Article

Machine-learning for Precision Breast Cancer Diagnosis and Prediction of the Nanoparticles Cellular internalization

Maha Alafeef, Indrajit Srivastava, and Dipanjan Pan

ACS Sens., **Just Accepted Manuscript** • Publication Date (Web): 28 May 2020

Machine-learning for Precision Breast Cancer Diagnosis and Prediction of the Nanoparticles Cellular internalization

Maha Alafeef[1, 2], Indrajit Srivastava[1], and Dipanjan Pan[1,3]*

[1]Bioengineering Department, University of Illinois at Urbana-Champaign, Urbana, 61801, USA

[2]Biomedical Engineering Department, Jordan University of Science and Technology, Irbid, Jordan

[3]Department of Radiology and Chemical, Biochemical and Environmental Engineering, University of Maryland, Baltimore and University of Maryland Baltimore County, Baltimore, USA

KEYWORDS: Carbon nanoparticles; breast cancer; pattern recognition; cancer diagnosis; artificial neural network, biosensor

ABSTRACT: In the field of theranostics, diagnostic nanoparticles are designed to collect highly patient-selective disease profiles, which is then leveraged by a set of nanotherapeutics to improve the therapeutic results. Despite their early promise, high inter-patient and intra-tumoral heterogeneities make any rational design and analysis of these theranostics platforms remained extremely problematic. Recent advances in deep learning-based tools may help bridge this gap, using pattern recognition algorithms for better diagnostic precision and therapeutic outcome. Triple-negative breast cancer (TNBC) is a conundrum because of the complex molecular diversity, making its diagnosis and therapy challenging. To address these challenges, we propose a method to predict the cellular internalization of nanoparticles (NPs) against different cancer stages using artificial intelligence. Here we demonstrate for the first time that a combination of machine learning (ML) algorithm and characteristic cellular uptake responses for individual cancer cell types can be successfully used to classify various cancer cell types. Utilizing this approach, we can optimize the nanomaterials to get an optimum structure–internalization response for a given particle. This methodology predicted the structure–internalization response of the evaluated nanoparticles with remarkable accuracy (Q^2=0.9). We anticipate that it can reduce the effort by minimizing the number of nanoparticles that need to be tested and could be utilized as a screening tool for designing nanotherapeutics. Following this, we have proposed a diagnostic nanomaterial-based platform used to assemble a patient-specific cancer profile with the assistant of machine-learning (ML). The platform is comprised of eight carbon nanoparticles (CNPs) with multifarious surface chemistries that can differentiate healthy breast cells from cancerous cells and then subclassify TNBC cells vs. non-TNBC cells, within the TNBC group. The artificial neural network (ANN) algorithm has been successfully used in identifying the type of cancer cells from 36 unknown cancer samples with an overall accuracy of >98%, providing potential applications in cancer diagnostics.

Nanomedicine-based technologies for cancer detection and treatments have been evolving rapidly for early disease detection and to improve drug efficacy with reduced side effects.[1] Despite all the promise, nanomedicine so far has found limited success in clinical translation due to high heterogeneities presented at the inter-patient and intra-tumoral levels.[2] It is also now known that various physiological states amongst patients contribute to their heterogeneity. Furthermore, diversity in vascular permeability leads to different drug responses in the patient population. Similar unpredictability exists in homing agents' expression in the case of molecularly targeted nanomedicine. As an important example, breast cancer is one of the most prevalent invasive malignancies and the leading cause of death among women.[3-5] Although the outcome of early screening and treatment of breast cancer has improved significantly in recent years, current diagnostic methods and treatment options still have unconquered limitations, especially when it comes to triple-negative breast cancer (TNBC).[6] TNBC, defined by the lack of expression or low expression of progesterone receptor (PR), estrogen receptor (ER), and human epidermal growth factor receptor 2 (HER2),[7-8] account for 12-17% of breast cancer cases.[9] As a highly heterogeneous group of breast cancer,[10] TNBC demonstrate aggressive clinical behavior,[11] high risks of metastasis, and death within five years after diagnosis.[12] Currently, over a hundred drugs based on nanostructures, are in various stages of clinical uses.[13] However, until now, no effective standard therapy has been approved for TNBC by the U.S. Food and Drug Administration (FDA).[6-7, 11]

Nanoparticles (NPs) are known for their multifunctionality and tunable surface chemistries.[14] However, little emphasis has been paid to show how structural changes affect their interaction with various subtypes of breast cancer cells, e.g. TNBC. Some of these properties, include their ability to cross targeted-cell membranes without the need for transfection vesicle via selective endocytosis pathway. The particle nanoscopic dimension, highly active surface, and the ability to interact differently at targeted cellular levels open up many possibilities. The process of finding the optimum design of NPs is dominated by trial and error methodology, consuming a significant amount of time, effort, and resources. The integration of nanotechnology with artificial intelligence (AI) may offer critical insights and help to seek specific information for guiding decisions

in diagnosis and therapy.[15-17] In fact, AI, including machine learning (ML) and deep learning approaches are already showing promises to improve the translational success of nanomedicines by exploiting large data sets and transforming them into decision-making tools, while avoiding the conventional trial-and-error process.[18-22]

In this work, we report the use of machine learning to estimate the cellular internalization of nanoparticles based on their surface design. Further, we develop an ML-based model to sense breast cancer cells by exploiting the internalization of eight differently functionalized CNPs. Quantitative measures of NPs cellular internalization were evaluated via different endocytic pathways assessed. Using AI approaches, we were able to get a rank of the order of the NPs properties to reach the optimum design which may help to improve therapeutic efficacy. This model can accurately predict the internalization of CNPs based on their structural features. In addition to that, we utilized the cellular internalization profile of the proposed diagnostic nanoparticles to assemble a patient-specific cancer profile using ML. For our studies, carbon nanoparticles were selected because of their high stability,[23] easily tunable surface chemistry,[24] and its biocompatibility.[25-29] The nanoparticle surface chemistry was functionally engineered to tune nanoparticle-cell interactions.[30-31] Five cell lines, one non-cancerous (normal), one breast cancer (non-TNBC), and three TNBC (BT-549, HCC 1806 and HCC1143) were used as the target. Utilizing these components, we designed a highly efficient platform to render distinct fingerprints for individual cancer cell types.

