marginal contribution of features to the model output and to measure the importance and impact of different features. According to the established ML model, this study presented an example for uncovering the hidden mechanism of nanozyme features and enzyme-like activity through sensitive analysis, which can facilitate the prediction and synthesis of nanomaterials with desirable enzyme-like performances or improved enzyme-like activities. As shown in Figure 6, the data mining was performed to uncover the hidden relationship between different periods of transition metals and their enzyme-like performance. In other words, as "black box" information, Ru-based nanomaterials with high POD-like activity and Mn-based nanomaterials with high OXD-like activity were accurately predicted according to data mining from the established models. As data-driven models, there were several limitations to this study. First, the data were extracted from published papers, which were highly heterogeneous, owing to the diversity of nanozyme assessment methodologies in different laboratories. Second, the currently available data volume resulted in the difficulty of quantitative prediction of SOD-like and CAT-like activities. Third, the datasets extracted from the same paper were typically focused on one certain condition and ignored other factors, which limited the ML model to uncover hidden mechanisms from these studies. Last, in addition to metal types, we sought to explore the effect of other factors on the activity of nanozymes, such as size, shape and metal valence. Unfortunately, the currently extracted data were limited in their capacity to reveal the importance of this valuable information. Thus, we suggest that the establishment of standard procedures of nanozyme assessment methods and the development of more nanomaterials with enzyme-like activities are necessary to realize accurate guidance in the design of nanozymes with desirable functionalities.

4. Experimental Section

Data Extraction and Dataset Construction: The publications were collected by retrieving data from ISI Web of Knowledge using keywords as below: TS = "nanozyme*" or ("nano*" and ("*ase-like" or "enzyme-like" or "enzyme-mim*" or "*ase-mim*")). At least 5355 published papers were found by 2021 May 30. After screening, ≈1000 papers were defined and the data regarding enzyme-like activity of nanomaterials were further extracted. Given the heterogeneity of the retrieval data, the publications we obtained were further filtered by the following conditions: i) the full text was available; ii) the study is focused on the enzyme-like activity of nanozyme or the biomedical applications of nanozyme; iii) some material characterization data and experimental conditions were listed; iv) the enzyme-like types must be given clearly; v) at least one index for determining the level of enzyme-like activity must be included. Finally, 920 pieces of data were collected for establishment of nanozyme database from over 300 papers. The extracted data were divided into internal factors (i.e., metal type, metal valence, metal-containing number, size, shape, nonmetal element, surface modification, enzyme-like type) and external factors (i.e., buffer pH, temperature, and substrate). To build classification model, the enzyme-mimicking types of nanozymes were divided into four types, including POD, OXD, CAT and SOD. For the quantitative models, the level of enzyme-like activity was determined by the coefficients including K_m, v_{max}, $k_{cat,}$ k_{cat}/K_m, and IC_{50}.

Data Processing: To better enable the data read by the computer, some variables in the database were digitally converted. Briefly, the variables regarding nonnumeric data were manually labeled using integer values via a LabelEncoder of Python. The K_m, v_{max}, k_{cat}, k_{cat}/K_m and IC_{50} values were processed by logarithm transformation, followed by a min–max normalization. To better exhibit the relationship between the variables and the enzyme-like activity, the heterogeneous distribution of the variables was performed in the classification and quantitative models by the R package for tabplot visualization. For established 14 models (one for classification model, 13 for quantitative models), the same enzyme-mimicking type shared the identical values of the independent variables. The data for each model were randomly divided into two datasets: the training dataset and the validation dataset. The former was used to build the neural network models, while the latter was applied to verify the accuracy of the established models.

Fully Connected Deep Neuron Network (DNN): The classification and quantitative models were built using fully connected DNN approach. The learning experiments were conducted using AIforScience produced by Beijing Diji Tech, a platform to accelerate implementing and explaining deep-learning algorithms. Typically, the DNN consists of three layers, input layer, hidden layer and output layer. The first layer and the last layer are generally used as the input layer and the output layer, respectively. The other layers are called hidden layer. All the layers are fully connected, that is, any unit in the ith layer is connected to all units in the $(i + 1)$ th layer. In other words, the input layer receives features, which are further processed and analyzed in the following layers. The output of the previous layer can be used to calculate the output of the next layer. The

structure of DNN follows a linear relationship and activation functions are used to add nonlinear factors to solve problems that cannot be solved by linear models

