Quantum machine learning-based electrokinetic mining for the identification of nanoparticles and exosomes with minimal training data

Abhimanyu Thakur [a,b,*,1], Pedro Correia Santos Bezerra [c], Abhishek [d], Shihao Zeng [e],
Kui Zhang [a,b], Werner Treptow [f], Alexander Luna [g], Urszula Dougherty [g], Akushika Kwesi [g],
Isabella R. Huang [h], Christine Bestvina [i], Marina Chiara Garassino [j], Fuyu Duan [a,b],
Yash Gokhale [a,b], Bin Duan [k], Yin Chen [l], Qizhou Lian [m,n,o,p], Marc Bissonnette [g],
Jianpan Huang [e,**], Huanhuan Joyce Chen [a,b,***]

[a] Pritzker School of Molecular Engineering, University of Chicago, Illinois, USA
[b] Ben May Department for Cancer Research, University of Chicago, Illinois, USA
[c] The University of Chicago Booth School of Business, Chicago, IL, USA
[d] Department of Analytics, Northeastern University, Boston, MA, USA
[e] Department of Diagnostic Radiology, The University of Hong Kong, 5 Sassoon Road, Hong Kong, China
[f] Theoretical and Computational Biology Lab (LBTC) from Universidade de Brasília (UnB), Brazil
[g] Section of Gastroenterology, Hepatology and Nutrition, Department of Medicine, The University of Chicago, Chicago, IL, USA
[h] The Laboratory Schools, The University of Chicago, Chicago, IL, USA
[i] Department of Medicine, Section of Hematology/Oncology, The University of Chicago, Chicago, IL, USA
[j] Department of Hematology/Oncology, The University of Chicago, Chicago, IL, USA
[k] Mary & Dick Holland Regenerative Medicine Program and Division of Cardiology, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, USA
[l] Department of Pharmacology and Toxicology, College of Pharmacy, University of Arizona, Tucson, AZ, USA
[m] Faculty of Synthetic Biology, Shenzhen University of Advanced Technology, Key Laboratory of Quantitative Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[n] Center for Translational Stem Cell Biology, Hong Kong, China
[o] HKUMed Laboratory of Cellular Therapeutics, University of Hong Kong, Hong Kong 999077, China
[p] Department of Surgery, The University of Hong Kong Shenzhen Hospital, Shenzhen 518053, China

ARTICLE INFO

ABSTRACT

Synthetic and naturally occurring particles, such as nanoparticles (NPs) and exosomes; a type of extracellular vesicles (EVs), have garnered widespread attention across various fields, including biomaterials, oncology, and delivery systems for drugs and vaccines. Traditional methods for identifying NPs and EVs, such as transmission electron microscopy, are often prohibitively expensive and labor-intensive. As an alternative, the assessment of electrokinetic attributes such as zeta potential or electrophoretic mobility, conductance, and mean count rate, offers a more cost-effective, rapid, and reliable means of characterizing these particles. In this context, we introduce the first application of a quantum machine learning (QML)-based electrokinetic mining for the identification of green-synthesized iron- and cobalt-based NPs, as well as exosomes derived from human embryonic stem cells (hESC), human lung cancer (A549) cells, and colorectal cancer (CRC) cells, based solely on their electrokinetic attributes. Comparative analyses involving cross-validation, train-test splits, confusion matrices, and Receiver Operating Characteristic (ROC) curves revealed that classical ML techniques could accurately identify the types of NPs and EVs. Notably, QML demonstrated proficiency in differentiating between various NPs and EVs, including the distinction of EVs in the plasma of CRC patients versus those of healthy individuals. Furthermore, QML's application has been extended to the identification of NPs along with EVs in the plasma of

CRC patients and experimental mice, achieving higher prediction performance even with a minimal training dataset, demonstrating that QML based electrokinetic mining could identify NPs or EVs with minimal training data, thereby facilitating novel clinical development in the realm of liquid biopsies.

1. Introduction

Both nanoparticles (NPs) and extracellular vesicles (EVs) have widespread applications in several disciplines of biomedical sciences, including targeted delivery for cancer therapy [1], tissue engineering [2], vaccine delivery, and biosensing [3]. NPs are broadly composed of materials in the size range of approximately less than 100 nm [4]. EVs are a heterogeneous group of small, membrane-bound nanovesicles, such as exosomes and microvesicles, that vary in size, composition, and function [5–7]. Most applications of NPs and EVs require their identification and detailed characterization such as *in vivo* imaging, pharmacokinetics study and in liquid biopsies [8–10]. Alternative to the well-established quantitative structure–property relationship (QSPR) technique [11–15], modern approaches for identification and characterization of NPs currently rely on a combination of machine learning (ML) algorithms and cutting-edge molecular biophysics techniques, including X-ray diffraction (XRD), surface enhanced Raman spectroscopy (SERS), tunable resistive pulse sensing (TRPS), nuclear magnetic resonance (NMR) spectroscopy, thermogravimetric analysis (TGA), transmission electron microscopy (TEM), scanning electron microscopy (SEM), and atomic force microscopy (AFM) [16]. EVs are often characterized by similar methods, including TEM, immunogold-EM, nano-tracking analyzer (NTA), and biochemical techniques such as Western blotting, enzyme linked immunosorbent assay (ELISA), surface plasmon resonance (SPR), and AFM [5,6]. While these methods yield high-resolution data for characterizing NPs and EVs, several of these approaches necessitate manual imaging and/or feature analysis. These processes are not only costly and time-consuming but also demand specialized equipment and the expertise of highly trained personnel, which hinder automation. Current ML-based technologies suffer from lack of generalizability and scalability [17], as the development of large, well-annotated training datasets is also resource intensive.

ML-based methods have the potential to provide rapid, accurate, and reproducible characterization of NPs and EVs. However, the lack of large training datasets demands the development of novel approaches to improve *(i)* the scalability of experimental routines and/or *(ii)* the predictive power of ML models. To this end, we investigate the performance of quantum ML (QML) and classical ML to better characterize NPs (iron and cobalt) and EVs (from hESC, A549, and Caco-2 cells) based on electrokinetic properties (Supplementary Figs. S1A–C).

Electrokinetic properties have proven utility for the characterization of NPs [18,19], and EVs [6,20]. Electrokinetic properties including zeta potential (ZP), electrophoretic mobility (Mob), conductivity (Cond) and mean count rate (MCR), have been used to characterize NPs subjected to an applied electric field. Electrokinetic analysis provides valuable information regarding the characteristics of NPs as an important technique to quantify the chemical constituents of NPs. As NPs continue to grow in applications in the world of medical technology, it is essential to ensure they are safe and effective [21,22]. ZP, one of the electrokinetic properties, reflecting electrostatic interactions between particles, is crucial for distinguishing cancer and non-cancer cells, and their EVs. Cancer patients and healthy individuals exhibit different EVs due to variations in cellular origin, tumor microenvironment, and genetic alterations [23]. Factors like pH, hypoxia, and oncogenic signaling influence EV biogenesis and cargo, altering their electronic kinetic properties including ZP [6,24]. These changes affect EV surface charge, size, and membrane composition, impacting their function and interaction with recipient cells [25]. Altered surface charge in cancer cells, influenced by molecular changes, depicting malignancy, can be detected through ZP measurements. This distinction in electrokinetic properties serves as a valuable biomarker for cancer detection, facilitating liquid biopsy based early diagnosis [6,20]. Electrokinetic properties can be measured conveniently using dynamic light scattering (DLS). DLS offers several advantages including non-invasive measurements, accessibility even with small sample quantities and sensitivity to trace aggregation, thus achieving scalability, cost savings and more rapid analysis *(i)*. ML models have been to be very useful in disclosing complex patterns in several biological datasets [26–31]. As such, ML has been utilized in various aspects of NP studies including synthesis [32,33], evaluation of dynamics [34], and delivery to tumor [35]. ML has also been applied to study EVs in connection to cancer diagnosis via SERS [36,37] and elucidation of proteins' interactions contributing to cancer stemness and integrity of EVs [38]. The variety of applications to characterize NPs and EVs necessitates the need for new training data sets collected from different experiments. Therefore, for ML methods to be useful, the resource burden required to develop high-quality training data must be reduced.

QML is expected to require smaller training datasets than classical ML to achieve similar performance. This reduction in necessary training data stems from the allowed superposition of states in the two-level quantum-mechanical system. Classical ML algorithms use classical computers, which operate on bits that can be in one of two states, 0 or 1. Quantum computers, on the other hand, use qubits, which can be in the superposition of both states (0 and 1) at the same time. This allows QML algorithms to solve problems that are intractable for classical computers. Variational Quantum Classifier (VQC), which is a hybrid QML and has also been applied in different biomedical research such as dementia prediction [39]. The VQC has several advantages over classical ML algorithms, for example, quantum computers can represent probability distributions more accurately than classical computers. In addition, QML shows improved generalization from limited data compared to classical methods. Caro et al. demonstrated that QML models can effectively learn from small datasets, potentially offering advantages in various applications requiring minimal training data [40,41]. Therefore, successfully applying QML would address both scalability and predictive power with minimal training dataset.

NPs have been widely utilized as the vaccine adjuvants and delivery platforms, enhancing the immunogenicity and stability of antigens, and allowing for slow release and targeted delivery [42–44]. Distinguishing between different NPs is crucial owing to their distinct properties, toxicity profiles, reactivities, and performance characteristics [45]. Therefore, a robust method is required for the identification of type of NPs. Detecting NPs in blood plasma enables assessment of biodistribution, toxicity, drug delivery efficacy, and interactions with blood components, supporting nanomedicine development and ensuring safety in biomedical applications [46]. Moreover, the heterogeneity and overlapping characteristics of these NPs along with exosomes make their accurate identification and characterization challenging, particularly in the context of cancer detection and monitoring [7,47]. In the current study, we validated an electrokinetic approach to identify and characterize NPs and EVs in plasma samples of colorectal cancer (CRC) patients and experimental mice. Given the low cost, and ease of implementation and scalability of this technology, we anticipate that QML-based electrokinetic mining will likely become an auxiliary method to characterize NP or EV for biomedical research – and could uncover other types of biological nanovesicles suitable to for liquid biopsy research.

2. Results

2.1. Size and electrokinetic characterization of NPs and EVs

The morphology and size of green synthesized iron- and cobalt- NPs were characterized by TEM and NTA. Most of the NPs were found to be round, and a few of them look slightly elongated. The diameter of both types of NPs was found to be around 20 nm. (Fig. 1A and B). Electrokinetic analysis provides valuable information regarding the characteristics of NPs. It is one of the most important techniques used to quantify the chemical constituents of NPs. As NPs continue to gain popularity in the world of medical technology, it is essential to ensure they are safe and effective [21,22]. Besides, electrokinetic analysis can be used to identify and quantify the concentration of chemical components (such as hydroxyl groups and carboxyl groups) in NPs and assess their stability under various conditions [48]. This helps to understand the various biological and chemical properties of NPs and helps to improve the design process of these futuristic technologies. This technique provides valuable insight into the size, ZP, surface charge and stability of NPs. It also helps to identify chemical constituents present in NPs [49–52]. Therefore, it is imperative to examine electrokinetic characteristics to distinguish between different types of NPs. Fig. 1C, D shows the distribution of ZP of iron- and cobalt- NPs, which were found to be approximately 10 and 26 mV, respectively (Fig. 1E). Interestingly, the ZP of iron NPs was found to be distributed at a much wider range as compared to that of cobalt NPs (Fig. 1A and B). The electrophoretic mobility (Mob) of iron NPs was found to be significantly lower than that of cobalt NPs (Fig. 1F), which might be related to the lower deposition rate of cobalt NPs. Notably, the molecular mass of iron and cobalt are 55.845 and 58.93 Da, respectively. Therefore, as other factors, including their size seems to be similar based on the TEM analysis (Fig. 1A and B), the possible significant difference between Mob could be due to the difference in their molecular masses. In general, larger particles with higher mobility often achieve a lower deposition rate, whereas smaller particles with lower mobility often achieve a higher deposition rate [53, 54]. The Cond of Co NP was found to be significantly lower than that of Fe NP (Fig. 1G). Further, the record of time-dependent phase response for iron NPs was maximum around 5 rad, which was significantly lower than that of cobalt NPs (maximum around 60 rad) (Fig. 1H and I), inferring that the particle with steep curve responds rapidly, whereas the particle with a shallow curve is responding slowly. Similarly, silver nanoparticles (AgNPs) were characterized by TEM to depict morphology, and DLS-based electrokinetic analysis (Supplementary Figs. S2A–C). Further, Supplementary Figs. S3A–C depict the size distribution of EVs from hESC, A549, and Caco-2.

2.2. Identification of distribution patterns and relation among different electrokinetic features of NPs and EVs

A scatter plot analysis of four different electrokinetic features, namely ZP, Mob, Cond, and MCR, reveals the relationships between variables compared in a pairwise fashion (Fig. 2). For both NPs and EVs, ZP and Mob are highly correlated (Spearman correlation, P = 1) (Fig. 2Ai, Aii, Bi, Bii, Ei, Eii, Fi, Fii). This can be attributed to the direct proportionality between ZP and Mob defined by the Smoluchowski equation [55]. In the case of a plot of four given parameters, the data points are plotted along each axis to visualize the correlation between the variables (Fig. 2A–D i-iv). As shown in the scatterplot, ZP and Mob are clearly correlated. The data points along the ZP axis are clustered around the same line as Mob, which indicates that the relationship between these two variables is linear. By contrast, the other parameter pairs show no visible correlation. This could indicate that there is no significant relationship between them. The scatter plot of the four given parameters suggests that it is possible to classify the data points into two different types using conductance and mean count rate, together with one or both of ZP and mobility. Similarly, the scatterplot of the four

different electrokinetic attributes of EVs from hESC or A549 has been shown in Fig. 2E–H i-iv, and that of EVs from hESC or Caco-2 has been shown in Supplementary Figs. S4A–D.