EXPERIMENTAL SECTION

All materials, reagents, instruments, and experimental methods used in this study are described in detail in the Supporting Information.

RESULTS AND DISCUSSION

Synthesis of Model Nanoparticles and their Physicochemical Characterization

In this work, carbon nanoparticles (CNPs) have been studied as a model NPs platform because they can be efficiently synthesized with controlled size, biocompatibility, and tunable surface chemistry.[32-33] Carbon nanoparticles are luminescent materials mainly composed of either graphitic or amorphous carbon as building blocks, occasionally in the presence of some doped heteroatoms. CNPs are characterized by their excellent biocompatibility and interesting multi-color fluorescence properties, which enable them as potential candidates for wide ranges of biomedical applications from imaging to biosensing. We investigated the internalization pattern of eight CNPs, synthesized by utilizing two main techniques of passivation (pre or past passivation). The CNPs have been derived from Agave nectar based on protocols established in our lab. [34-35] The surface of these as-synthesized nanoparticles were passivated with linear or branched organic macromolecules used for the -passivation of the nanoparticles. To understand the influence of surface charges on the cellular internalization, neutral, negative, and positively charged macromolecules were used for surface modification. To make the carbon surface either positively charged or neutral, nanoparticles surface was passivated with PEI, and PEG (400, 10K, and 20K), respectively. To make the surfaces negatively charged, phosphate (-PO_4), sulfonates (–SO_3H), and thiol (–SH) functionalities were used (Figure 1a). Zeta potential measurements were conducted and revealed a

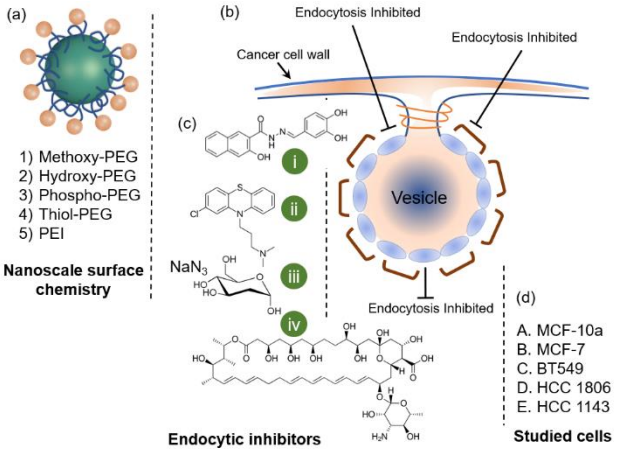

Figure 1. (a) CNPs architecture and surface chemistries. (b) Illustration of the effect of the endocytosis inhibitors on the cell. (c) Type of the endocytosis inhibitors utilized the CNPs-cellular internalization protocol including (i) Dynasore for inhibiting the Dynamin-GTPase pathway, (ii) Chlorpromazine for inhibiting the clathrin-mediated endocytosis, (iii) NaN_3 + DOG for ATP depletion, and (iv) Nystatin for the inhibition of clathrin-independent endocytosis. (d) The type of targeted cell lines studied throughout this work.

negative zeta potential for bare CNPs (−22±02 mV) owing to the hydroxyl and carboxyl groups on the CNPs surface. Zeta-

the potential of CNP-PEG400, CNP-PEG10K, CNP-PEG20K, CNP-PO_4, CNP-SO_3H, CNP-SH were recorded to be negative due to the passivation of neutral or anionic macromolecules. Such surface modification resulted in more negative zeta potential due to the presence of ethylene glycols, phosphates, sulfonates, and thiols on the CNPs surface. On the other hand, the bare CNP passivated with a cationic polymer, PEI (CNP-PEI), possesses a positive zeta potential (19±6mV), which can be explained by the presence of primary, secondary, and tertiary amines (Figure 2b, Table S2). The negative zeta potential on the surface of the pristine CNPs (no surface coating) has been attributed to the over-abundance of carboxyl, hydroxyl, and carbonyl groups on their nanoscale surface. However, when a polymeric coating is wrapped around these CNPs, it leads to the lowering of their "effective" zeta potential values, as was evident from the case where PEG 400, PEG 10K, and PEG20K were wrapped around CNPs and resulted in lowering of the zeta potential. Irrespective of the molecular weight of PEG, these are charge-neutral molecules and do not add any charge to passivated nanoparticles, but due to their extended chains affect their layer of hydration to a significant extent. This layer of hydration presumably is masked the surface charge on CNPs, based on longer or smaller chain length in higher or lower MW PEGs. Similarly, subsequent passivation of CNPs with anionic macromolecules resulted in the lowering of the zeta potential compared to pristine CNPs but was still negative owing to the profusion of phosphate groups (PEG-PO_4) or sulfonate groups (PEG-SO_3H). The physicochemical characterization of the surface-modified CNPs and control (pristine/ bare CNPs) were investigated by applying several techniques. To reveal their size, as well as the morphology of the synthesized CNPs, dynamic light scattering (DLS) and transmission electron microscopy (TEM) measurements, were carried out. Bare and surface modified CNPs exhibit a number average hydrodynamic diameter ranging from 8-40 nm (Figure 2a, Table S1). Transmission electron microscopy (TEM) was performed to investigates the anhydrous state diameter and morphologies. Representative TEM images of the