$$\mathbf{z} = \sum \boldsymbol{\omega}_i \mathbf{x}_i + \mathbf{b} \tag{1}$$

while linear relationship coefficient matrices $\boldsymbol{\omega}$ and bias vector $\mathbf{b}$ are used to perform a series of linear operations with the input value vector $\mathbf{x}$, and select the suitable activation function for activation operations. The structure of the certain DNN model is carefully defined via a trial-and-error experiment with the corresponding dataset to obtain the best-performed model. The activation function we used is the logarithmic sigmoid, relu, tanh function, and the network type of each layer is dense

$$\text{Sigmoid: } f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

$$\text{Relu: } f(x) = \begin{cases} -x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{3}$$

$$\tanh: f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{4}$$

where x is the input signal to the neuron and f is the activation function. All of the optimized parameters are listed in Table S2 (Supporting Information).

Data Dimension Reduction: The dimensionality reduction of the extracted data was conducted by Uniform Manifold Approximation and Projection (UMAP) method. UMAP was performed on the data matrix using umap package of the R software. Each of the individual spot on the 2D space was annotated by distinctive color according to their enzyme-like type.

Loss Function: Here, a common loss function, mean square error (MSE), was used to show the training process of the model and estimate the deviation between the predicted value and the real value. MSE is a more convenient method to measure the "average error". The smaller the value of MSE, the better the accuracy of the prediction model to describe the experimental data

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

where y_i is the real data, $\hat{y}_i$ is the fitted data, and n is the number of samples.

Model Validation Analysis: Model validation in this study consists of two parts, including data validation and experimental validation. Data validation was performed by randomly extracted 20% data from the full dataset. For experimental validation, the nanomaterials were first synthesized in terms of requirement of research. The characterization parameters were input into the well-trained machine learning models to obtain the predicted results. The enzyme-mimicking types or the level of enzyme-like activity of the nanomaterials were evaluated by the standard procedure. The validation was finally conducted by comparison of the model prediction and the experimental result.

The following metric was used to evaluate the accuracy of the classification model

$$\text{Accuracy} = \frac{N'}{N} \times 100\% \tag{6}$$

N is the number of samples, N' is the number of the sample whose model outputs is consistent with the real value in the validation set.

For confusion matrix analysis of the classification model, the records were summarized in the form of a matrix according to two categories: real enzyme-mimicking types and predicted enzyme-mimicking types. The row of the matrix represents the predicted enzymatic types by the model, and the column is the real enzymatic types. Kappa consistency check (κ) is a general statistic that can be used for classification systems. The κ analysis was conducted based on confusion matrix. The κ coefficient measures the agreement between classification and truth values.

$$\kappa = \frac{\frac{\sum_{i=1}^{r} x_{ii}}{N} - \frac{\sum_{i=1}^{r} x_i . x_i}{N^2}}{1 - \frac{\sum_{i=1}^{r} x_i . x_i}{N^2}} \tag{7}$$

where N is the number of all sample values and x_i is the number at the corresponding position of the confusion matrix.

The accuracies of the quantitative models (i.e., the level of enzymatic activity) were evaluated by the R^2 coefficient as below

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{8}$$

where x_i is the observed value, y_i is the predicted value, and $\overline{x}$ is the mean value of x. The R^2 shows the variance between the predicted and the observed enzymatic activity.

Sensitivity Analysis: The sensitivity analysis of the models was determined by SHapley Additive exPlanation (SHAP), which represents the potential contribution of the input data for the model output.[20] Briefly, the activation degree of each neuron in the neural network was compared with the reference value, and the score was assigned to infer the contribution of the input. For each predicted sample, the model produced a predicted value, and the SHAP value was the contribution assigned to each feature in the sample. Assuming that the jth feature of the ith sample is x_{ij}, the model predicts y_i for the ith sample, and the baseline of the entire model (usually the mean value of the target variable of all samples) is y_{base}, then the SHAP value obeys the following equation

$$y_i = y_{\text{base}} + f(x_{i1}) + f(x_{i2}) + \cdots + f(x_{ij}) \tag{9}$$

Intuitively, $f(x_{ij})$ is the contribution value of the j^{th} feature in the i^{th} sample to the final predicted value y_i.

The SHAP values were given by SHAP package of Python. For the classification model, the sensitivity of the independent variables was ranked in Figure 2c, as evidenced by SHAP values. The sensitivity analysis of the quantitative models was conducted and listed in Figure 5a,b.