2.3. Cross-validation, train-test split, confusion matrix and ROC curve analysis established effectiveness and predictive power of classical ML models to identify NPs and EVs

Cross-validation is a resampling method used to evaluate ML models on a limited data set. Cross-validation ensures that the model is free of bias due to the data used to train it by preventing the risk of overfitting [56]. As ZP and Mob are linearly correlated [55], which is also supported by our analysis (Fig. 2, Supplementary Figs. S4A–D), we excluded Mob from downstream analysis. K-fold cross validation [57], was used with 10, 100, and 200 folds to train classical ML models to differentiate between our two populations of NPs and two populations of EVs (Supplementary Table S1). Specifically, we trained five classical ML models for each classification task: linear regression (LR), support vector machine (SVM), K-Nearest Neighbors (KNN), decision tree (DT), and eXtreme Gradient Boosting (XGB). The data is divided into k folds, and the model was trained on k-1 folds and evaluated on the remaining fold. This process was repeated k times, and the average performance of the model is used to evaluate its performance. The train-test split is a method of training and evaluation of the performance of ML algorithms by splitting the data into 70:30 percent ratio [58]. A trained LR model showed the lowest prediction performance and the highest variance, whereas DT and XGB achieved the best prediction performance with minimal variance. The 200-fold cross-validation prediction performance of LR, SVM, KNN, DT, a XGB were found to be 0.7554 ± 0.227, 0.9108 ± 0.142, 0.9687 ± 0.098, 0.9987 ± 0.018, and 0.9987 ± 0.018, respectively (Supplementary Figs. S5A–C). Similar trends were found with 100- and 10- fold cross-validation (Supplementary Figs. S4D–I). When training with a train/test split, the prediction performance of LR, SVM, KNN, DT, and XGB on the held-out test set were found to be 0.75, 0.91, 0.96, 0.99, and 0.99, respectively for the identification of NPs (Supplementary Table S1). Similar strategy was used for the identification of hESC- or A549- EVs, and the prediction performance was determined by the LR, SVM, KNN, DT, and XGB on the held-out test. Apparently, the cross-validation of different ML models at different folds suggests their applicability for the prediction of NPs and EVs across the classical ML models. Confusion matrices for all the trained classical ML models for NPs and EVs (from hESC and A549 cells) are shown in Fig. 3A–E,I,M and Supplementary Fig. S6A-C, S7A-C. For each model, these demonstrate similar levels of false positives and false negatives. The precision, recall, and F1-score for each model is shown in Fig. 3B–F, J,N and Supplementary Fig. S6D-F, S7D-F. Area under the curve of ROC (AUC-ROC) for LR, SVM, KNN, DT, and XGB were 0.74, 0.89, 0.97, 1.0, and 1.0, respectively for the identification of NPs (Fig. 3C–G, Supplementary Figs. S6G–I). The AUC-ROC for LR, SVM, KNN, DT, and XGB for the identification of hESC- and A549 cells-derived EVs have been shown in Fig. 3K–O, and Supplementary Figs. S7G–I. All predictions from each model are displayed in three-dimensional feature space in Fig. 3D–H,L, P, and Supplementary Fig. S6J-L, S7J-L. By similar analysis, the confusion matrix, precision, recall, and F1-score, AUC-ROC, and three-dimensional scatter plots for hESC- or Caco-2- EVs have been shown in Supplementary Figs. S8A–O. A learning curve analysis shows that the performance of classical ML models such as LR, kNN, and DT were somewhat affected, and that of SVM and XGB were lower than the QML, when the training set was gradually reduced whereas the prediction performance for the prediction via QML was persistent, as depicted by the heatmap and plot (Supplementary Figs. S9A and B). In addition, the classical ML as well as QML models could distinguish AgNPs from FeNPs or CoNPs (Supplementary Figs. S10A–D). A gamified SHAP value analysis showed that Cond and MCR are the most significant features for classifying NPs and EVs, respectively (Fig. 3Q and R).
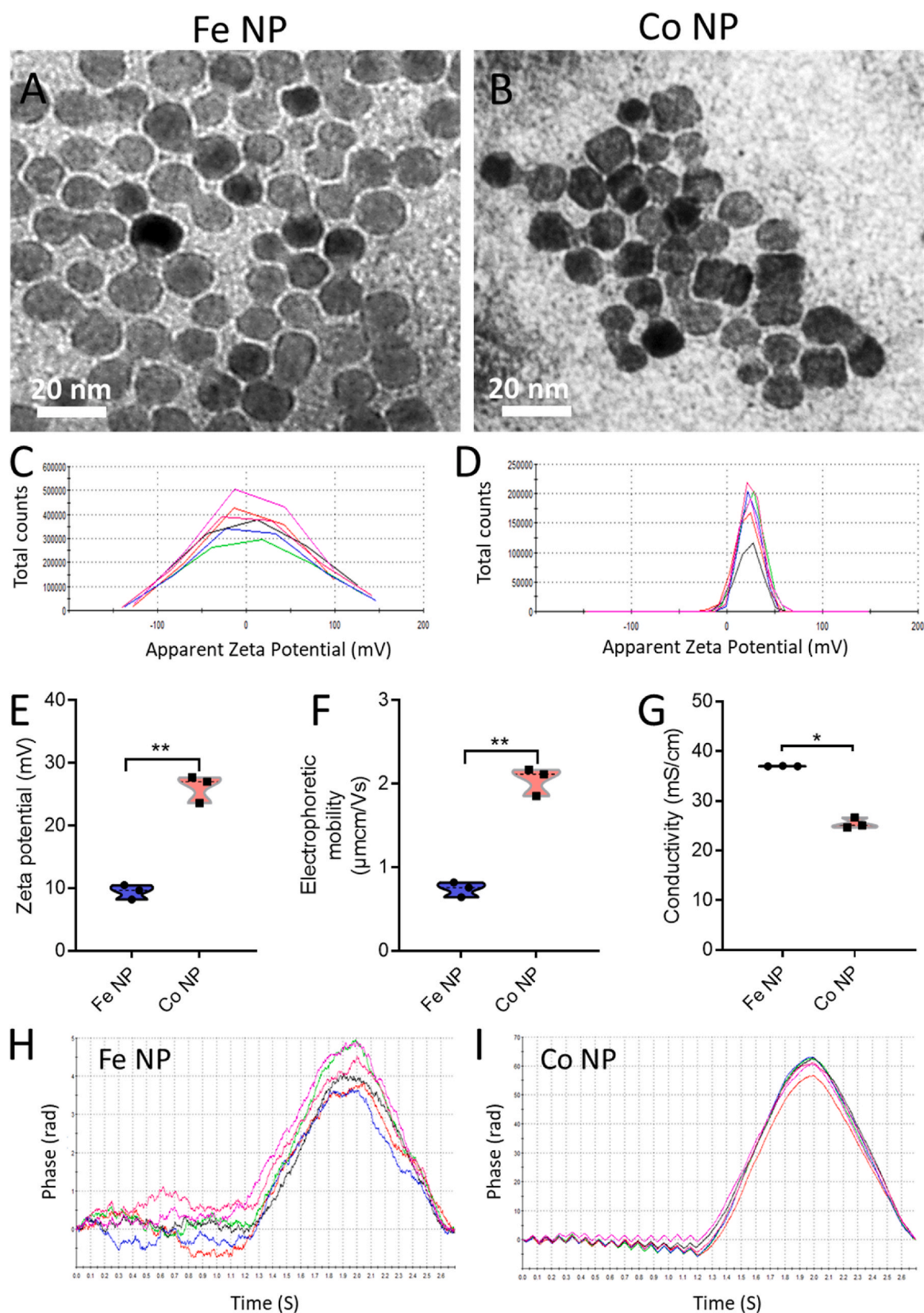
Fig. 1. Characterization and Electrokinetic analysis of iron- and cobalt-nanoparticles. **(A, B)** Representative transmission electron microscopy (TEM) image of (A) iron NPs, and (B) cobalt NPs. Scale bar: 20 nm. Both NP populations are mostly spherical in shape. **(C–D)** Representative graphs showing distribution of apparent ZP vs total count of (C) iron NPs, and (D) cobalt NPs for six samples of NPs. Each color represents a sample of NPs. **(E–G)** Bar graphs showing the mean - (E) zeta potential (ZP), (F) electrophoretic mobility (Mob), and (G) conductivity for iron (black) and cobalt (gray) NPs. Error bars display the standard error of the mean (N = 3). **(H, I)** Time-dependent phase of (H) iron NPs, and (I) cobalt NPs. Student's t-test was applied for comparison: iron NPs vs. cobalt NPs; Statistical significance was set at *P < 0.05, **P < 0.01.
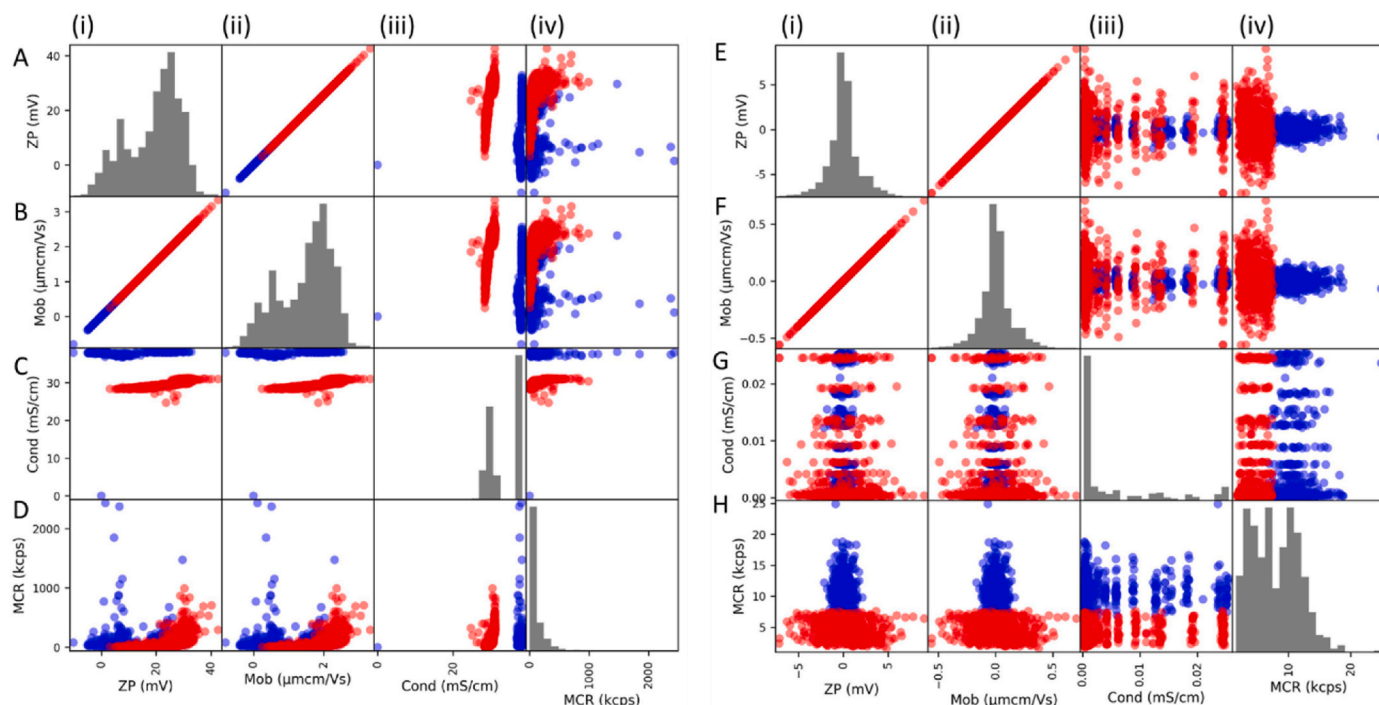
Fig. 2. Visualization of the data distribution of the raw datasets using scatter plot analy**sis. (A–D)** Pairwise comparisons of electrokinetic features, namely ZP, Mob, Cond, and MCR-for iron (blue) and cobalt (red) NPs (N = 1016). **(E–H)** Pairwise comparisons of electrokinetic features for A549- (blue) and hESC- (red) EVs (N = 1033).

2.4. Quantum ML based electrokinetic mining identifies NPs and EVs with fewer training sets

Classical ML algorithms use classical computers, which operate on bits that can be in one of two states, 0 or 1. Quantum computers, on the other hand, use qubits, which can be in a superposition of both states (0 and 1) at the same time (Supplementary Fig. S11). This allows QML algorithms to solve problems that are intractable for classical computers. VQC, which is a QML algorithm and has also been applied to different biomedical research problems such as dementia prediction [39]. The VQC has several advantages over classical ML algorithms, for example, quantum computers can represent probability distributions more accurately than classical computers. Intriguingly, the confusion matrix and AUC-ROC curve for the prediction of NPs via VQC with 70 % training set yielded an prediction performance of 94 % for the held out 30 % of test set (Fig. 4A–D,G). In addition, the VQC can be trained on much smaller datasets than classical ML algorithms, because the VQC can learn the parameters of the quantum circuit directly from the data, without the need for a large number of training set [39]. To evaluate the power of VQC for predicting NPs with lower training set, we reduced the training set to 50 %, and 10 % of training set, which still yielded the prediction performance of 94 %, and 92 %, respectively for the held out 30 % of test set (Fig. 4B,C,E,F, Supplementary Fig. S12). This demonstrates that VQC, which is a hybrid QML, could predict the NPs with better prediction performance even after training the QML with electrokinetic mining of fewer training set (Fig. 4G–I), inferring its advantage over classical ML. The comparative analysis of prediction performance via QML was found to be significantly better than the classical ML models with much lower training sets (Fig. 4J–P, Supplementary Fig. S13), suggesting towards the applicability of QML even with extremely minimal training set for electrokinetic mining to identify NPs.

A similar strategy, of QML based electrokinetic mining was utilized for the identification of EVs from hESC, A549, and Caco-2 cells (Supplementary Figs. S14 and S15). EVs play a significant role in intercellular communication and can carry various biological molecules, including proteins, lipids, DNA, and different RNA forms, which mimic the constituents of their parent cells [59]. hESC can be differentiated into specific cell lineages such as pulmonary neuroendocrine cells (PNECs) and progenitor cells, which can be utilized for disease modeling, drug discovery, and potential therapeutic applications [60–63]. Distinguishing among EVs from hESC, A549, and Caco-2 cells by QML based electrokinetic mining would offer a better approach for identification of EVs requiring lower training sets. Notably, the QML-based electrokinetic mining provided improved prediction performance for the identification of types of EVs, even with minimal training set (Supplementary Fig S14A-K, S15A-E).