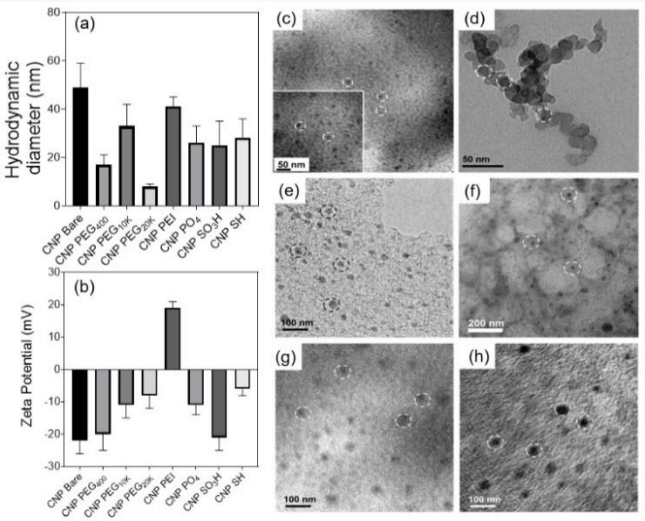

Figure 2. Physiochemical characterization of the as-synthesized CNPs (a) Hydrodynamic diameter measurements of via dynamic light scattering (DLS) (b) Zeta potential measurements for all CNPs, (c-h) Transmission electron microscopy images of CNP (c), CNP PEG10K (d) CNP PO_4 (e) CNP PEI (f) CNP SH (g) and CNP SO_3H (h).

eight particles are shown in Figure 2c-h. The synthesized CNPs possess different morphological structure and dynamic dimeters accompanying with different passivation. The synthesized CNPs via post- and pre-passivation result in different hydrodynamic diameters, which has been confirmed through several studies done by our group.[34-36] For example, CNPs which are post-passivated with dendrimers such as polyamidoamine (PAMAM,) and bis(hydroxymethyl) propionic acid (bis-MPA), etc. were found to be significantly larger than pristine CNPs due to their large bushy chains, which gets extended when they are suspended in aqueous media. However, for pre-passivation, where both the precursor of CNPs (sucrose) and PEG polymers (such as PEG400 or PEG20K) were taken, we envision a tightly wrapped thin layer of the PEGylated polymer around the CNPs leading to a decreased hydrodynamic diameter. The as-synthesized CNPs were found to be well suspended in an aqueous medium without forming an aggregate as it has been confirmed using the DLS measurements (Table S1) and the TEM images (Figure 2c-h). Although the physicochemical characterization of the synthesized particles showed minor variation, they were shown to maintain a consistent measure in terms of hydrodynamic diameters, zeta-potential, and polydispersity index (PDI).

To further investigate the effect of the pre- and post- passivation synthetic strategy on the optical properties of as-synthesized CNPs, UV-Vis absorption spectroscopy and fluorescence spectroscopy measurements were carried out. UV-Vis spectra exhibit characteristic peaks between 250-300 nm (Figure S1). The fluorescence spectra of the bare- and modified CNPs at 365 nm excitation are shown in Figure S2. For evaluating the surface chemistry of CNPs, we utilized Fourier-transform infrared (FT-IR) spectroscopy (Figure S3). Taking all together, the surface modification provides the CNPs with distinct physicochemical properties in terms of size, morphology, surface properties, and optical characteristics, foreshadow distinguish cellular internalization response.

Mechanism of Cellular Internalization

Nanoparticles are found to enter cells through multiple pathways, with one of the more common routes being endocytosis.[37] This mechanism of intracellular transport depends on the functional groups presented on the surface of nanoparticles and often varies for different cell types.[38] The intracellular internalization has been investigated by utilizing different endocytosis inhibitors (Table S3) prior to their entry by energy, caveolar, and clathrin-mediated pathways. Dynasore, which is an inhibitor for the dynamin GTPase could be utilized to inhibit the dynamin-dependent endocytosis in cells.[39-40] Dynamin-dependent pathways are considered as clathrin-independent endocytosis.[41-42] The clathrin-dependent path can be blocked by using chlorpromazine (CPM), which a cationic amphiphilic drug and known to reduce the formation of clathrin-coated pits on the cell surface.[43-45] The uptake of nanoparticles by the energy dependence pathway can be studied using sodium azide (NaN_3) and 2-deoxyglucose (DOG) which inhibits glycogenolysis and cellular respiration, respectively.[38, 46-47] A sterol-binding agent, nystatin, which disassembles cholesterol in the membrane, has been used to study clathrin-independent inhibition of endocytosis.[48-49]

To analyze the mechanism of CNPs internalization and get a better understanding of how the physicochemical properties affect the cellular internalization in the case of normal and different cancer stages, we utilized internalization inhibition assay applied on CNPs established in our previous work.[50] It is a high-throughput assay that allows for fast and quantitative evaluation of the CNPs cellular internalization. The protocol is based on culturing different types of cell lines and treated each well with a distinct kind of CNPs (i.e., CNP1-CNP8) after selective blocking of endocytic pathways to investigate their cellular internalization. The working hypothesis of this screening protocol was that each inhibitor could block a particular endocytic pathway and consequently block the entry of the CNPs, using this specific pathway. Therefore, decreasing the cellular internalization eventually leads to increases in the cell viability.

The output value from this screening protocol in terms of cell viability was transformed to fold increase values—referred to as the internalization response throughout the manuscript. Using this protocol, we estimated the internalization response to eight CNPs against four inhibitors at five cell lines (Figure 1c). Three biological replicates were used to acquire the fold-increase in cell uptake for each sample. In this work, more than 450 cell culture wells, in addition to the controls, have been used, and more than 450 experiments and measurements were analyzed.