Synthesis of Nanozymes: Various metal-based nanozymes were prepared by different methods. The detailed preparation processes are listed as below.

Fe_3O_4 Nanoparticles:[21] $FeCl_3 \cdot 6H_2O$ were mixed with ethylene glycol under magnetic stirring. NaAc and poly(acrylic acid) were added into the mixture, followed by vigorously stirring for 40 min. Then the mixture was sealed in a 25 mL autoclave, heated at 200 °C for 12 h. After cooling to room temperature, the resulting products were collected by carefully washing several times with dehydrated ethanol and then vacuum dried at 60 °C for 12 h.

CeO_2 Nanoparticles:[22] 1 mL acetic acid was dropwise added into a solution of 0.5 g $Ce(NO_3) \cdot 6H_2O$ in 30 mL solvent (distilled water: ethylene glycol = 1:2), following by the addition of 0.8 g poly(vinylpyrrolidone). Consequently, the whole mixture was sealed in a Teflon autoclave and solvothermal treated at 180 °C for 22 h. After cooling to room temperature, the precipitation was washed by dehydrated ethanol for three times and dried at 60 °C for 10 h. The resulting product was annealed at 500 °C for 2 h for further use.

CuO Nanoparticles: 4 g $CuAc_2 \cdot H_2O$ was dissolved in 50 mL distilled water to obtain solution A. Meanwhile, 1.6 g NaOH was dissolved in 40 mL distilled water to get solution B. Consequence, solution A and solution B were well mixed and further sealed in Teflon autoclave and hydrothermally treated at 110 °C for 2 h. Thoroughly washing and vacuum drying were conducted to purify the product.

MnO_2 Nanoparticles:[23] 94 mg $MnSO_4$ and 873 mg $Na_2S_2O_8$ were dissolved in 15 mL distilled water containing 1mL H_2SO_4. After vigorous stirring for about 10 min, the mixture was sealed in Teflon autoclave and hydrothermally treated at 110 °C for 1 h. The obtained product was washed thoroughly and dried at 60 °C for 10 h for further use.

$Co(OH)_2$ Nanoparticles: 0.05 M NaOH solution was dropwise added to the 0.2 M $Co(NO_3)_2{\cdot}6H_2O$ solution with a ratio of Co^{2+}:OH^- = 4:1. The resulting mixture was sealed in Teflon autoclave and hydrothermally treated at 180 °C for 6 h. After routine purification, $Co(OH)_2$ nanoparticles were obtained for further use.

Ag Nanoparticles:[24] 0.5 g of poly(vinylpyrrolidone) was dispersed in 15 mL ethylene glycol and heated at 180 °C under vigorous stirring for 30 min. Then the temperature was adjusted to and kept at 120 °C, followed by adding the solution of 0.1 g $AgNO_3$ in 5 mL ethylene glycol. The reaction was kept at 120 °C for 3 min, with the generation of brown colloidal which indicates the formation of Ag nanoparticles. After cooling down to room temperature, acetone was added for product precipitation, which was further purified by repeated washing and finally redispersed in distilled water for further use.

Ru Nanoparticles:[25] 0.5 mmol NaOH was mixed with a solution of 0.1 mmol $RuCl_3{\cdot}H_2O$ in 15 mL ethylene glycol. The mixture was heated at 160 °C for 2 h under vigorous stirring and Ar atmosphere. After purification and concentration, Ru nanoparticles were obtained.

Pd Nanoparticles:[26] 15 mL H_2PdCl_4 (2.0×10^{-3} M) was added to the ethanol: water (volume ratio, 1:4) solution, followed by adding poly(vinylpyrrolidone). Subsequently, the resulting solution was refluxed for 4 h to remove ethanol. The resulting mixture dialyzed against deionized water. After purification and concentration, Pd nanoparticles were obtained for further use.

Rh Nanoparticles:[27] 42 mg $RhCl_3{\cdot}xH_2O$, 336 mg TTAB and 444 mg poly(vinylpyrrolidone) were added into a three-necked flask containing 20 mL ethylene glycol and mixed thoroughly. The mixture was deaired in an Ar atmosphere and was maintained at 80 °C for ≈20 min until the solution turned dark brown. The temperature was then raised to 185 °C for 1.5 h under an Ar atmosphere. When the reaction completed, an excess of acetone was added at room temperature to form a cloudy black suspension before separation of nanoparticles by centrifugation. After purification and concentration, Rh nanoparticles were obtained for further use.