2.5. QML detects NPs in mouse plasma and in both healthy and cancer patients via electrokinetic mining with minimal training data

To examine the feasibility of using QML for identifying the NPs in plasma, we utilized pre-clinical and clinical plasma samples without or with NPs to record their electrokinetic parameters. For pre-clinical samples, 10 mice were injected with 100 µl of AgNP solution via tail vein for 2 consecutive days, followed by isolation of EVs from their plasma and EVs from the plasma of 10 untreated mice were used as control group. Then, the electrokinetic parameters of samples from both the groups were recorded using Malvern Zetasizer Nano, followed by segregating the training and test data in the percentage ratio of 70:30. Furthermore, after keeping the test data separate, the percentage of training data was reduced, followed by training the classical ML and QML on different percentages of training set (Fig. 5A). Based on ROC analysis, the performance of QML were 85 % (AUC = 0.85, 95 % CI: 0.75–0.94), 80 % (AUC = 0.80, 95 % CI: 0.69–0.91), 70 % (AUC = 0.85, 95 % CI: 0.70–0.84), 63 % (AUC = 0.63, 95 % CI: 0.48–0.76) by training set at 70 %, 50 %, 10 %, and 1 %, respectively, to identify NPs containing plasma of mice. This was further illustrated with 3D scatter plots showing the separation between two groups by respective training sets **(5B-E).** The quantitative bar graph demonstrated that QML could maintain prediction performance even with mining of massively
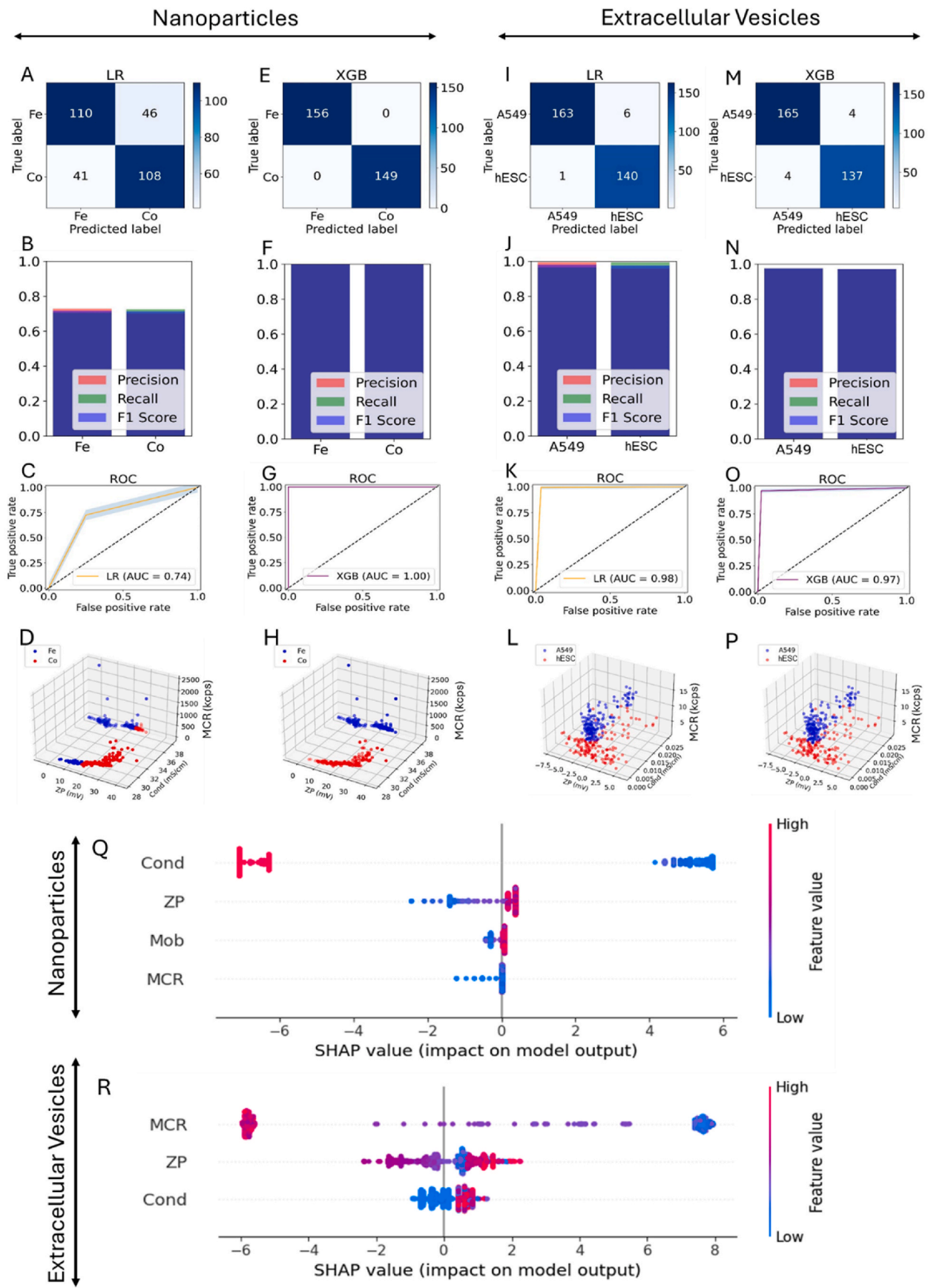
Nanoparticles

Extracellular Vesicles

(caption on next page)

Fig. 3. Performance evaluation of classical ML models to predict the types of NPs and EVs via confusion matrix and ROC cu**rve. (A–H)** Confusion matrices, binary classification metrics (precision, recall, F1 score), ROC, and three-dimensional scatter plots for each ML model trained for the identification of NPs. **(I–P)** Confusion matrices, binary classification metrics (precision, recall, F1 score), ROC, and three-dimensional scatter plots for each ML model trained for the identification of EVs. (A, E, I, M) The confusion matrix is a table that summarizes the performance of a classification model. The rows of the confusion matrix represent the true NP classes, while the columns represent the predicted NP classes by the ML models. (B, F, J, N) Precision measures the proportion of correctly predicted positive instances out of all predicted positive instances, while recall measures the proportion of correctly predicted positive instances out of all actual positive instances. The F1 score is the harmonic mean of precision and recall, and it is a measure of a model's overall performance. (C, G, K, O) The AUC-ROC is a measure of the overall performance of the model. A perfect model would have an AUC of 1, while a random model would have an AUC of 0.5. (D, H, L, P) Three-dimension scatter plots showing the distribution of predicted data via LR, and XGB on the held out 30 % test set. **(Q, R)** SHAP value analysis showing the contribution of major attributes for LR-based identification of (Q) NPs and (R) EVs. For the degree of uncertainty in sample was represented in (C,G,K,O) ROC plots via confidence interval (CI) with probability limit of 95 %.

reduced training set (Fig. 5F).

For the clinical samples, plasma was obtained from the ten healthy subjects and ten colorectal cancer (CRC) patients. Patients consented under IRB-10-209-A approved by the University of Chicago Institutional Review Board. The CRC patients ranged in age from 42 to 82, half of them were male and the other half were female. They were majorly from two different self-identified races including White, and African American, however one patient declined to identify race. In the context of ethnicity, except for one patient, all were non-Hispanic. All the cancers were identified as either low grade, invasive poorly differentiated, or invasive moderately differentiated. They were also categorized based on the tumor, node, and metastasis (TNM) based stages, and AJCC stage as reported in Supplementary Table S2. The control group consisted of 10 healthy subjects in the age range of 49–81 with half males and half females, as age-matched controls. The age-matched control group of healthy subjects were also either White or African American, in the context of both age and race-matched to patients. Like the group of CRC patients, all were non-Hispanic, except one (Supplementary Table S2). To examine the capacity of QML to identify NPs in plasma samples of healthy and cancer patients, NPs were spiked in the plasma samples, followed by the electrokinetic recording of control- and NPs-spiked healthy and cancer patients' plasma, and prediction performance by training QML with 70 %, 50 %, 10 %, and 5 % data (Fig. 5G). As evident from the ROC curves analysis with 70 % and 5 % training set, the QML could identify the NPs spiked plasma samples of healthy and cancer patients (Fig. 5H–J). Quantitative bar graphs showed that the prediction performance by the QML could be sustained even after massively reducing the percentage training data (Fig. 5I–K), suggesting that QML based electrokinetic mining could be a better approach for the identification of NPs in plasma samples.

2.6. QML distinguishes EVs and NP in healthy individuals vs. cancer patients

To examine how accurately QML based electrokinetic mining could distinguish between the EVs in the plasma of heathy subjects and CRC patients, the electrokinetic parameters of plasma-derived EVs were measured using the Malvern Zetasizer Nano, followed by segregating the training and test data in the percentage ratio 70:30. Further, after keeping the test data separate, the percentage of training data was reduced sequentially, followed by training the classical ML and QML on different percentages of training set. Additionally, the plasma derived EV samples were spiked with NPs, to determine if the QML based electrokinetic mining could distinguish healthy-from cancer- EVs with or without NPs (Fig. 6A). The prediction performance of different classical models; LR, kNN, SVM, DT, and XGB for the identification of NPs with EVs from the plasma of cancer patients are shown in Supplementary Figs. S16A and B. Importantly, the electrokinetic properties of NPs also change when they are diluted in plasma (Supplementary Figs. S17A–D). As per the ROC analysis, the prediction accuracies by QML were 69 % (AUC = 0.69, 95 % CI: 0.44–0.65) and 73 % (AUC = 0.73, 95 % CI: 0.55–0.77) by training set at 70 % and 5 %, respectively, for distinguishing healthy EVs versus cancer EVs. This suggests that, unlike classical ML, the VQC QML could distinguish the EVs from the plasma of

healthy subjects and cancer patients even with mining of 5 % training set (Fig. 6B–E). Similar analysis was carried out to distinguish between healthy EVs and healthy EVs spiked with NPs. As evident from the ROC analysis, the prediction accuracies by QML were 89 % (AUC = 0.89, 95 % CI: 0.68–0.86) and 77 % (AUC = 0.77, 95 % CI: 0.45–0.64) by training set at 70 % and 5 %, respectively, for distinguishing healthy EVs versus healthy EVs with NPs. This suggests that, unlike classical ML, the VQC QML could distinguish the EVs from the plasma of healthy subjects and cancer patients even with mining of 5 % training set (Fig. 6C–E). A comparable investigation was conducted to differentiate between cancer EVs and cancer EVs spiked with NPs. The prediction accuracies of QML were 73 % (AUC = 0.73, 95 % CI: 0.52–0.74) and 65 % (AUC = 0.65, 95 % CI: 0.65–0.89) by training set at 70 % and 5 %, respectively, for distinguishing cancer EVs versus cancer EVs with NPs, suggesting that QML could also distinguish between the plasma derived EVs mixed with or without NPs (Fig. 6D and E). Supplementary Figs. 18A–E showed the QML based electrokinetic mining with fewer observations could distinguish healthy EVs with NPs vs cancer EVs with NPs. The quantitative bar graph depicting the training set (%) versus prediction performance (%), demonstrated that QML could provide superior prediction performance even with mining of massively reduced training set (Fig. 6E).

3. Discussion

NPs and EVs are pivotal in a wide array of industrial domains, including biomedical research, nano-biomedicine, drug delivery, and vaccine development. They serve not only as effective diagnostic tools but also as efficient vehicles for drug delivery. With their ability to target specific molecules and structures, they can be used to identify and quantify biomarkers, remove waste, and help mitigate diseases. In addition, they could contribute to developing advanced materials like more durable paints, coatings, and additives. NPs also have a great potential in the energy sector, providing materials with higher energy outputs [22,26,28].

NPs have unique physical and chemical properties compared to larger particles, due to their small size. Therefore, appropriate identification and characterization of these particles is very important. This will help scientists understand their potential applications, risk assessment of their use, and help design effective strategies for safe handling. It is also important for drug manufacturers to ensure the quality of nanomaterials used in their products [22,32,64].

EVs are increasingly recognized for their potential in liquid biopsy applications, particularly in the field of oncology. EVs are secreted by most of the cell types including stem cells and cancer cells, and circulate in biofluid like blood, carrying a diverse array of biomolecules reflective of their cells of origin [5,65]. This makes them valuable for non-invasive cancer diagnosis, monitoring treatment response, and detecting therapeutic resistance through blood samples. The complexity and heterogeneity of EVs owing to their diverse size, biogenesis pathway, and overlapping biomarkers, pose huge challenges for their identification and analysis [66]. Intriguingly, ZP, an intrinsic electrokinetic property, illuminates the electrostatic interplay among particles, playing a pivotal role in discerning cancer from non-cancer cells and their EVs. The modified surface charge in cancer cells, shaped by molecular dynamics
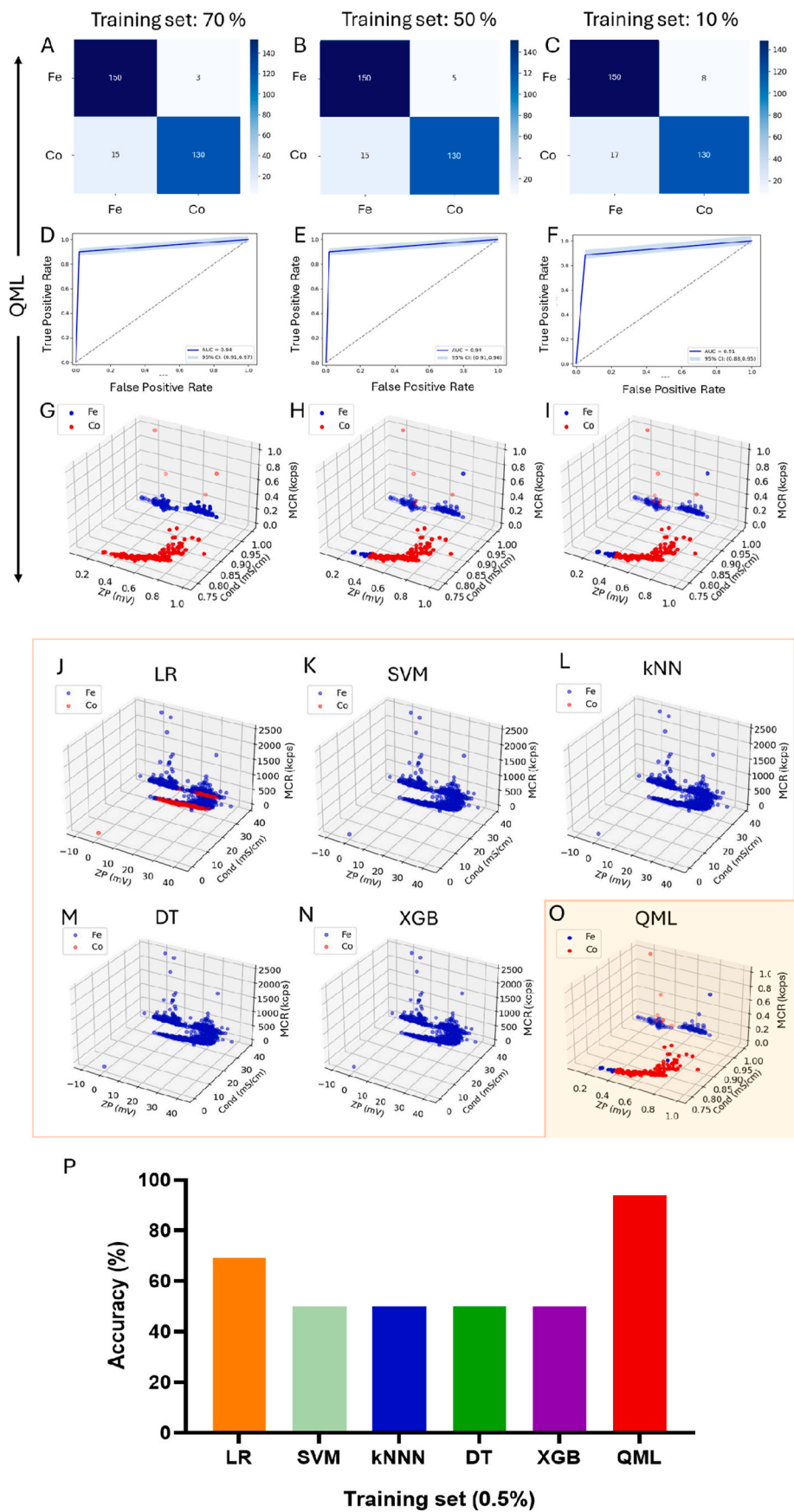
(caption on next page)