These data revealed a distinct pattern of CNPs internalization for each cell line, indicating their metastasis stage. Figure S4 illustrates that for MCF-10a (non-cancerous cells), the clathrin-mediated pathway was utilized for uptake irrespective of the surface chemistry, whereas for MCF-7 (early-stage cancerous cells), only PEGylated CNPs internalized via a combination of both energy-dependent and clathrin-mediated pathways. Interestingly, for BT-549 (TNBC cell line), a combination of all the four investigated routes were active, suggesting better success rates for internalization. In contrast, this was not the case for HCC 1806 or HCC 1143 (both TNBC cell lines). In these cells, a lesser response from these endocytic inhibitors was found, with CNP-PO_4 for the HCC 1143 cell lines being internalized profoundly via the dynamin-dependent pathway. This indicates that engineering the CNPs by modifying their surface chemistry will affect their cellular internalization.

Comparing the cellular internalization among cells reveal a distinct entry mechanism that is dominated by one or more pathways.

To gain insight into the effect of nanoparticle properties on their cellular behavior, we investigate the cellular internalization responses at the nanoparticles level (Figure S5). Figure S6 depicts the cellular internalization heat maps of neutral, positive, and negative CNPs. CNP-PEI and CNP-PEG20K exhibiting a preference toward the clathrin-dependent pathway for internalizing BT-549 cells. Further, CNP-PEG400 exhibits a favor to internalize through the energy-dependent pathway in BT-549. CNPs-PO_4 exhibits a favor to internalize the BT-549 cell line via both clathrin-dependent and dynamin-dependent endocytic pathways. However, in HCC-1143, the internalization of CNPs-PO_4 is dominated by the dynamin-dependent endocytosis pathway. This result indicates the importance of CNPs structural parameters in determining cellular internalization. Each CNPs exhibits a dominant cellular mediated endocytosis pathway that is dependent on both cancer stage and physicochemical properties.

To further confirm the cellular uptake of CNPs in cancer cells and the validity of our MTT assay results, we performed a quantitative uptake study of CNP-PEG10K using flow cytometry.[51-52] Figure S6 shows the quantitative cellular internalization of CNP-PEG10K in the presence of various pharmacological inhibitors (Figure S7). These results corroborated our MTT results as well. Establishing a correlation or standard curve between the amount of the CNPs entering the cells and the fold increase value from our MTT assay will be considered for follow-up work. It is worth mentioning that, in the cell culture media or fetal bovine serum (FBS), a significantly smaller number of proteins are present, and in spite of forming a corona shell around nanoparticles does not affect its physiochemical properties throughout the course of our inhibitor study (Table S4). However, the protein corona formation in blood plasma or blood serum samples is more rapid because of the presence of a multitude of proteins and lipids, and thus, it should be taken into consideration for potential *in vivo* studies.[53-54]

These internalization patterns were found to be unique for each cell. However, no clear trend has been concluded to explain the relationship between the NPs size, surface chemistries, and other design parameters with cancer or its stages. From here, a critical need arises for developing a holistic model that can predict the cellular internalization of each particle and reveal the non-intuitive relationships between the drug carries properties—nanoparticle in this case—and the targeted cell. Such a model can provide the necessary information on dominating chemistry governing the internalization of nanoparticles in a specific cell line. Thus, delegating better treatment and imaging strategies beforehand with higher success rate.[55]

Artificial Neural Network Confirms Trends of the Importance of Nanomaterial Properties in the Cellular Internalization.

In silico methods of hierarchical clustering (HCA),[56] linear discrimination analysis (LDA),[57] k-nearest neighbors (kNN),[58] and principal component analysis (PCA)[58] are the most common methods used to design predictive quantitative nano-bio activity relationship model. However, the performance of the statistical techniques is limited by their inherent algorithm, which needs to be explicitly programmed and cannot be generalized to compensate for the complexity in attributes governing the nanoparticle internalization activity. The main reason is that the nanoparticle's physiochemical characteristics possess an intricate non-intuitive internalization pattern, which may vary with the cell's subtypes. From a modeling point of view, HCA,

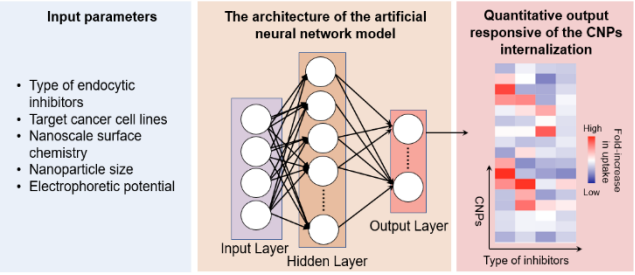

Figure 3. Training of the artificial neural network model using the design parameters of eight CNPs to predicts their cellular internalization. The particle size, ζ-potential, and the surface chemistry of the eight CNPs, combined with the inhibitor and the targeted cell, were used as inputs for the network. The model assigns each CNP a quantitative number for the cellular internalization through the specific endocytosis pathway.