Preparation of FTn-Based Nanozymes: For the synthesis of FTn-based nanozymes, different metals were incorporated into FTn protein scaffold using the identical procedure.

FTn protein scaffolds were first obtained by our previous reports.[28] Next, metal-based nanoparticles were in situ biomineralized into the cavity of FTn. Briefly, FTn in NaCl solution were deaired by Ar atmosphere, followed heating at 65 °C in water bath. MCl_x (M = Fe, Co, Ni, Mn) solution were mixed with the above mixture. H_2O_2 solution was slowly added into the solution at constant pH 8.5. The resulting solution was further purified with PD-10 desalting columns. For other metals, FTn in PBS buffer was mixed with metal-containing solutions (Ru, Rh, Pt, Pd, Ir, Au). After vigorous stirring for 1 h, the mixture was purified by PD-10 desalting column to remove the free metal ions. Sodium borohydride was slowly added into the mixture, following by magnetic stirring under ice bath for 2 h. The resulting solution was purified for further use.

Characterization of Nanozymes: The morphology of the nanozymes was characterized by transmission electron microscopy (TEM) (HITACHI HT7700 Exalens) and high-resolution transmission electron microscopy (HR-TEM) (JEOL-JEM-2800). The native FTn was characterized by TEM with the negative staining of the specimen with 1% uranyl acetate for 1 min, and FTn-based nanozymes were observed by HR-TEM without negative staining. The X-ray powder diffraction (XRD) spectrum was measured by Rigaku SmartLab (Rigaku, Japan), with a diffractometer using Cu Kα radiation (40 KV and 150 mA) and scan rate 15° min^{-1}, 2–80° 2 theta range. The incorporation of metals into FTn was determined by measuring metal concentration with inductively coupled plasma atomic emission spectroscopy (ICP-OES) (Thermo Fisher ICAP 7400) and FTn protein concentration via SDS-PAGE gel electrophoresis.

Measurement of Enzyme-Like Activity: The enzyme-like activities of nanomaterials were determined by catalyzing the corresponding substrates in the standard procedure.

The POD-like activity test was carried out in 0.2 M sodium acetate buffer solution (pH 4.5) with 3′,3′,5,5′-tetramethylbenzidine (TMB) as the substrate in the presence of H_2O_2. The absorbance of the oxidized TMB was recorded under 650 nm every 10 s for 10 min via a Multiskan FC microplate spectrophotometer (Thermo Fisher Scientific). The kinetic parameters were determined based on the Michaelis–Menten equation by GraphPad Prism v8.0.

For the OXD-like activity, the tests were conducted in 0.2 M sodium acetate buffer solution (pH 4.5) with TMB as the substrate in the absence of H_2O_2. The absorbance of the oxidized TMB was recorded under 650 nm every 10 s for 10 min via a microplate spectrophotometer. The kinetic parameters were determined based on the Michaelis–Menten equation by GraphPad Prism v8.0.

For measurement of SOD-like activity, different nanozymes were added to 0.1 M PB buffer (pH 7.8) containing 13.2×10^{-3} M L-methionine, 2×10^{-6} M riboflavin, 0.25×10^{-3} M Nitrotetrazolium Blue chloride (NBT), then the solutions were irradiated by white light (0.03 W cm^{-2}) for 5 min. $O_2^{\bullet-}$ was captured by NBT as it was eliminated by nanozymes. Therefore, the SOD-like activity of nanozymes was expressed as the inhibition percentage of the competitive NBT reaction with nanozymes.

The CAT-like activity was performed at room temperature by measuring the generated oxygen using a specific oxygen electrode on multiparameter analyzer (JPBJ-609L, Leici China). The kinetic assay of nanozymes with H_2O_2 as the substrate was performed in 8.0 mL 0.1 M PB (pH 7.0) buffer solution by adding 80 μg (0.01 mg mL^{-1}) nanozymes and different concentrations of H_2O_2 (0–176×10^{-3} M). The generated O_2 solubility was measured every 5 s for 5 min. The kinetic parameters were determined based on the Michaelis–Menten equation by GraphPad Prism v8.0.