Fig. 4. Performance evaluation of quantum ML model and comparative analysis of its prediction performance with classical ML models with extremely small training set for the identification of NPs. Representative **(A–C)** confusion matrix, **(D–F)** ROC curves with confidence interval, and **(G–I)** three-dimensional scatter plots for QML with (A, D, G) 70 % training set, (B, E, H) 50 % of initial training set, and (C, F, I) 10 % of initial training set. The test accuracies were evaluated on the held out 30 % test set. The confusion matrix shows the number of correctly and incorrectly classified NPs for each type of NP. The ROC curve shows the trade-off between the true positive rate (TPR) and the false positive rate (FPR) for different thresholds. As can be seen from the figures, the VQC model can achieve high prediction performance in identifying the types of NPs, even with a smaller training set. **(J–P)** Representative three-dimensional scatter plots showing the NP prediction ability via (J–N) classical ML models: LR, SVM, KNN, DT, and XGB, and (O) VQC, a QML with 0.5 % training set. (P) Bar graph showing that QML outperformed the classical ML models for the identification of NPs, even when trained with 0.5 % training set. For precision estimation, the degree of uncertainty in sample was represented in (D–F) ROC plots via confidence interval (CI) with probability limit of 95 %.

indicative of malignancy, is unveiled through ZP analysis. This nuanced electrokinetic differentiation stands as a precious biomarker for cancer detection, enabling timely liquid biopsy-driven diagnoses [6,20]. In addition, the ZP of EVs derived from cancer cells, have been found to correlate with malignant characteristics of their parent cells, suggesting the utility of ZP as a surrogate biomarker of cancer [5,6,67,68]. The narrow range of negative surface charge in most EVs [20], reflects their consistent biophysical properties due to shared lipid and protein compositions. Despite this limited range, advanced ML models can extract realistic information by integrating electrokinetic data with other EV characteristics such as size, cargo, and origin to capture subtle variations and functional implications. Importantly, by leveraging multi-dimensional data and pattern recognition, ML models identify subtle correlations and variations, enhancing predictive accuracy despite limited surface charge range [28,69,70]. EVs from a single cell type exhibit heterogeneity in size, shape, and composition [66]. These factors influence electrokinetic properties like zeta potential by affecting surface charge distribution and hydrodynamic behavior. Electrokinetic measurements thus provide a comprehensive assessment of EV diversity [20].

EVs and NPs can interact when co-secreted, with NPs potentially adhering to EV surfaces or being encapsulated [71–73]. This interaction creates complex nanostructures with distinct electrokinetic properties [74]. ML models can differentiate silver NPs from EVs by analyzing subtle variations in electrokinetic parameters which also consider size, shape, surface charge, and biomolecular composition [75,76]. ML algorithms can recognize intricate patterns in these electrokinetic parameters, enabling accurate classification. This approach leverages the heterogeneity of EV populations and the unique properties of NPs to achieve precise separation and characterization [77,78], advancing our understanding of EV-NP interactions and their potential applications in nanomedicine.

Conventional techniques for the identification and characterization of NPs and EVs include TEM, SEM, AFM, XRD and DLS. TEM and SEM provide insights into the shape, size distribution and properties of the NPs. AFM provides information on the surface morphology of the NPs. XRD provides structural information on the crystalline structure of the NPs. DLS is used to characterize the size and polydispersity of the NPs. The drawbacks of these techniques include the inability to identify and characterize nanomaterials in complex matrices. The techniques are also expensive and require specialized equipment and personnel. Furthermore, these techniques are limited by imaging resolution, which is often too low to identify nanomaterials. Finally, some of these techniques require a sample to be exposed to radiation of high energy, which can alter the properties of the sample [11–16,79–82].

The application of ML models for the identification and characterization of NPs and EVs is critical for understanding their physicochemical properties and functionalities. ML models extensively utilize large-scale data mining approaches to capture and analyze physicochemical characteristics, like size, shape, and composition. ML models are also able to characterize and identify NPs through features like surface reactivity and safety. The advantages of using ML models for the identification and characterization of NPs include the high speed and accuracy in which the data can be visualized and characterized. Additionally, ML models can help identify difficulty detecting properties in NPs to better understand their structure and behavior in different contexts.

Ultimately, ML models applied to electrokinetic features offer an efficient and cost-effective way to identify and characterize NPs which would otherwise be difficult to acquire through conventional methods [26,27,34,83,84].

Electrokinetic analysis and data mining with the application of ML models provide a powerful way to identify and characterize NPs and EVs. This approach is useful in industries related to nanotechnology and biomedicine for assessing the particle size distribution, surface charge density, and concentration of nanomaterials. Additionally, ML models allow us to process large sets of data and identify and characterize nanomaterial characteristics quickly and accurately. This approach offers faster and more accurate outcomes with remarkably lower cost that enable better decision-making and efficient NP characterization [85–88].

The No-Free-Lunch Theorem implies that, when averaging over all possible objective functions, no optimization algorithm outperforms any other in all cases. There are many algorithms and hyperparameters, so it is impractical to test all possibilities. We investigated several algorithms, including LR, SVM, KNN, DT, and XGB. These five classical ML models were chosen for electrokinetic data mining due to their effectiveness with numerical data, interpretability, and efficiency. While the deep learning approach excels in complex analysis, these models offer transparency and establish a performance baseline [89]. The data were randomly split into 70 % training and 30 % testing sets. Each model was trained and tested by 10-, 100-, and 200- fold cross-validation. 10-fold cross-validation involves splitting the data into 10 subsets and using 9 subsets for training and 1 subset for testing. This was repeated 10 times, and the results were averaged. The dataset was shuffled to prevent overfitting (Supplementary Fig. S4, Supplementary Table S1).

The application of ML algorithms to electrokinetic analysis has been successful in yielding meaningful results. It could detect patterns in large datasets and make predictions based on the electrokinetic parameters of NPs, such as ZP, Mob, Cond, and MCR. The electrokinetic properties of NPs change when diluted in plasma due to the formation of a protein corona, altering their surface charge and mobility [90]. This affects their interactions with cells and tissues, impacting drug delivery efficacy and potential toxicity in clinical applications.

Using five types of supervised classical ML models namely LR, KNN, SVM, DT, and XGB, accuracy of 74 %, 89 %, 97 %, 100 %, and 100 % were respectively achieved for the identification of NPs (Fig. 3A–H). Similarly, classical ML models were found to be effective in identifying the types of EVs (Fig. 3I–P). This illustrates the effectiveness of classical ML-driven mining of electrokinetic parameters in identifying the NPs or EVs. As a limitation, the identical performance of DT and XGB could be potential due to overfitting. Typically, XGB outperforms single DTs on complex tasks due to its ensemble nature and boosting technique [91]. The lack of improvement observed here could suggest that either the dataset is relatively simple, allowing even a single DT to capture most patterns, or that both models might be overfitting to noise in the training data. This emphasizes the need for further investigation into model generalizability and dataset complexity in future studies. In addition, this could potentially be avoided by the implementation of batch-based data split in future.

QML is a promising new field that uses quantum mechanics to solve ML problems. Unlike classical ML, which uses bits that can be in one state or the other, QML uses qubits, which can be in a superposition of
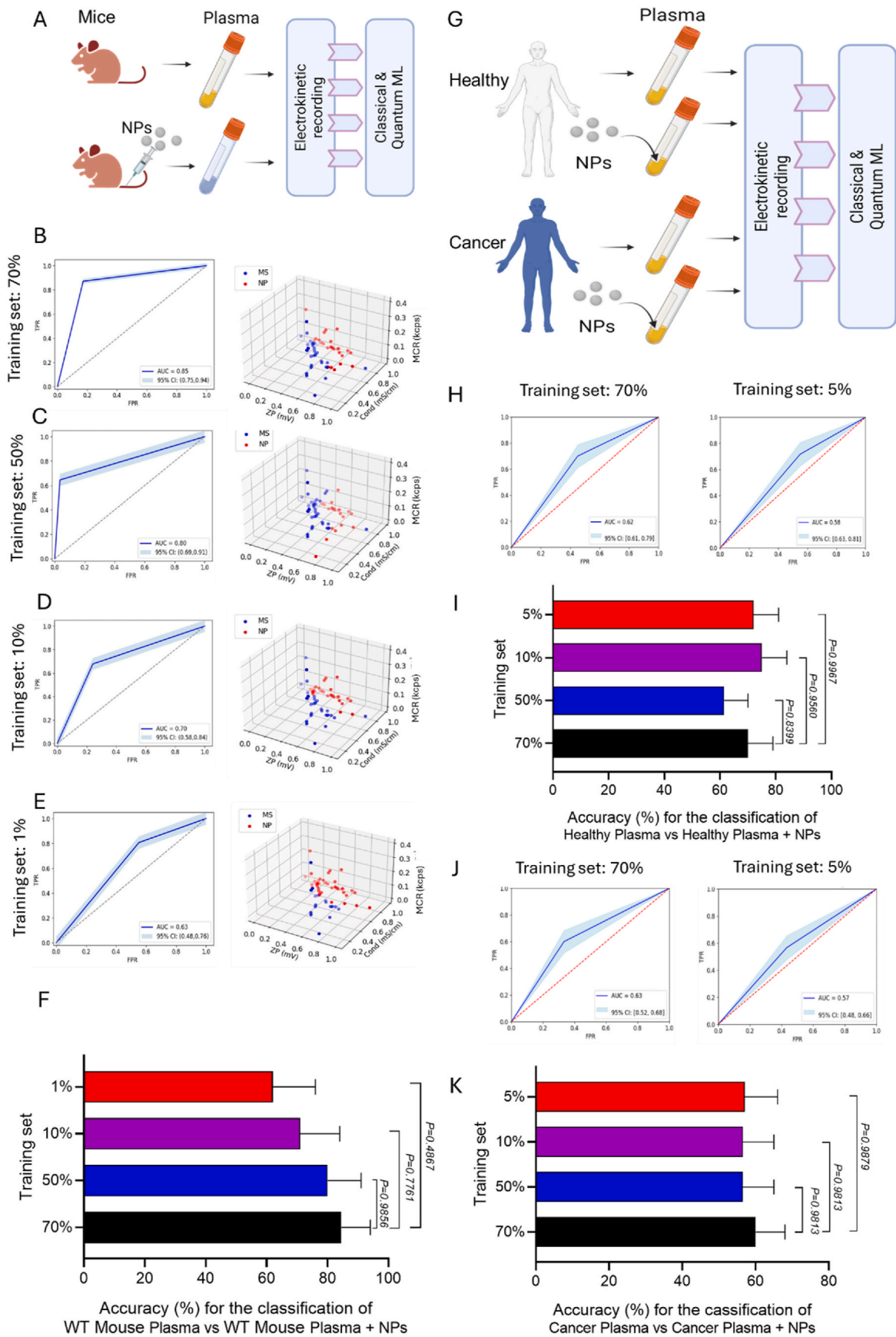
(caption on next page)