LDA, KNN, and PCA are not efficient when dealing with high dimensionality data and in cases where huge variations and differences are observed in the data points.[59-60] To circumvent this problem, we utilized a supervised machine learning algorithm known as an artificial neural network (ANN), to estimate the internalization of nanoparticles *in vitro* and used these patterns to predict the stage of cancer progression as it is effective at revealing complex relationships in big data. ANN is an artificial-intelligent algorithm that mimics the human brain in recognizing the sophisticated non-linear relationships of features that differentiate two or more classes of objects or events.[61-66] ANN surpasses traditional pattern recognition algorithms because of their ability to 'learn' from the input data on their own, and therefore, there is no need to program them with formulas. The neural network model creating multiple intermediary transformation layers known as hidden layers that compute the weight of each neuron comprises the network to predict input-output relationships which granted ANN their high sensitivity. Even in the case of data with small variations, the network estimates the backpropagating error of these non-intuitive differences to update the network architecture during the training process until the error reaches a minimum.[61] This adaptive approach is promising for analyzing multiparametric data without prior knowledge of the input-output relationships. Thus, the ANN model can efficiently map relationships in data with high dimensionality in an unbiased manner even if subtle differences are presented between data points. As our aim in this study is to build a predictive quantitative NPs internalization activity relationship (QNIAR) model for nanoparticle internalization. Previous studies have been unable to achieve high accuracy in correlating NPs physiochemical characteristics and the targeted cell type inputs to biological behavior through conventional analyses tools.[67-68] Further, because the studied parameters were inter-dependent, the conventional method may unable to reveal the non-intuitive complex relationship between the inputs and output. Utilizing machine learning as a pattern recognition algorithm will address these limitations and also eliminates the time-consuming trial and error process in identifying the potential particle candidate with optimum tumor internalization characteristics.

In this work, we used supervised techniques to automatically estimates the internalization behavior of CNPs against cells with the expectation that attributes related to cell-nanoparticles interaction would enhance the performance of the model. The network architecture comprises of five nodes (five inputs, Figure 3) and one hidden layer that reveals the input-output relationships using the hyperbolic tangent activation

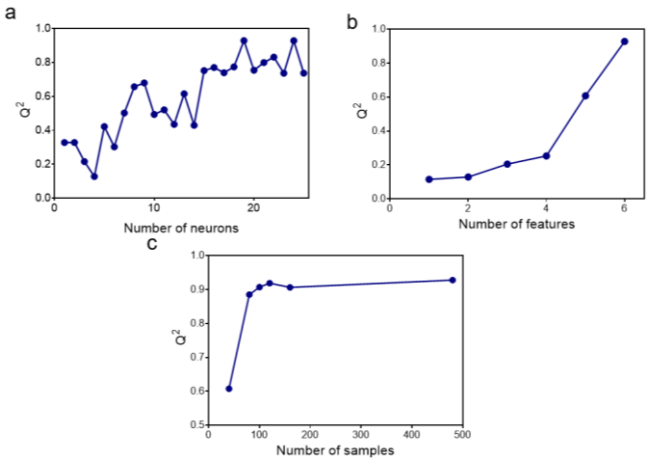

Figure 4. Analysis by a supervised artificial neural network. (a) The optimization of the network architecture in term of the number of neurons which provides the highest Q^2 value. (b), Highest Q^2 statistics of several attributes' combinations are shown across different numbers of parameters. (c) Q^2 performance when train the network on randomly selected CNPs subsample and testing predictions on the remaining CNPs samples.

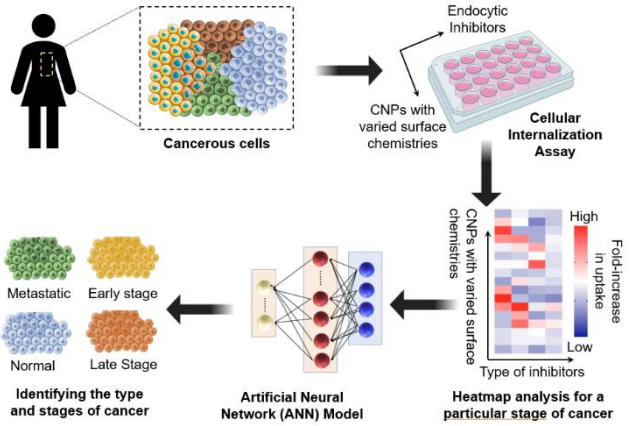

Figure 5. Schematic representation of the diagnosis of breast cancer using the proposed sensor array. The internalization response of each CNPs against each cell type was quantified by the proposed MTT assay in the presence of four inhibitors. These responses, known as fold-increase in the cellular uptake were classified using the ANN model to differentiate the normal from cancerous cells and identify the type of cancer.

function. To start with, we construct the QNIAR network, by first analyzing our training data set that consists of both the physicochemical characteristics data of the CNPs, and the targeted cell type for inputs and cellular internalization as the outputs (Figure 4). The quantitative cellular internalization was estimated using the aforementioned high-throughput screening protocol to establish CNPs libraries, the cellular internalization of eight CNPs toward five targeted cell lines are assessed by the cellular internalization screening protocol. The model inputs are 1) nanoparticle size, 2) ζ-potential, 3) surface chemistry of the eight CNPs, 4) type of inhibitor and 5) targeted cell. Whereas, the model outputs is the internalization response of various CNPs, quantified using the screening protocol described previously, by utilizing different stages of breast cancer cell lines, such that it included both TNBC cell lines, i.e. BT-549, HCC 1806 and HCC1143 along with non-TNBC (MCF-7, early-stage breast cancer cell line) cell lines and non-tumorigenic breast cells (MCF-10a).