Statistical Analysis: All statistical analysis was conducted by R software and GraphPad Prism v8.0. One-way analysis of variance (ANOVA) and Kruskal–Wallis tests were used to compare among multiple groups.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Y.W. and J.W. contributed equally to this work. We acknowledge the support by the National Natural Science Foundation of China (91959129, 82072054, and 31870999), Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICIP-KJGG-014-03), Nankai University Hundred Young Academic Leaders Program.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

machine learning, nanomaterials, nanozyme

Received: February 22, 2022
Revised: April 20, 2022
Published online: June 3, 2022

[1] a) D. Jiang, D. Ni, Z. T. Rosenkrans, P. Huang, X. Yan, W. Cai, *Chem. Soc. Rev.* **2019**, *48*, 3683; b) M. M. Liang, X. Y. Yan, *Acc. Chem. Res.* **2019**, *52*, 2190; c) J. J. X. Wu, X. Y. Wang, Q. Wang, Z. P. Lou, S. R. Li, Y. Y. Zhu, L. Qin, H. Wei, *Chem. Soc. Rev.* **2019**, *48*, 1004.

[2] a) J. Orozco, V. Garcia-Gradilla, M. D'Agostino, W. Gao, A. Cortes, J. Wang, *ACS Nano* **2013**, *7*, 818; b) D. M. Duan, K. L. Fan, D. X. Zhang, S. G. Tan, M. F. Liang, Y. Liu, J. L. Zhang, P. H. Zhang, W. Liu, X. G. Qiu, G. P. Kobinger, G. F. Gao, X. Y. Yan, *Biosens. Bioelectron.* **2015**, *74*, 134.

[3] a) Z. Chen, H. Ji, C. Liu, W. Bing, Z. Wang, X. Qu, *Angew. Chem., Int. Ed.* **2016**, *55*, 10732; b) G. Fang, W. F. Li, X. M. Shen, J. M. Perez Aguilar, Y. Chong, X. F. Gao, Z. F. Chai, C. Y. Chen, C. C. Ge, R. H. Zhou, *Nat. Commun.* **2018**, *9*, 129.

[4] F. M. Wang, Y. Zhang, Z. W. Liu, J. S. Ren, X. G. Qu, *Nanoscale* **2020**, *12*, 14465.

[5] a) S. S. Li, L. Shang, B. L. Xu, S. H. Wang, K. Gu, Q. Y. Wu, Y. Sun, Q. H. Zhang, H. L. Yang, F. R. Zhang, L. Gu, T. R. Zhang, H. Y. Liu, *Angew. Chem., Int. Ed.* **2019**, *58*, 12624; b) K. L. Fan, J. Q. Xi, L. Fan, P. X. Wang, C. H. Zhu, Y. Tang, X. D. Xu, M. M. Liang, B. Jiang, X. Y. Yan, L. Z. Gao, *Nat. Commun.* **2018**, *9*, 1440.

[6] L. Gao, J. Zhuang, L. Nie, J. Zhang, Y. Zhang, N. Gu, T. Wang, J. Feng, D. Yang, S. Perrett, X. Yan, *Nat. Nanotechnol.* **2007**, *2*, 577.

[7] a) T. F. Liu, B. W. Xiao, F. Xiang, J. L. Tan, Z. Chen, X. R. Zhang, C. Z. Wu, Z. W. Mao, G. X. Luo, X. Y. Chen, J. Deng, *Nat. Commun.* **2020**, *11*, 2788; b) M. J. Manto, P. F. Xie, C. Wang, *ACS Catal.* **2017**, *7*, 1931; c) M. Diez-Castellnou, F. Mancin, P. Scrimin, *J. Am. Chem. Soc.* **2014**, *136*, 1158; d) A. Chevalier, Y. Zhang, O. M. Khdour, J. B. Kaye, S. M. Hecht, *J. Am. Chem. Soc.* **2016**, *138*, 12009; e) Y. Y. Huang, J. S. Ren, X. G. Qu, *Chem. Rev.* **2019**, *119*, 4357.

[8] a) H. J. Zhang, X. Liang, L. Han, F. Li, *Small* **2018**, *14*, 1803256; b) P. Zhang, D. Sun, A. Cho, S. Weon, S. Lee, J. Lee, J. W. Han, D. P. Kim, W. Choi, *Nat. Commun.* **2019**, *10*, 940.

[9] W. Feng, X. Han, H. Hu, M. Chang, L. Ding, H. Xiang, Y. Chen, Y. Li, *Nat. Commun.* **2021**, *12*, 2203.

[10] J. Suntivich, H. A. Gasteiger, N. Yabuuchi, H. Nakanishi, J. B. Goodenough, Y. Shao-Horn, *Nat. Chem.* **2011**, *3*, 546.