Fig. 5. QML based electrokinetic mining could detect the NPs in mouse plasma, and plasma from healthy individuals and cancer pat**ients. (A)** A schematic diagram showing the experimental plan for the collection of plasma from control- and NPs injected-mice, followed by recording of electrokinetic data, and the ML based training and prediction modules. **(B–E)** ROC and 3D scatter plots, and **(F)** the quantitative bar graph, showing the prediction performance to identify NPs in the plasma of mice using QML trained with 70 %, 50 %, 10 %, or 1 % data. **(G)** A schematic diagram showing the experimental plan for the collection of plasma from healthy individuals and CRC patients, and separately spiking NPs in a separate group of plasma from healthy individuals and CRC patients, followed by measurements of electrokinetic data, and the ML based training and prediction modules. **(H–K)** ROC and 3D scatter plots showing the QML based identification of NPs in plasma of (H) healthy and (J) cancer patients, and the corresponding bar graphs, showing the prediction performance of QML trained with 70 %, 50 %, 10 %, and 5 % data, to identify NPs in the plasma of healthy and cancer patients. (F, I, K) Bar graphs showed that QML could maintain high prediction performance, even with massive reductions in training set size. For precision estimation, the degree of uncertainty in sample was represented in ROC plots via CI (showed with confidence shape) with probability limit of 95 %.
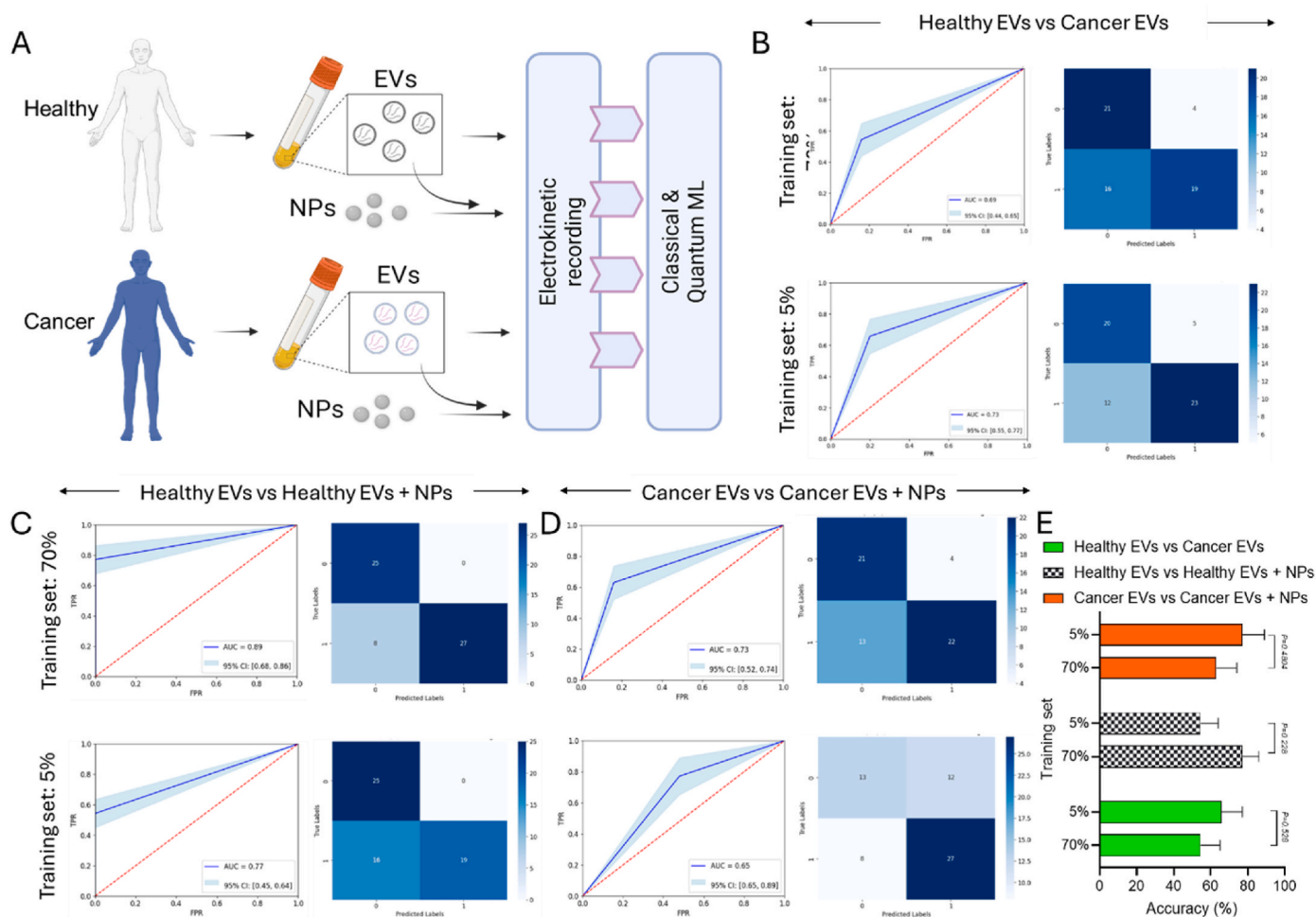
Fig. 6. QML based electrokinetic mining with fewer observations identifies EVs from the plasma of healthy and cancer patients, and EVs mixed wit**h NPs. (A)** A schematic diagram showing the experimental plan for the collection of plasma from healthy and cancer patients, followed by the isolation of EVs and their spiking with NPs. Then, electrokinetic recording was conducted on the plasma derived EVs without or with NPs. **(B–D)** ROC curves and confusion matrices showing the classification of (B) healthy EVs vs cancer EVs, (C) healthy EVs vs healthy EVs spiked with NPs, and (D) cancer EVs vs cancer EVs spiked with NPs by QML trained with 70 % and 5 % data. **(E)** Quantitative bar graphs showing the prediction performance by QML at training set of 70 % and 5 %, inferring that QML could sustain the percentage of prediction performance, even with massive decrease in training set. For precision estimation, the degree of uncertainty in sample was represented in ROC plots via CI (showed with confidence shape) with probability limit of 95 %. In Fig. 6E, the statistical significance was determined via multiple t-tests, followed by the Holm-Sidak method, with P value = 0.05.

both states at the same time. This allows QML algorithms to solve problems that are intractable for classical computers [39,92–95]. QML through VQC offers distinct advantages over traditional algorithms. Quantum systems excel in representing complex probability distributions, enhancing accuracy. Moreover, QML demonstrates superior generalization from limited datasets, a crucial benefit in data-scarce scenarios. Research by Caro et al. highlights QML's efficacy in learning from minimal training data, potentially tackling both scalability challenges and predictive power limitations, particularly when working with constrained datasets, positioning QML as a promising

frontier in computational learning [40,41]. We demonstrated the first use of QML alongside classical ML to identify NPs or EVs via electrokinetic mining. More specifically, green synthesis allied to DLS allowed characterization of 504 and 512 samples of Fe- and Co- NPs, respectively. According to train-test split and k-fold cross-validation techniques, VQC was found to predict NP types with accuracies of 94 %, 94 %, and 91 %, when respectively trained on datasets of 70 %, 50 %, and 10 % of the initial size. Notably, the QML showed the highest prediction performance compared to the classical ML models at the 1 % training set (Supplementary Figs. S12 and 13). Similarly, the electrokinetic data of

EVs were analyzed by the implementation of QML, and notably with mere 0.5 % training set QML could identify the type of EVs with prediction performance of 71 % (Supplementary Figs. S14A–K). These findings indicate that QML based electrokinetic mining shows promising potential for accurately identifying the cellular origins of EVs, a pivotal step in analyzing the diverse populations of EVs present in biofluids such as plasma and CSF. Our research underscores the remarkable potential of QML in discerning EVs originating from cancerous and non-cancerous cells. This discovery opens exciting avenues for leveraging this technology to differentiate EVs present in clinically relevant biofluids. Our current focus lies in investigating the feasibility of utilizing QML to distinguish between EVs extracted from the plasma of cancer patients and those from individuals without cancer. By doing so, we aim to enhance the diagnostic capabilities of EV analysis, paving the way for more precise and effective cancer detection methods. This endeavor has led us to harness EVs sourced from both healthy subjects and individuals diagnosed with cancer, marking a crucial step towards harnessing the full potential of QML in advancing cancer diagnostics.

In expanding the scope of our approach, we analyzed the electrokinetic properties of pre-clinical and clinical plasma samples. The clinical samples included plasma from 10 healthy subjects and 10 CRC patients, from which EVs were isolated. Notably, the QML based electrokinetic mining was found to be effective in distinguishing heathy EVs and cancer EVs (Fig. 6A–C, E). Additionally, to assess the potential of employing QML to distinguish NPs containing plasma derived EVs, NPs were spiked into the EV samples from the plasma of 10 CRC patients in a separate group. Further, we used pre-clinical samples to assess the potential of employing QML to distinguish EVs in plasma containing NPs, 10 mice received 100 μl of AgNP solution intravenously. Subsequently, EVs were isolated from their plasma, with EVs from the plasma of 10 untreated mice serving as the control group. Notably, the QML based electrokinetic mining could identify NPs with EVs from the plasma of CRC patients (Fig. 6A–D, E), and mice (Figs. S5A–F). Evidently, compared to the QML, the benchmarks for LR, KNN, SVM, DT) and XGB indicate that the performance of VQC profoundly outperforms classical ML models making it especially useful for predictive purposes with a small training set. Interestingly, the QML based analysis using only 0.5 % training set demonstrated that even a single feature would be enough to learn and predict, contributing towards the identification of NPs or EVs. This is a substantial finding that could lead to new applications in areas such as drug discovery and environmental remediation. Some of the challenges that need to be addressed before VQCs can be widely used are noise, scalability, and software. In the context of noise, quantum computers are susceptible to noise, which can cause errors in the computation. In terms of scalability, quantum computers are currently very small, and it is not yet clear how to scale them up to the size that would be needed for practical applications. When it comes to the software, there is a lack of software development tools for VQCs [96]. Therefore, these aspects may require future studies to make progress in the application of QML. Additionally, incorporating ML models through application programming interfaces (APIs) into Internet of Things (IoT) applications, especially by utilizing sophisticated QML algorithms, holds the potential to substantially elevate the precision in predicting cancer susceptibility, recurrence, and survival rates. This integration would serve as a crucial asset for the early diagnosis and prognosis evaluation of cancer, marking a substantial advancement in the medical diagnostics and patient care [97,98].

In future, the scope of ML algorithms in nano-biomedicine for designing targeted therapy for various diseases is exciting and varied. ML algorithms can be used to develop advanced nanomaterials that can recognize and bind specifically to diseased proteins and cellular components, for imaging and tracking of NPs and EVs *in vivo*, to analyze large datasets to identify biomarkers, to identify optimal dosing, and to predict the response of a patient to a particular drug regimen. Ultimately, ML algorithms can be used to help optimize targeted therapies based on individual patient data, with the goal of achieving improved outcomes and improved quality of care. Furthermore, with the tremendous advances in ML technology, there is an opportunity for researchers to develop more sophisticated algorithms that can predict the response of a patient to a particular therapeutic intervention and guide dosing regimen. This could be particularly useful for personalized medicine approaches.

4. Materials and methods

4.1. Chemicals

Cobalt nitrate hexahydrate ($Co(NO_3)_2 \cdot 6H_2O$) (Product no.: 1.02536), and ferrous iron (III) chloride hexahydrate ($FeCl_3 \cdot 6H_2O$) (Product no.: 1.03943) were obtained from Millipore Sigma. The chemicals were obtained at 99.9 % purity in analytical reagent (AR) grade.

4.2. Green synthesis of cobalt- and iron- NPs

The Curcuma longa plant's rhizome extract was obtained from Sigma-Aldrich (Product no.: NIST3300). For the green synthesis of NPs, 0.1 M Cobalt nitrate hexahydrate or Ferrous iron (III) chloride was used along with 10 mL plant extract of Curcuma longa. Both sources were mixed by using a magnetic stirrer at constant rpm (750 rpm) for 30 min. The color deviations from light brown to dark brown specify the nucleation and growth of Co_3O_4 NPs. The obtained color-modified NP solutions were centrifuged at 10,000 rpm for 10 min (repeated three times), and their obtained pellet was washed with double distilled water. Finally, the collected NP precipitate was filtered (with Whatman No. 1 filter paper) and dried in the oven at 100 °C for 60 min. The dried NP samples were processed for further characterization and electrokinetic analysis.

4.3. TEM analysis

The morphology of the Fe- or Co- NPs was visualized using TEM machine, at a 200 kV accelerating voltage and an ultra-high resolution of 0.2 nm and a magnification level of 2,000,000 x. A 3 mm diameter TEM grid was prepped by placing 5 μL of NP solutions onto carbon-coated copper grids. The grids were then dried with a mercury lamp before analysis. Image magnifying software was used to determine the size of the NPs to obtain morphological data.

4.4. Cell culture

All embryonic stem cell studies were approved by the Institutional Review Board (IRB) at the University of Chicago. All cells were purchased from ATCC or WiCell in the past 3 years and were negative for mycoplasma. The hESC lines were regularly checked for chromosome abnormalities and maintained with normal chromosome numbers.

hESC was cultured as described previously [61,99]. The hESC line-RUES2 or H1 was cultured on irradiated mouse embryonic fibroblasts (Global Stem) at a density of 20,000–25,000 cells/cm^2 in a medium of DMEM/F12, 20 % knockout serum replacement (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma Aldrich) and 20 ng/ml bFGF (R&D Systems), and medium was changed daily. hESC cultures were maintained in an undifferentiated state at 37 °C in a 5 % CO_2/air environment until stem cells reached about 90 % confluence.

Human lung cancer (A549)- and Colorectal adenocarcinoma (Caco-2)- cells were cultured in a medium of DMEM supplemented with 10 % FBS, and 1 % Penicillin-Streptomycin. Cell cultures were maintained in an undifferentiated state at 37 °C in a 5 % CO_2.

4.5. Isolation of EVs from cell cultured condition medium

EVs from the culture conditioned medium (CCM) of hESC, A549, or

Caco-2 were isolated using Total Exosome Isolation reagent (Invitrogen, catalog no. 4478359) as per the manufacturer's protocol. The CCM from cultured hESC or Caco-2 were first centrifuged at $2000 \times g$ for 30 min. Then, the resultant supernatant was mixed with 0.5 vol of a total exosome isolation reagent. The mixture was incubated at 4 °C overnight, followed by its centrifugation at $10,000 \times g$ for 1 h at 4 °C. The resultant supernatant was aspirated and discarded, and the remaining exosome pellet was diluted with $1 \times$ PBS for NTA analysis.

4.6. Clinical samples of plasma

Plasma samples were obtained from ten healthy subjects and ten colorectal cancer patients (background information is listed in Supplementary Table S2), and patients consented under IRB-10-209-A approved under the University of Chicago Institutional Review Board. For the group consisting of NPs with EVs, the plasma-derived EV samples were spiked with NPs (at the dilution of 1:100).

4.7. Pre-clinical samples of plasma

All procedures in mice were in accordance with the animal protocol approved by the Institutional Animal Care and Use Committee at the University of Chicago, Illinois, USA.

Blood samples were withdrawn from control and experimental group of C57BL/6 mice (10 mice/group) as described previously [100,101]. In experimental group of mice, 100 μl volume of AgNPs suspension (Sigma-Aldrich, Cat# 730785) in PBS (stock concentration = 0.5 mg/ml) was injected via tail vein for 2 consecutive days.

4.8. Isolation of EVs from plasma

EVs from plasma were isolated using Total Exosome Isolation kit (from plasma) (Invitrogen, catalog no. 4484450) as per the manufacturer's protocol. The reagent works by sequestering water molecules, which causes less soluble components like vesicles to precipitate out of the solution. These can then be collected through a brief, low-speed centrifugation process. Initially, the plasma was treated with Proteinase K for 10 min. Next, the reagent was added, and the mixture was incubated for half an hour at 4 °C, followed by a standard centrifugation at room temperature ($10,000 \times g$ for 5 min), to recover the precipitated exosomes. Finally, the resulting pellet is resuspended in PBS for downstream analysis.