The pharmacological inhibitors used in the screening protocol were selected to avoid causing any toxicity effect by themselves during the incubation period (Figure 1b). Moreover, they possess no impact on the actin cytoskeleton post-treatment.[50, 69] The workflow-enabled us to train the ANN model using CNPs design parameters and the targeted-cell type as inputs and nanoparticle internalization activities as outputs (Figure 3). By feeding these inputs attributes to the developed model and providing nanoparticle internalization, we could capture nonlinearity of CNPs-cellular internalization activity and map these attributes with the cellular internalization without the need for explicit programming the model. The model has been trained through an iterative feedforward and Levenberg-Marquardt backpropagation learning based on K-fold cross-validation, and an early termination rule has been used to avoid the overfitting. Further, the testing cross-validation was performed using the Q^2 statistic.[70] Q^2 is a parameter used to assess the accuracy of the estimated cellular internalization response against actual values, and it is ranging from $-\infty$ to 1. Zero indicates no estimation power, which is in a sense the same as guessing the mean, and 1 indicates perfect estimation.[71] ANN algorithm designed to imitate intelligent human cognition behavior in deciding on the action in response to input data for achieving a predetermined aim. To get the maximum possible efficiency of the developed model, we optimized the architecture of the hidden layer by experimentally tune the number of neurons comprising the model. The hidden layers represent the computational engine for the network architecture which can be tuned to optimize the learning capability of the model. The learning performance of the model found to be enhanced—increase Q^2 statistics—by increasing the neuron number, however, this trend found to be non-linear (Figure 4a). The network architecture which is consists of 24 neurons exhibits the best learning performance and provides the highest Q^2 value. Therefore, we use this network architecture for further data analysis and input-output relationship mapping throughout the manuscript. Once we successfully trained the network by performing N-1 cross-validation,[72] we then subsequently evaluated their performance using a blinded test set, independent from the training set, where the internalization pattern against each cell type was not known a priori. The testing set was used to validate the model accuracy and ensure no overfitting has occurred.

We challenged the model to predict the fold increase value, which represents the CNPs-cellular internalization. The prediction accuracy quantified by Q^2 statistics and the model estimates the cellular internalization of CNPs with Q^2=0.9, indication the reliability and the efficiency of the developed network. It noteworthy to mention that the neural network model was accurate in predicting the cellular internalization with mean squared error (MSE) of 3.3284×10^{-4}.

Next, we trained the network with several combinations of properties or inputs (i.e., two at a time, three at a time, and so on) and analyzed their Q^2 statistics (Figure 4b). An increment in the Q^2 performance was observed when additional attributes were added to the ANN models. However, we found the maximum increase in Q^2 when features 5 and 6 (nanoparticles surface chemistry and the stage of the cancer metastasis) were added. This presumably indicates that the CNPs-cell internalization activity is predominantly governed by the particle surface chemistry and the type of the targeted cell. It worth mentioning that each input attribute has been treated as an independent variable, thus the order of the

attributes does not contribute to the accuracy, specificity, or any other model performance.

Capturing the Full Images of Cellular Internalization with Minimum CNPs Synthesis

It is known that capturing the maximum design-internalization response with minimum synthesis and testing is a challenging aspect in nanomedicine. In an attempt to evaluate whether an equivalent Q^2 value is maintained with less, randomly selected CNPs with different properties and surface chemistries, a randomly selected portion of the data has been introduced to the ANN model. This scenario is found in nanomedicine screening when synthesis and testing of full chemistry libraries are unrealistic. However, the examination of ample nanoparticles structure space is highly desired. In this situation, a proper solution is to synthesize and design a random batch that would capture the most significant design parameters and then evaluate their design performance. To simulate this methodology, we trained the ANN model on an arbitrary selection of CNPs with different surface chemistry and testing predictions on the rest of the points (Figure 4c). We identify the diminishing returns point (Figure 5c), which indicates the minimum amount of tested CNPs with the highest Q^2 performance. This point is 100 CNPs data points (out of ~ 500 sample point) with Q^2 = 0.9030 for CNPs with different surface chemistry of. These sample numbers represent 20.83% of the total number of CNPs-cancer cell samples. To assure that the presence of triplicated points is not affecting the generalization of the model, we tested the performance of the model using 160 data points, by taking the average of each triplicate, which provides Q^2 = 0.95. Almost similar predicting performance (Q^2= 0.94) has been obtained from 32 samples, indicating the capability of the ML in predicting the internalization of small data samples with a performance comparable to a large dataset, which further supports our previous findings. These results are suggesting that the ANN model can predict design–internalization relationships of a relatively large CNPs library using a small number of randomly selected CNPs samples. In practice, the combination of ease-to engineer CNPs design and characterization approach and the artificial intelligent analysis exhibit that hybridizing both the experimental and computational methods can predict the design–internalization in the targeted cancer cell of tens of thousands of nanoparticles structures using a small nanoparticles subset.

Supervised Machine-Learning for Cancer Diagnosis and Staging

Once the different internalization characteristics of CNPs within the targeted cells were established, we utilize these patterns to accurately identify the presence of cancer and staging them. To this end, the particle-inhibitor conjugates were used to determine the cancer cell type. Five targeted cancer cells were utilized to cover normal, and breast cancer (TNBC and non-TNBC) groups. Within this set, there were three targeted cells that belonged to the TNBC group, providing a challenging testbed for cancer diagnosis. The results obtained from these studies, for the first time, highlight the feasibility of utilizing the cellular internalization of the nanoparticles to indicates the presence of cancer cells and their stage. The CNPs internalization signatures has been used to predict the cell type and estimate the stage of cancer progression. Figure 5 illustrates a schematic representation of the process of predicting the type of cancer cell in the sample using the proposed sensor array combined with artificial intelligence. We utilized the strength of neural network inter-neuron connections as a quantitative weight of the input features and activation threshold for releasing information[66, 73-74] to map the relationship between the input and

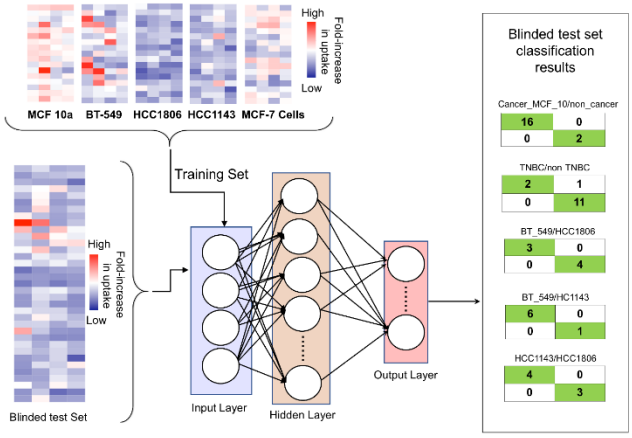

Figure 6. The artificial neural network model for cancer diagnosis. The model was trained using 70% of the data. Next, the performance of the network was evaluated using a blinded test set represents 30% of the data. The confusion matrix of the test set shows that one data point was misclassified with an overall accuracy of 98.1%.

output. This analysis superior to statistical methods such as LDA and provides the advantages of revealing the complex-nonlinear relationship that underlies CNPs internalization associated with different breast cancer cases.[74-75] The network takes into account the internalization of CNP in the presence of the four-endocytosis inhibitors individually as input parameters to predict the type of the cancer cell as output. Figure S8 depicts the architecture of the ANN used for breast cancer diagnosis and staging of cancer progression.