[11] a) K. Hornik, M. B. Stinchcombe, H. White, *Neural Networks* **1989**, *2*, 359; b) Y. LeCun, Y. Bengio, G. E. Hinton, *Nature* **2015**, *521*, 436.

[12] a) J. M. Granda, L. Donina, V. Dragone, D. L. Long, L. Cronin, *Nature* **2018**, *559*, 377; b) Y. Zhao, L. Wang, J. Luo, T. Huang, S. Tao, J. Liu, Y. Yu, Y. Huang, X. Liu, J. Ma, *Environ. Sci. Technol.* **2019**, *53*, 13238; c) G. Yamankurt, E. J. Berns, A. Xue, A. Lee, N. Bagheri, M. Mrksich, C. A. Mirkin, *Nat. Biomed. Eng.* **2019**, *3*, 318.

[13] S. Coombes, J. G. Taylor, *Neural Networks* **1996**, *9*, 837.

[14] a) Z. Ban, P. Yuan, F. Yu, T. Peng, Q. Zhou, X. Hu, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 10492; b) X. Wang, X. J. Gao, L. Qin, C. Wang, L. Song, Y. N. Zhou, G. Zhu, W. Cao, S. Lin, L. Zhou, K. Wang, H. Zhang, Z. Jin, P. Wang, X. Gao, H. Wei, *Nat. Commun.* **2019**, *10*, 704; c) H. C. Tao, T. Y. Wu, M. Aldeghi, T. C. Wu, A. Aspuru-Guzik, E. Kumacheva, *Nat. Rev. Mater.* **2021**, *6*, 701.

[15] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, *Nature* **2016**, *533*, 73.

[16] a) L. McInnes, J. Healy, J. Melville, arXiv:1802.03426, **2018**; b) S. Sakaue, J. Hirata, M. Kanai, K. Suzuki, M. Akiyama, C. Lai Too, T. Arayssi, M. Hammoudeh, S. Al Emadi, B. K. Masri, H. Halabi, H. Badsha, I. W. Uthman, R. Saxena, L. Padyukov, M. Hirata, K. Matsuda, Y. Murakami, Y. Kamatani, Y. Okada, *Nat. Commun.* **2020**, *11*, 1569.

[17] M. T. Ribeiro, S. Singh, C. Guestrin in *KDD '16: Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, ACM, New York **2016**, pp. 1135–1144.

[18] K. Fan, C. Cao, Y. Pan, D. Lu, D. Yang, J. Feng, L. Song, M. Liang, X. Yan, *Nat. Nanotechnol.* **2012**, *7*, 459.

[19] C. Rudin, *Nat. Mach. Intell.* **2019**, *1*, 206.

[20] a) S. M. Lundberg, G. G. Erion, S.-I. Lee, arXiv:1802.03888, **2018**; b) S. M. Lundberg, S.-I. Lee, arXiv:1705.07874, **2017**.

[21] H. Deng, X. Li, Q. Peng, X. Wang, J. Chen, Y. Li, *Angew. Chem., Int. Ed.* **2005**, *44*, 2782.

[22] X. Jiang, J. Zhang, L. Yu, R. Chen, X. Xu, *Micro Nano Lett.* **2016**, *11*, 137.

[23] S. Zhao, T. Liu, D. Shi, Y. Zhang, W. Zeng, T. Li, B. Miao, *Appl. Surf. Sci.* **2015**, *351*, 862.

[24] L. Li, J. Sun, X. Li, Y. Zhang, Z. Wang, C. Wang, J. Dai, Q. Wang, *Biomaterials* **2012**, *33*, 1714.

[25] Y. Chen, K. Y. Liew, J. Li, *Mater. Lett.* **2008**, *62*, 1018.

[26] T. Teranishi, M. Miyake, *Chem. Mater.* **1998**, *10*, 594.

[27] Y. Zhang, M. E. Grass, J. N. Kuhn, F. Tao, S. E. Habas, W. Huang, P. Yang, G. A. Somorjai, *J. Am. Chem. Soc.* **2008**, *130*, 5868.

[28] a) X. Huang, J. Chisholm, J. Zhuang, Y. Xiao, G. Duncan, X. Chen, J. S. Suk, J. Hanes, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E6595; b) X. Huang, J. Zhuang, S. W. Chung, B. Huang, G. Halpert, K. Negron, X. Sun, J. Yang, Y. Oh, P. M. Hwang, J. Hanes, J. S. Suk, *ACS Nano* **2019**, *13*, 236.