4.9. NTA analysis

The number and size distribution of EVs were characterized using NTA with a Malvern NanoSight NS300 instrument. A monochromatic laser beam at 405 nm was applied to 500 μL of EV solutions loaded into the sample chamber. Three video captures of exosome movements were recorded over a 30-s period and analyzed using NTA software. The software was optimized to identify and track exosomes on a frame-by-frame basis. The number of EVs released from hESC or Caco-2 cells was calculated using NTA analysis.

4.10. Electrokinetic analysis by DLS

The colloidal suspension of Fe- or Co- NPs was (10 times) diluted with sterile water, followed by homogenization with an ultrasonic homogenizer for 15 min at 20 Hz to break up aggregations of NPs. The solution of NP was passed through a Millipore filter (0.22 μm) and analyzed through Malvern Zetasizer Nano. EVs were analyzed by Malvern Zetasizer Nano using equally diluted samples prepared with equal amount (50 μg/ml) within PBS for each group. Similarly, the electrokinetic analysis of EVs plasma from control- and NP injected- C57BL/6 mice (10 mice/group), and EVs from the plasma of 10 healthy subjects, 10 colorectal cancer patients (without and with NPs spiking), was

accomplished via Malvern Zetasizer Nano.

4.11. Data collection

A large dataset of electrokinetic features (ZP, Mob, Cond, and MCR) of two different green synthesized Fe NPs (N = 504) and Co NPs (N = 512), two different types of EVs; hESC-derived EVs (N = 500), Caco-2-derived EVs (N = 500), and DLD-1-derived EVs (N = 500), clinical samples: Control- and NPs spiked-plasma samples from 10 healthy subjects (N = 200), 10 colorectal cancer patients (N = 200). EVs from the plasma of 10 healthy subjects (N = 100), 10 colorectal cancer patients (N = 100), and NPs along with EVs from the plasma of 10 colorectal cancer patients (N = 100), and pre-clinical samples: EVs from the plasma of control C57BL/6 (N = 100), and NP injected C57BL/6 mice, were characterized via DLS. The data was subsequently used for the downstream ML analysis.

4.12. Classical machine learning analysis

The NP dataset included 1016 data points in two types (Fe and Co) with four given parameters, namely ZP, Mob, Cond, and MRC. The EV dataset included 1533 data points in three types (hESC-, A549-or Caco-2- EVs). Among them, only ZP, Cond, and MRC were used due to the linear correlation between ZP and Mob. The structure of the dataset was first analyzed using scatter plot analysis and principal component analysis (PCA). Then we used a train-test split and k-fold cross validation to evaluate the performance of the models. For the train-test split, the dataset was shuffled, then divided into two sets with sizes of 70 % and 30 % for training and testing, respectively. For the k-fold cross validation, we used 10-, 100, and 200- folds. In this study, we used LR, SVM, KNN, DT, and XGB to classify the NP's and EVs' types based on three parameters: ZP, Mob, and MCR. In this work, all the ML models were first trained using a training set and then tested on the test set. The performance of these models was evaluated using various metrics such as cross validation accuracy, confusion matrix, precision, recall, F1-score, and AUC-ROC.

Optimization: The parameters of the classical ML models were optimized via hyperparameter tuning using GridSearchCV.

4.13. Quantum machine learning analysis via variational quantum classifier

The VQC is a key supervised QML algorithm (Supplementary Fig. S9) that is widely used for classification problems in noisy intermediate-scale quantum (NISQ) devices [93–95]. It allows us to obtain exploratory results on NISQ devices without the need for additional error-correction techniques. The cost function is calculated using iterative device measurements, which helps to mitigate errors by incorporating noisy data into the optimization computations. This quantum method uses the mapping of classical input data to an increasingly large quantum feature space, which is based on quantum circuits that are difficult to replicate classically.

The QML model was implemented in three steps: state preparation, model circuit, and measurement. VQC started with the initial state preparation of data. Then, a quantum feature map, which is a mathematical mapping that embedded the data into higher dimensional spaces was applied to convert classical data into quantum states. The variational circuit was then constructed, and its parameters were optimized using a classical optimizer. The number of measurements and dimensions determined the size of the variational circuit. The measured value was transmitted back to the circuit as feedback, which was used to improve the variational circuit's parameters. Finally, the process was repeated until the desired prediction performance was achieved.

Feature Map: A feature map was used as the first step. This is a type of mathematical function that maps data from one space to another, and in the context of QML, it was used to map classical data into quantum

states. This allowed quantum computers to process and learn from classical data in a new way. A quantum feature map was used, which is a function that maps a classical data point x_i to a quantum state $|\phi(x_i)\rangle$. The quantum state $|\phi(x_i)\rangle$ is a superposition of states that represent the different possible values of the classical data point x_i. Once the classical data was mapped to quantum states, the quantum computer was used to process and learn from the data, such as using a classifier to find a hyperplane that separated the data into two classes. In this research paper, a Pauli feature map with a linear entanglement strategy was used, which is depicted in Supplementary Fig. S19 as a quantum circuit for a Pauli feature map with two data encoding repetitions over four features.

Model Circuit: The second step is the model circuit, or the classifier strictly speaking. A parameterized unitary operator U(w) is created such that $|\psi(x:\theta)\rangle = U(w)|\psi(x)\rangle$ The model circuit is constructed from gates that evolve the input state. The circuit is based on unitary operations and depends on external parameters which will be adjustable. Given a prepared state $|\psi_i\rangle$ the model circuit, U(w) maps $|\psi_i\rangle$ to another vector $|\psi_i\rangle = U(w)|\psi_i\rangle$. In turn U(w) consists of a series of unitary gates. The weights of the model circuit can be optimized using a classical optimization algorithm. The optimization algorithm tries different values for the weights and measures the performance of the model for each value. The value of the weights that produce the best performance is then used for the final model. We used the State vector simulator to test the circuit on 32 qubits, which can be represented as.

$$|\psi(x:\theta)i = U(\theta)|\varphi(x)i$$

Supplementary Fig. S20 depicts the variational circuit used in this work.

Measurement: This is the final step in a QML algorithm. It estimates the probability of a data point belonging to a particular class by measuring the quantum state of the data point. This is equivalent to sampling multiple times from the set of all possible values that the data point could take on and averaging the results.

Optimization: The parameters of the VQC are updated using an optimization routine. This is done in a classical loop that trains the parameters until the cost function's value decreases. Here, we used gradient-free stochastic optimization algorithm including Constrained Optimization by Linear Approximation (COBYLA), Powell's Method or Powell's Conjugate Direction Method (POWELL), Simultaneous Perturbation Stochastic Approximation (SPSA), or Univariate Marginal Distribution Algorithm (UMDA) optimizer. By varying the range of maximum iterations of these optimizers, we ensured the maximum prediction performance of the given model.

The data was prepared for training and testing by encoding the data into qubits and by shuffling and splitting it into two sets: a training set and a testing set. Next, the feature map was defined, which is a circuit that maps the input data to a set of qubits that can be used by the VQC algorithm. Next, the ansatz was defined, which is a circuit that is used to approximate the target function. Then, the VQC algorithm was trained by iteratively varying the parameters of the ansatz and measuring the prediction performance of the classifier on the training set (70 % initial training set, 50 % of initial, and 10 % of initial). Eventually, the VQC algorithm was tested to evaluate the performance of the classifier via confusion matrix and ROC analysis.

Similarly, data points of three types of EVs (hESC-, A549-, or Caco-2-EVs) with their given parameters, namely ZP, Cond, and MRC were used for the implementation of VQC. In addition, the same strategy was employed for mouse- and human-plasma or EVs, with or without NPs. Augmentation techniques were employed for distinguishing cancer plasma vs cancer plasma with NPs, followed by training the QML model.

4.14. SHAP value analysis

SHAP value analysis was conducted to explain the output of ML based electrokinetic mining for the identification of NPs or EVs. SHAP values were calculated by measuring the contribution of each attribute to the prediction ability of ML model.

4.15. Statistical analysis

All statistical tests, no adjustments were made for comparison. For all parametric statistical analyses, data were determined to be normally distributed. In Fig. 2E–G, the data are presented as mean $\pm$ standard error mean (S.E.M.) of 3 data points. Student's t-test was applied for statistical analysis with significance level at *P $<$ 0.05, **P $<$ 0.01. The degree of uncertainty in sample was represented in ROC plots via CI (showed with confidence shape) with probability limit of 95 %. In Fig. 6E, the statistical significance was determined via multiple t-tests, followed by the Holm-Sidak method, with P value = 0.05.

CRediT authorship contribution statement

Abhimanyu Thakur: Writing – original draft, Visualization, Project administration, Methodology, Conceptualization. **Pedro Correia Santos Bezerra:** Writing – review & editing, Visualization, Methodology. **Abhishek:** Visualization, Methodology. **Shihao Zeng:** Visualization, Methodology. **Kui Zhang:** Investigation. **Werner Treptow:** Investigation. **Alexander Luna:** Methodology. **Urszula Dougherty:** Methodology. **Akushika Kwesi:** Methodology. **Isabella R. Huang:** Resources. **Christine Bestvina:** Writing – review & editing. **Marina Chiara Garassino:** Writing – review & editing. **Fuyu Duan:** Visualization. **Yash Gokhale:** Writing – review & editing, FNU. **Bin Duan:** Writing – review & editing. **Yin Chen:** Writing – review & editing. **Qizhou Lian:** Resources. **Marc Bissonnette:** Methodology. **Jianpan Huang:** Writing – review & editing, Visualization, Supervision, Methodology. **Huanhuan Joyce Chen:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Data and materials availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper can be requested from the corresponding authors. The code used to generate the results in this paper is available at https://github.com/Abhimanyu2023/nanoQML.

Ethics approval and consent to participate

All embryonic stem cell studies were approved by the Institutional Review Board (IRB) at the University of Chicago. Plasma samples were obtained from ten healthy subjects and ten colorectal cancer patients, and patients consented under IRB-10-209-A approved under the University of Chicago Institutional Review Board.

Funding

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

the UChicago, IL, USA).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi. org/10.1016/j.bioactmat.2025.03.023.

References

[1] J.K. Patra, G. Das, L.F. Fraceto, E.V.R. Campos, M. del P. Rodriguez-Torres, L. S. Acosta-Torres, L.A. Diaz-Torres, R. Grillo, M.K. Swamy, S. Sharma, S. Habtemariam, H.-S. Shin, Nano based drug delivery systems: recent developments and future prospects, J. Nanobiotechnol. 16 (2018) 71, https://doi. org/10.1186/s12951-018-0392-8.

[2] M. Fathi-Achachelouei, H. Knopf-Marques, C.E. Ribeiro da Silva, J. Barthès, E. Bat, A. Tezcaner, N.E. Vrana, Use of nanoparticles in tissue engineering and regenerative medicine, Front. Bioeng. Biotechnol. 7 (2019), https://doi.org/ 10.3389/fbioe.2019.00113.

[3] H. Malekzad, P. Sahandi Zangabad, H. Mirshekari, M. Karimi, M.R. Hamblin, Noble metal nanoparticles in biosensors: recent studies and applications, Nanotechnol. Rev. 6 (2017), https://doi.org/10.1515/ntrev-2016-0014.

[4] I. Khan, K. Saeed, I. Khan, Nanoparticles: properties, applications and toxicities, Arab. J. Chem. (2017), https://doi.org/10.1016/j.arabjc.2017.05.011.

[5] A. Thakur, X. Ke, Y.-W. Chen, P. Motallebnejad, K. Zhang, Q. Lian, H.J. Chen, The mini player with diverse functions: extracellular vesicles in cell biology, disease, and therapeutics, Protein Cell 13 (2022) 631–654, https://doi.org/10.1007/ s13238-021-00863-6.

[6] A. Thakur, G. Qiu, C. Xu, X. Han, T. Yang, S.P. Ng, K.W.Y. Chan, C.M.L. Wu, Y. Lee, Label-free sensing of exosomal MCT1 and CD147 for tracking metabolic reprogramming and malignant progression in glioma, Sci. Adv. 6 (2020), https:// doi.org/10.1126/sciadv.aaz6119.

[7] A. Thakur, D.C. Parra, P. Motallebnejad, M. Brocchi, H.J. Chen, Exosomes: small vesicles with big roles in cancer, vaccine development, and therapeutics, Bioact. Mater. 10 (2022) 281–294, https://doi.org/10.1016/j.bioactmat.2021.08.029.

[8] L. Gong, Y. Weng, W. Zhou, K. Zhang, W. Li, J. Jiang, J. Zhu, In vivo CT imaging of gold nanoparticle-labeled exosomes in a myocardial infarction mouse model, Ann. Transl. Med. 9 (2021) 504, https://doi.org/10.21037/atm-21-981.

[9] O. Betzer, N. Perets, A. Angel, M. Motiei, T. Sadan, G. Yadid, D. Offen, R. Popovtzer, In vivo neuroimaging of exosomes using gold nanoparticles, ACS Nano 11 (2017) 10883–10893, https://doi.org/10.1021/acsnano.7b04495.

[10] X. Fang, Y. Wang, S. Wang, B. Liu, Nanomaterials assisted exosomes isolation and analysis towards liquid biopsy, Mater, Today Bio 16 (2022) 100371, https://doi. org/10.1016/j.mtbio.2022.100371.

[11] N. Sizochenko, J. Leszczynski, Review of current and emerging approaches for quantitative nanostructure-activity relationship modeling, J. Nanotoxicology Nanomedicine. 1 (2016) 1–16, https://doi.org/10.4018/JNN.2016010101.

[12] T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T.P. Dasari, A. Michalkova, H.- M. Hwang, A. Toropov, D. Leszczynska, J. Leszczynski, Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles, Nat. Nanotechnol. 6 (2011) 175–178, https://doi.org/10.1038/nnano.2011.10.

[13] N. Sizochenko, B. Rasulev, A. Gajewicz, V. Kuz'min, T. Puzyn, J. Leszczynski, From basic physics to mechanisms of toxicity: the "liquid drop" approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles, Nanoscale 6 (2014) 13986–13993, https://doi.org/10.1039/ C4NR03487B.