The network was cycled through an iterative process of feedforward and Levenberg-Marquardt backpropagation learning, where the classifier was trained in a hierarchal manner. We first generated a prediction algorithm by running ANN on a training set (ST) (Figure 6), where the targeted class of each input data point is a priori known. Throughout the training process, the network updated the weight and the activation threshold of the hidden layer's neurons.[20, 76]

The training procedure has been done in three consecutive steps. First, the internalization response patterns against the CNPs sensor array were first used for the sensor training, which was classified into two distinct groups according to the presence of cancer (normal/cancerous). These results illustrate that ANN allows the discrimination of breast cancer cases based on their internalization responses. Second, the classifier was trained to further group the breast cancer data into TNBC and non-TNBC classes. Finally, a further step was taken to classify the cancer cells within the TNBC group. The targeted cell types were classified into either BT-549, HCC1143, or HCC1806. Once the network is successfully trained, a blinded study has been performed to validate the model accuracy and to deny the model overfitting. We evaluate the performance of the sensor array against 36 unknown samples, acquired from the proposed screening assay. Based on the network established by the training sets of the 36 unknown cell samples, only one sample was incorrectly identified, affording an identification accuracy of 98.1%. Figure 6 depicts the confusion matrix of each subclass. These results confirmed that our sensor array shows substantial promise for both the identification of cancerous from healthy subjects and determining the stage of cancer cell and its underlying subtypes.

The system need only four readouts from each sample to give the decision. These findings have been supported by the results of the training phase. This assay outperforms previously reported sensors arrays[77-78] in terms of simplicity and efficiency. Thus, we anticipate

that our system holds a promise to translate to clinical practice. We envisioned that the end-user would only record the absorbance of the MTT reagent for one of the CNPs in the presence of the four inhibitors, and the ML model will map these data to identify the cancer type and then display the results to the end-user. The developed system represents a simple platform for breast cancer diagnosis with potential use in clinical practice.

In practical situations, usually, the biopsy from the patient has a blend of cancer cells at different stages. The algorithm can be further improved to analyze a blended sample of cells by adjusting the network architecture in term of the number of hidden layers, the number of neurons in each layer, and the training function.[63-64, 79-80]

The technology proposed in this work is designed to address an urgent need. Thus, it is essential for the system to be cost-effective, easy-to-use, and user-friendly. To this end, we used the absorbance signal as read-out obtained from the sensor, which will allow the integration of our technology into a portable system. We envision that the end-user will add the sample to a well-plate containing a mixture of the desired CNPs, endocytosis inhibitors, and the MTT reagent. The samples will be incubated for a while, then the absorbance will be recorded and transmitted, processed, and finally, the results will be displayed to the end-user. Therefore, this technology provides an affordable system that can be used in cancer staging, which is user-friendly, cost-effective with great potential, and applicability. The developed technology possessed many advantages over conventional techniques—like flow cytometry—in terms of the possibility to be used at the patient point-of-care. Flow cytometry requires extensive sample preparation and expatriates. Further, reliance on the fluorescence as a readout increases the cost and needs specialized equipment.

CONCLUSION

Nanomedicines holds the promise of transforming today's healthcare by offering efficient therapies and early-stage diagnostics. However, their clinical translation faces numerous challenges. Representative hurdles that are hindering the clinical translation of nanomedicines include scaling-up, *in vivo* instability and bioavailability, potential toxicity, immune response, regulatory barriers, and disease heterogenicity. AI-based tools integrated with a full range of structure–internalization relationships study is disposed to improve the complete nanomedicine development process by exploiting large, multisource data, to transform them into usable and actionable knowledge and decision-making tools, while leaving behind years of trial-and-error in drug development. The proposed study discussed the importance of considering the full range of structure–internalization relationships when designing nanomedicine by high-throughput methodology. Our study revealed that the internalization signature of each CNPs against different cancer stages have been found to be unique. These unique patterns were classified using a robust machine learning algorithm known as the artificial neural network. Through the application of ANN, we are able to sort the stage of various cancer stages and differentiate within the TNBC classes with an overall accuracy of 98.1%. Our study used supervised machine learning to analyze the data generated for ~500 combinations of CNPs structures and cancer stages. We examined five different cell lines, and further work will be done in the future to expand the work to cover other cancer types. We utilized the ANN independently for libraries defined by the selection of size, surface charge, and surface functionality with the benefit of knowing that these features influence the cellular internalization. Integrating AI tools with 'nano-bio' structure-activity relationships can be generalized by expanding the training set to cover other platforms and includes more chemical features. This paper presents a standalone sensor array built on machine learning techniques for cancer screening based on nanoparticle-cell internalization. The system offers rapid, efficient, and highly accurate identification of tumor cells and can be extended to define numerous other nanoplatforms and potentially different types of diseased cells.

AUTHOR INFORMATION

Corresponding Author

E-mail: dipanjan@som.umaryland.edu; Dipanjan Pan*

Conflicts of interest

Prof. Dipanjan Pan is the founder/co-founder of three University-based start-ups. None of these entities, however, supported this work.