[14] R. Liu, R. Rallo, S. George, Z. Ji, S. Nair, A.E. Nel, Y. Cohen, Classification NanoSAR development for cytotoxicity of metal oxide nanoparticles, Small 7 (2011) 1118–1126, https://doi.org/10.1002/smll.201002366.

[15] R. Tantra, C. Oksel, T. Puzyn, J. Wang, K.N. Robinson, X.Z. Wang, C.Y. Ma, T. Wilkins, Nano(Q)SAR: challenges, pitfalls and perspectives, Nanotoxicology 9 (2015) 636–642, https://doi.org/10.3109/17435390.2014.952698.

[16] S. Mourdikoudis, R.M. Pallares, N.T.K. Thanh, Characterization techniques for nanoparticles: comparison and complementarity upon studying nanoparticle properties, Nanoscale 10 (2018) 12871–12934, https://doi.org/10.1039/ C8NR02278J.

[17] K. Li, B. DeCost, K. Choudhary, M. Greenwood, J. Hattrick-Simpers, A critical examination of robustness and generalizability of machine learning prediction of materials properties, npj Comput. Mater. 9 (2023) 55, https://doi.org/10.1038/ s41524-023-01012-9.

[18] Handbook of Nanotechnology Applications, Elsevier, 2021, https://doi.org/ 10.1016/C2019-0-01115-6.

[19] D.F. Quevedo, C.J. Lentz, A. Coll de Peña, Y. Hernandez, N. Habibi, R. Miki, J. Lahann, B.H. Lapizco-Encinas, Electrokinetic characterization of synthetic protein nanoparticles, Beilstein J. Nanotechnol. 11 (2020) 1556–1567, https:// doi.org/10.3762/bjnano.11.138.

[20] G. Midekessa, K. Godakumara, J. Ord, J. Viil, F. Lättekivi, K. Dissanayake, S. Kopanchuk, A. Rinken, A. Andronowska, S. Bhattacharjee, T. Rinken, A. Fazeli, Zeta potential of extracellular vesicles: toward understanding the attributes that determine colloidal stability, ACS Omega 5 (2020) 16701–16710, https://doi. org/10.1021/acsomega.0c01582.

[21] I. Gaurav, T. Singh, A. Thakur, G. Kumar, P. Rathee, P. Kumari, K. Sweta, Synthesis, in-vitro and in-silico evaluation of silver nanoparticles with root extract of withania somnifera for antibacterial activity via binding of penicillin-

[22] binding protein-4, Curr. Pharm. Biotechnol. 21 (2020) 1674–1687, https://doi. org/10.2174/1389201021666200702152000.

[22] A. Thakur, Nano therapeutic approaches to combat progression of metastatic prostate cancer, Adv. Cancer Biol. - Metastasis. 2 (2021) 100009, https://doi.org/ 10.1016/j.adcanc.2021.100009.

[23] M. Tkach, C. Théry, Communication by extracellular vesicles: where we are and where we need to go, Cell 164 (2016) 1226–1232, https://doi.org/10.1016/j. cell.2016.01.043.

[24] I. Parolini, C. Federici, C. Raggi, L. Lugini, S. Palleschi, A. De Milito, C. Coscia, E. Iessi, M. Logozzi, A. Molinari, M. Colone, M. Tatti, M. Sargiacomo, S. Fais, Microenvironmental pH is a key factor for exosome traffic in tumor cells, J. Biol. Chem. 284 (2009) 34211–34222, https://doi.org/10.1074/jbc.M109.041152.

[25] J. Skog, T. Würdinger, S. van Rijn, D.H. Meijer, L. Gainche, W.T. Curry, B. S. Carter, A.M. Krichevsky, X.O. Breakefield, Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers, Nat. Cell Biol. 10 (2008) 1470–1476, https://doi.org/10.1038/ ncb1800.

[26] A. Thakur, A.P. Mishra, B. Panda, D.C.S. Rodríguez, I. Gaurav, B. Majhi, Application of artificial intelligence in pharmaceutical and biomedical studies, Curr. Pharm. Des. 26 (2020) 3569–3578, https://doi.org/10.2174/ 1381612826666200515131245.

[27] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, S. Zhao, Applications of machine learning in drug discovery and development, Nat. Rev. Drug Discov. 18 (2019) 463–477, https://doi.org/10.1038/s41573-019-0024-5.

[28] A. Thakur, A.P. Mishra, B. Panda, K. Sweta, B. Majhi, Detection of disease-specific parent cells via distinct population of nano-vesicles by machine learning, Curr. Pharm. Des. 26 (2020) 3985–3996, https://doi.org/10.2174/ 1381612826666200422091753.

[29] L. Chen, M. Schär, K.W.Y. Chan, J. Huang, Z. Wei, H. Lu, Q. Qin, R.G. Weiss, P.C. M. van Zijl, J. Xu, In vivo imaging of phosphocreatine with artificial neural networks, Nat. Commun. 11 (2020) 1072, https://doi.org/10.1038/s41467-020- 14874-0.

[30] J. Huang, J.H.C. Lai, K.-H. Tse, G.W.Y. Cheng, Y. Liu, Z. Chen, X. Han, L. Chen, J. Xu, K.W.Y. Chan, Deep neural network based CEST and AREX processing: application in imaging a model of Alzheimer's disease at 3 T, Magn. Reson. Med. 87 (2022) 1529–1545, https://doi.org/10.1002/mrm.29044.

[31] X. Huang, B. Liu, S. Guo, W. Guo, K. Liao, G. Hu, W. Shi, M. Kuss, M.J. Duryee, D. R. Anderson, Y. Lu, B. Duan, SERS spectroscopy with machine learning to analyze human plasma derived sEVs for coronary artery disease diagnosis and prognosis, Bioeng. Transl. Med. 8 (2023), https://doi.org/10.1002/btm2.10420.

[32] H. Tao, T. Wu, M. Aldeghi, T.C. Wu, A. Aspuru-Guzik, E. Kumacheva, Nanoparticle synthesis assisted by machine learning, Nat. Rev. Mater. 6 (2021) 701–716, https://doi.org/10.1038/s41578-021-00337-5.

[33] H. Lv, X. Chen, Intelligent control of nanoparticle synthesis through machine learning, Nanoscale 14 (2022) 6688–6708, https://doi.org/10.1039/ D2NR00124A.

[34] L. Yao, Z. Ou, B. Luo, C. Xu, Q. Chen, Machine learning to reveal nanoparticle dynamics from liquid-phase TEM videos, ACS Cent. Sci. 6 (2020) 1421–1430, https://doi.org/10.1021/acscentsci.0c00430.

[35] Z. Lin, W.-C. Chou, Y.-H. Cheng, C. He, N.A. Monteiro-Riviere, J.E. Riviere, Predicting nanoparticle delivery to tumors using machine learning and artificial intelligence approaches, Int. J. Nanomed. 17 (2022) 1365–1379, https://doi.org/ 10.2147/IJN.S344208.

[36] H. Shin, B.H. Choi, O. Shim, J. Kim, Y. Park, S.K. Cho, H.K. Kim, Y. Choi, Single test-based diagnosis of multiple cancer types using Exosome-SERS-AI for early stage cancers, Nat. Commun. 14 (2023) 1644, https://doi.org/10.1038/s41467- 023-37403-1.

[37] H. Shin, S. Oh, S. Hong, M. Kang, D. Kang, Y. Ji, B.H. Choi, K.-W. Kang, H. Jeong, Y. Park, S. Hong, H.K. Kim, Y. Choi, Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes, ACS Nano 14 (2020) 5435–5444, https://doi.org/10.1021/acsnano.9b09119.

[38] E.K. Ramos, C.-F. Tsai, Y. Jia, Y. Cao, M. Manu, R. Taftaf, A.D. Hoffmann, L. El- Shennawy, M.A. Gritsenko, V. Adorno-Cruz, E.J. Schuster, D. Scholten, D. Patel, X. Liu, P. Patel, B. Wray, Y. Zhang, S. Zhang, R.J. Moore, J. V Mathews, M. J. Schipma, T. Liu, V.L. Tokars, M. Cristofanilli, T. Shi, Y. Shen, N.K. Dashzeveg, H. Liu, Machine learning-assisted elucidation of CD81–CD44 interactions in promoting cancer stemness and extracellular vesicle integrity, Elife 11 (2022), https://doi.org/10.7554/eLife.82669.

[39] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, J. R. McClean, Power of data in quantum machine learning, Nat. Commun. 12 (2021) 2631, https://doi.org/10.1038/s41467-021-22539-9.

[40] M.C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, P. J. Coles, Generalization in quantum machine learning from few training data, Nat. Commun. 13 (2022) 4919, https://doi.org/10.1038/s41467-022-32550-3.

[41] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, J.R. McClean, Quantum advantage in learning from experiments, Science 376 (2022) 1182–1186, https://doi.org/10.1126/ science.abn7293.

[42] C.-M.J. Hu, R.H. Fang, B.T. Luk, L. Zhang, Nanoparticle-detained toxins for safe and effective vaccination, Nat. Nanotechnol. 8 (2013) 933–938, https://doi.org/ 10.1038/nnano.2013.254.

[43] Y. Zhao, X. Zhao, Y. Cheng, X. Guo, W. Yuan, Iron oxide nanoparticles-based vaccine delivery for cancer treatment, Mol. Pharm. 15 (2018) 1791–1799, https://doi.org/10.1021/acs.molpharmaceut.7b01103.

[44] S.-Y. Liu, W. Wei, H. Yue, D.-Z. Ni, Z.-G. Yue, S. Wang, Q. Fu, Y.-Q. Wang, G.-H. Ma, Z.-G. Su, Nanoparticles-based multi-adjuvant whole cell tumor vaccine for cancer immunotherapy, Biomaterials 34 (2013) 8291–8300, https://doi.org/10.1016/j.biomaterials.2013.07.020.

[45] K. Luyts, D. Napierska, B. Nemery, P.H.M. Hoet, How physico-chemical characteristics of nanoparticles cause their toxicity: complex and unresolved interrelations, Environ. Sci. Process. Impacts. 15 (2013) 23–38, https://doi.org/10.1039/C2EM30237C.

[46] J.P.M. Almeida, A.L. Chen, A. Foster, R. Drezek, In vivo biodistribution of nanoparticles, Nanomedicine. 6 (2011) 815–835, https://doi.org/10.2217/nnm.11.79.

[47] E.A. Grego, A.C. Siddoway, M. Uz, L. Liu, J.C. Christiansen, K.A. Ross, S.M. Kelly, S.K. Mallapragada, M.J. Wannemuehler, B. Narasimhan, in: Polymeric Nanoparticle-Based Vaccine Adjuvants and Delivery Vehicles, 2020, pp. 29–76, https://doi.org/10.1007/82_2020_226.

[48] S. Lupínková, M. Benkocká, P. Ryšánek, Z. Kolská, Enhancing immobilization of iron oxide particles on various polymer surfaces, Polym. Eng. Sci. 62 (2022) 1463–1472, https://doi.org/10.1002/pen.25935.

[49] J.L. Arias, V. Gallardo, M.A. Ruiz, in: Multifunctional Anticancer Nanomedicine Based on a Magnetically Responsive Cyanoacrylate Polymer, 2012, pp. 61–88, https://doi.org/10.1016/B978-0-12-391860-4.00004-5.

[50] X. Li, P. Zhou, Z. Luo, R. Feng, L. Wang, Hohenbuehelia serotina polysaccharides self-assembled nanoparticles for delivery of quercetin and their anti-proliferative activities during gastrointestinal digestion in vitro, Int. J. Biol. Macromol. 203 (2022) 244–255, https://doi.org/10.1016/j.ijbiomac.2022.01.143.

[51] X. Cheng, X. Wang, Z. Cao, W. Yao, J. Wang, R. Tang, Folic acid-modified soy protein nanoparticles for enhanced targeting and inhibitory, Mater. Sci. Eng. C. 71 (2017) 298–307, https://doi.org/10.1016/j.msec.2016.10.018.

[52] C. Cano-Sarmiento, D.I. Téllez-Medina, R. Viveros-Contreras, M. Cornejo-Mazón, C.Y. Figueroa-Hernández, E. García-Armenta, L. Alamilla-Beltrán, H.S. García, G. F. Gutiérrez-López, Zeta potential of food matrices, Food Eng. Rev. 10 (2018) 113–138, https://doi.org/10.1007/s12393-018-9176-z.

[53] B. De Berardis, M. Marchetti, A. Risuglia, F. Ietto, C. Fanizza, F. Superti, Exposure to airborne gold nanoparticles: a review of current toxicological data on the respiratory tract, J. Nanoparticle Res. 22 (2020) 235, https://doi.org/10.1007/s11051-020-04966-9.

[54] R.C. Flagan, Differential Mobility Analysis of Aerosols: A Tutorial, KONA Powder Part. J., vol. 26, 2008, pp. 254–268, https://doi.org/10.14356/kona.2008023.

[55] A.V. Delgado, F. González-Caballero, R.J. Hunter, L.K. Koopal, J. Lyklema, Measurement and interpretation of electrokinetic phenomena, J. Colloid Interface Sci. 309 (2007) 194–224, https://doi.org/10.1016/j.jcis.2006.12.075.

[56] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Int. Jt. Conf. Articial Intell., 1995, in: https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf.

[57] E. Alpaydın, Introduction to Machine Learning, The MIT Press, 2004. https://mit press.mit.edu/9780262012119/introduction-to-machine-learning/.

[58] Y. Xu, R. Goodacre, On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning, J. Anal. Test. 2 (2018) 249–262, https://doi.org/10.1007/s41664-018-0068-2.

[59] A. Iyaswamy, A. Thakur, X.-J. Guan, S. Krishnamoorthi, T.Y. Fung, K. Lu, I. Gaurav, Z. Yang, C.-F. Su, K.-F. Lau, K. Zhang, R.C.-L. Ng, Q. Lian, K.-H. Cheung, K. Ye, H.J. Chen, M. Li, Fe65-engineered neuronal exosomes encapsulating corynoxine-B ameliorate cognition and pathology of Alzheimer's disease, Signal Transduct. Targeted Ther. 8 (2023) 404, https://doi.org/10.1038/s41392-023-01657-4.

[60] H.J. Chen, E.E. Gardner, Y. Shah, K. Zhang, A. Thakur, C. Zhang, O. Elemento, H. Varmus, Formation of malignant, metastatic small cell lung cancers through overproduction of cMYC protein in TP53 and RB1 depleted pulmonary neuroendocrine cells derived from human embryonic stem cells, Elife (2024), https://doi.org/10.7554/eLife.93170.1.

[61] Q. Lian, K. Zhang, Z. Zhang, F. Duan, L. Guo, W. Luo, B.W.-Y. Mok, A. Thakur, X. Ke, P. Motallebnejad, V. Nicolaescu, J. Chen, C.Y. Ma, X. Zhou, S. Han, T. Han, W. Zhang, A.Y. Tan, T. Zhang, X. Wang, D. Xu, J. Xiang, A. Xu, C. Liao, F.-P. Huang, Y.-W. Chen, J. Na, G. Randall, H. Tse, Z. Chen, Y. Chen, H.J. Chen, Differential effects of macrophage subtypes on SARS-CoV-2 infection in a human pluripotent stem cell-derived model, Nat. Commun. 13 (2022) 2028, https://doi.org/10.1038/s41467-022-29731-5.

[62] R.C. Smith, V. Tabar, Constructing and deconstructing cancers using human pluripotent stem cells and organoids, Cell Stem Cell 24 (2019) 12–24, https://doi.org/10.1016/j.stem.2018.11.012.

[63] M. Zhang, J.J. Vandana, L. Lacko, S. Chen, Modeling cancer progression using human pluripotent stem cell-derived cells and organoids, Stem Cell Res. 49 (2020) 102063, https://doi.org/10.1016/j.scr.2020.102063.

[64] J. Zhu, S.K. Ozdemir, Y.-F. Xiao, L. Li, L. He, D.-R. Chen, L. Yang, On-chip single nanoparticle detection and sizing by mode splitting in an ultrahigh-Q microresonator, Nat. Photonics 4 (2010) 46–49, https://doi.org/10.1038/nphoton.2009.237.

[65] J.A. Welsh, D.C.I. Goberdhan, L. O'Driscoll, E.I. Buzas, C. Blenkiron, B. Bussolati, H. Cai, D. Di Vizio, T.A.P. Driedonks, U. Erdbrügger, J.M. Falcon-Perez, Q. Fu, A. F. Hill, M. Lenassi, S.K. Lim, M.G. Mahoney, S. Mohanty, A. Möller, R. Nieuwland, T. Ochiya, S. Sahoo, A.C. Torrecilhas, L. Zheng, A. Zijlstra, S. Abuelreich, R. Bagabas, P. Bergese, E.M. Bridges, M. Brucale, D. Burger, R.P. Carney, E. Cocucci, F. Colombo, R. Crescitelli, E. Hanser, A.L. Harris, N.J. Haughey, A. Hendrix, A.R. Ivanov, T. Jovanovic-Talisman, N.A. Kruh-Garcia, V. Ku'ulei-Lyn Faustino, D. Kyburz, C. Lässer, K.M. Lennon, J. Lötvall, A.L. Maddox, E.

S. Martens-Uzunova, R.R. Mizenko, L.A. Newman, A. Ridolfi, E. Rohde, T. Rojalin, A. Rowland, A. Saftics, U.S. Sandau, J.A. Saugstad, F. Shekari, S. Swift, D. Ter-Ovanesyan, J.P. Tosar, Z. Useckaite, F. Valle, Z. Varga, E. van der Pol, M.J.C. van Herwijnen, M.H.M. Wauben, A.M. Wehman, S. Williams, A. Zendrini, A. J. Zimmerman, M.I.S.E.V. Consortium, C. Théry, K.W. Witwer, Minimal information for studies of extracellular vesicles (MISEV2023): from basic to advanced approaches, J. Extracell. Vesicles 13 (2024), https://doi.org/10.1002/jev2.12404.

[66] H. Liu, S. Liu, Y. Xiao, W. Song, H. Li, L.W.C. Ho, Z. Shen, C.H.J. Choi, A pH-reversible fluorescent probe for in situ imaging of extracellular vesicles and their secretion from living cells, Nano Lett. 21 (2021) 9224–9232, https://doi.org/10.1021/acs.nanolett.1c03110.

[67] H. Mendivil-Alvarado, A.T. Limon-Miro, E. Carvajal-Millan, J. Lizardi-Mendoza, A. Mercado-Lara, C.D. Coronado-Alvarado, M.L. Rascón-Durán, I. Anduro-Corona, D. Talamás-Lara, A. Rascón-Careaga, H. Astiazarán-García, Extracellular vesicles and their zeta potential as future markers associated with nutrition and molecular biomarkers in breast cancer, Int. J. Mol. Sci. 24 (2023) 6810, https://doi.org/10.3390/ijms24076810.

[68] T. Akagi, T. Ichiki, Evaluation of zeta-potential of individual exosomes secreted from biological cells using a microcapillary electrophoresis chip, in: Encycl. Biocolloid Biointerface Sci. 2V Set, Wiley, 2016, pp. 469–473, https://doi.org/10.1002/9781119075691.ch37.

[69] E. van der Pol, L. de Rond, F.A.W. Coumans, E.L. Gool, A.N. Böing, A. Sturk, R. Nieuwland, T.G. van Leeuwen, Absolute sizing and label-free identification of extracellular vesicles by flow cytometry, Nanomed. Nanotechnol. Biol. Med. 14 (2018) 801–810, https://doi.org/10.1016/j.nano.2017.12.012.

[70] 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, Machine-learning-assisted materials discovery using failed experiments, Nature 533 (2016) 73–76, https://doi.org/10.1038/nature17439.

[71] L. Shang, Q. Xie, C. Yang, L. Kong, Z. Zhang, Extracellular vesicles facilitate the transportation of nanoparticles within and between cells for enhanced tumor therapy, ACS Appl. Mater. Interfaces 15 (2023) 42378–42394, https://doi.org/10.1021/acsami.3c10237.

[72] J.A. Peruzzi, T.F. Gunnels, H.I. Edelstein, P. Lu, D. Baker, J.N. Leonard, N. P. Kamat, Enhancing extracellular vesicle cargo loading and functional delivery by engineering protein-lipid interactions, Nat. Commun. 15 (2024) 5618, https://doi.org/10.1038/s41467-024-49678-z.

[73] L.W.C. Ho, C.K.W. Chan, R. Han, Y.F.Y. Lau, H. Li, Y.-P. Ho, X. Zhuang, C.H. J. Choi, Mammalian cells exocytose alkylated gold nanoparticles via extracellular vesicles, ACS Nano 16 (2022) 2032–2045, https://doi.org/10.1021/acsnano.1c07418.

[74] I. Gaurav, A. Thakur, G. Kumar, Q. Long, K. Zhang, R.K. Sidu, S. Thakur, R. K. Sarkar, A. Kumar, A. Iyaswamy, Z. Yang, Delivery of apoplastic extracellular vesicles encapsulating green-synthesized silver nanoparticles to treat citrus canker, Nanomaterials 13 (2023) 1306, https://doi.org/10.3390/nano13081306.

[75] E. Gómez-de-Mariscal, M. Maška, A. Kotrbová, V. Pospíchalová, P. Matula, A. Muñoz-Barrutia, Deep-learning-based segmentation of small extracellular vesicles in transmission electron microscopy images, Sci. Rep. 9 (2019) 13211, https://doi.org/10.1038/s41598-019-49431-3.

[76] M.N. Jensen, E.M. Guerreiro, A. Enciso-Martinez, S.G. Kruglik, C. Otto, O. Snir, B. Ricaud, O.G. Hellesø, Identification of extracellular vesicles from their Raman spectra via self-supervised learning, Sci. Rep. 14 (2024) 6791, https://doi.org/10.1038/s41598-024-56788-7.

[77] W. Phillips, E. Willms, A.F. Hill, Understanding extracellular vesicle and nanoparticle heterogeneity: novel insights and considerations, Proteomics 21 (2021), https://doi.org/10.1002/pmic.202000118.

[78] D.K. Jeppesen, Q. Zhang, J.L. Franklin, R.J. Coffey, Extracellular vesicles and nanoparticles: emerging complexities, Trends Cell Biol. 33 (2023) 667–681, https://doi.org/10.1016/j.tcb.2023.01.002.

[79] Singh Grover, Bakshi, Quantitative structure-property relationships in pharmaceutical research - Part 2, Pharmaceut. Sci. Technol. Today 3 (2000) 50–57. http://www.ncbi.nlm.nih.gov/pubmed/10664573.

[80] Singh Grover, Bakshi, Quantitative structure-property relationships in pharmaceutical research - Part 1, Pharmaceut. Sci. Technol. Today 3 (2000) 28–35. http://www.ncbi.nlm.nih.gov/pubmed/10637598.

[81] E. Skotadis, A. Kanaris, E. Aslanidis, N. Kalatzis, F. Chatzipapadopoulos, N. Marianos, D. Tsoukalas, Identification of two commercial pesticides by a nanoparticle gas-sensing array, Sensors 21 (2021) 5803, https://doi.org/10.3390/s21175803.

[82] L. He, Ş.K. Özdemir, J. Zhu, W. Kim, L. Yang, Detecting single viruses and nanoparticles using whispering gallery microlasers, Nat. Nanotechnol. 6 (2011) 428–432, https://doi.org/10.1038/nnano.2011.99.

[83] B. Sun, M. Fernandez, A.S. Barnard, Machine learning for silver nanoparticle electron transfer property prediction, J. Chem. Inf. Model. 57 (2017) 2413–2423, https://doi.org/10.1021/acs.jcim.7b00272.

[84] J.A.M. Sidey-Gibbons, C.J. Sidey-Gibbons, Machine learning in medicine: a practical introduction, BMC Med. Res. Methodol. 19 (2019) 64, https://doi.org/10.1186/s12874-019-0681-4.

[85] J. Stetefeld, S.A. McKenna, T.R. Patel, Dynamic light scattering: a practical guide and applications in biomedical sciences, Biophys. Rev. 8 (2016) 409–427, https://doi.org/10.1007/s12551-016-0218-6.

[86] Z. Jia, J. Li, L. Gao, D. Yang, A. Kanaev, Dynamic light scattering: a powerful tool for in situ nanoparticle sizing, Colloids and Interfaces 7 (2023) 15, https://doi.org/10.3390/colloids7010015.

[87] P.M. Carvalho, M.R. Felício, N.C. Santos, S. Gonçalves, M.M. Domingues, Application of light scattering techniques to nanoparticle characterization and development, Front. Chem. 6 (2018), https://doi.org/10.3389/fchem.2018.00237.

[88] S. Bhattacharjee, DLS and zeta potential – what they are and what they are not? J. Control. Release. 235 (2016) 337–351, https://doi.org/10.1016/j.jconrel.2016.06.017.

[89] K. Rjoob, R. Bond, D. Finlay, V. McGilligan, S.J. Leslie, A. Rababah, A. Iftikhar, D. Guldenring, C. Knoery, A. McShane, A. Peace, P.W. Macfarlane, Machine learning and the electrocardiogram over two decades: time series and meta-analysis of the algorithms, evaluation metrics and applications, Artif. Intell. Med. 132 (2022) 102381, https://doi.org/10.1016/j.artmed.2022.102381.

[90] S. Ibsen, A. Sonnenberg, C. Schutt, R. Mukthavaram, Y. Yeh, I. Ortac, S. Manouchehri, S. Kesari, S. Esener, M.J. Heller, Recovery of drug delivery nanoparticles from human plasma using an electrokinetic platform technology, Small 11 (2015) 5088–5096, https://doi.org/10.1002/smll.201500892.

[91] A. Moore, M. Bell, XGBoost, A novel explainable ai technique, in the prediction of myocardial infarction: a UK biobank cohort study, Clin. Med. Insights Cardiol. 16 (2022), https://doi.org/10.1177/11795468221133611.

[92] D. Gottesman, Fault-tolerant quantum computation with local gates, J. Mod. Opt. (2000). https://arxiv.org/pdf/quant-ph/9903099.pdf.

[93] M. Cerezo, A. Arrasmith, R. Babbush, S.C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, P.J. Coles, Variational quantum algorithms, Nat. Rev. Phys. 3 (2021) 625–644, https://doi.org/10.1038/s42254-021-00348-9.

[94] J.-M. Liang, S.-Q. Shen, M. Li, L. Li, Variational quantum algorithms for dimensionality reduction and classification, Phys. Rev. 101 (2020) 032323, https://doi.org/10.1103/PhysRevA.101.032323.

[95] D. Sierra-Sosa, J. Arcila-Moreno, C. Garcia-Zapirain, Begonya Castillo-Olea, A. Elmaghraby, Dementia Prediction Applying Variational Quantum Classifier, (n. d.). https://doi.org/https://doi.org/10.48550/arXiv.2007.08653.

[96] D. Franklin, F.T. Chong, Challenges in Reliable Quantum Computing, in: Nano, Quantum Mol. Comput., Kluwer Academic Publishers, Boston, n.d.: pp. 247–266. https://doi.org/10.1007/1-4020-8068-9_8.

[97] F.J. Shaikh, D.S. Rao, Prediction of cancer disease using machine learning approach, Mater. Today Proc. 50 (2022) 40–47, https://doi.org/10.1016/j.matpr.2021.03.625.

[98] Z. Cai, R.C. Poulos, J. Liu, Q. Zhong, Machine learning for multi-omics data integration in cancer, iScience 25 (2022) 103798, https://doi.org/10.1016/j.isci.2022.103798.

[99] H.J. Chen, A. Poran, A.M. Unni, S.X. Huang, O. Elemento, H.-W. Snoeck, H. Varmus, Generation of pulmonary neuroendocrine cells and SCLC-like tumors from human embryonic stem cells, J. Exp. Med. 216 (2019) 674–687, https://doi.org/10.1084/jem.20181155.

[100] A. Thakur, G. Qiu, S.-P. Ng, J. Guan, J. Yue, Y. Lee, C.-M.L. Wu, Direct detection of two different tumor-derived extracellular vesicles by SAM-AuNIs LSPR biosensor, Biosens. Bioelectron. 94 (2017) 400–407, https://doi.org/10.1016/j.bios.2017.03.036.

[101] A. Thakur, C. Xu, W.K. Li, G. Qiu, B. He, S.-P. Ng, C.-M.L. Wu, Y. Lee, In vivo liquid biopsy for glioblastoma malignancy by the AFM and LSPR based sensing of exosomal CD44 and CD133 in a mouse model, Biosens. Bioelectron. 191 (2021) 113476, https://doi.org/10.1016/j.bios.2021.113476.