ACKNOWLEDGMENT

Authors gratefully acknowledge funding from the University of Illinois at Urbana-Champaign, Univeristy of Maryland Baltimore, NIBIB R03 EB028026, and National Science Foundation.

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

Materials, reagents, experimental methods, nanoparticles characterization, and cellular internalization responses (PDF).

REFERENCES

(1) Mukherjee, P.; Misra, S. K.; Gryka, M. C.; Chang, H.-H.; Tiwari, S.; Wilson, W. L.; Scott, J. W.; Bhargava, R.; Pan, D. Tunable Luminescent Carbon Nanospheres with Well-Defined Nanoscale Chemistry for Synchronized Imaging and Therapy. *Small* **2015,** *11* (36), 4691-4703.

(2) Pan, D. Theranostic Nanomedicine with Functional Nanoarchitecture. *Molecular Pharmaceutics* **2013,** *10* (3), 781-782.

(3) Eckhardt, B. L.; Francis, P. A.; Parker, B. S.; Anderson, R. L. Strategies for the Discovery and Development of Therapies for Metastatic Breast Cancer. *Nature Reviews Drug Discovery* **2012,** *11*, 479.

(4) Siegel, R.; Naishadham, D.; Jemal, A. Cancer Statistics, 2013. *CA Cancer J Clin* **2013,** *63* (1), 11-30.

(5) Bertucci, F.; Ng, C. K. Y.; Patsouris, A.; Droin, N.; Piscuoglio, S.; Carbuccia, N.; Soria, J. C.; Dien, A. T.; Adnani, Y.; Kamal, M.; Garnier, S.; Meurice, G.; Jimenez, M.; Dogan, S.; Verret, B.; Chaffanet, M.; Bachelot, T.; Campone, M.; Lefeuvre, C.; Bonnefoi, H.; Dalenc, F.; Jacquet, A.; De Filippo, M. R.; Babbar, N.; Birnbaum, D.; Filleron, T.; Le Tourneau, C.; André, F. Genomic Characterization of Metastatic Breast Cancers. *Nature* **2019,** *569* (7757), 560-564.

(6) Podo, F.; Buydens, L. M. C.; Degani, H.; Hilhorst, R.; Klipp, E.; Gribbestad, I. S.; Van Huffel, S.; van Laarhoven, H. W. M.; Luts, J.; Monleon, D.; Postma, G. J.; Schneiderhan-Marra, N.; Santoro, F.; Wouters, H.; Russnes, H. G.; Sørlie, T.; Tagliabue, E.; Børresen-Dale, A.-L.; Consortium, F. Triple-Negative Breast Cancer: Present Challenges and New Perspectives. *Molecular oncology* **2010,** *4* (3), 209-229.

(7) Foulkes, W. D.; Smith, I. E.; Reis-Filho, J. S. Triple-Negative Breast Cancer. *New England Journal of Medicine* **2010,** *363* (20), 1938-1948.

(8) Shu, D.; Li, H.; Shu, Y.; Xiong, G.; Carson, W. E.; Haque, F.; Xu, R.; Guo, P. Systemic Delivery of Anti-Mirna for Suppression of Triple Negative Breast Cancer Utilizing Rna Nanotechnology. *ACS Nano* **2015,** *9* (10), 9731-9740.

(9) Guo, P.; Huang, J.; Wang, L.; Jia, D.; Yang, J.; Dillon, D. A.; Zurakowski, D.; Mao, H.; Moses, M. A.; Auguste, D. T. Icam-1 as a Molecular Target for Triple Negative Breast Cancer. *Proceedings of the National Academy of Sciences* **2014,** *111* (41), 14710-14715.

(10) Karaayvaz, M.; Cristea, S.; Gillespie, S. M.; Patel, A. P.; Mylvaganam, R.; Luo, C. C.; Specht, M. C.; Bernstein, B. E.; Michor, F.; Ellisen, L. W. Unravelling Subclonal Heterogeneity and Aggressive Disease States in Tnbc through Single-Cell Rna-Seq. *Nature Communications* **2018,** *9* (1), 3588.

(11) Hudis, C. A.; Gianni, L. Triple-Negative Breast Cancer: An Unmet Medical Need. *Oncologist* **2011,** *16 Suppl 1*, 1-11.

(12) Kassam, F.; Enright, K.; Dent, R.; Dranitsaris, G.; Myers, J.; Flynn, C.; Fralick, M.; Kumar, R.; Clemons, M. Survival Outcomes for Patients with Metastatic Triple-Negative Breast Cancer: Implications for Clinical Practice and Trial Design. *Clin Breast Cancer* **2009,** *9* (1), 29-33.

(13) Bobo, D.; Robinson, K. J.; Islam, J.; Thurecht, K. J.; Corrie, S. R. Nanoparticle-Based Medicines: A Review of Fda-Approved Materials and Clinical Trials to Date. *Pharm Res* **2016,** *33* (10), 2373-2387.
(14) Mosquera, J.; García, I.; Liz-Marzán, L. M. Cellular Uptake of Nanoparticles Versus Small Molecules: A Matter of Size. *Accounts of Chemical Research* **2018,** *51* (9), 2305-2313.
(15) Ho, D.; Wang, P.; Kee, T. Artificial Intelligence in Nanomedicine. *Nanoscale Horizons* **2019,** *4* (2), 365-377.
(16) Adir, O.; Poley, M.; Chen, G.; Froim, S.; Krinsky, N.; Shklover, J.; Shainsky-Roitman, J.; Lammers, T.; Schroeder, A. Integrating Artificial Intelligence and Nanotechnology for Precision Cancer Medicine. *Advanced Materials 0* (0), 1901989.
(17) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Advanced Materials 0* (0), 1902765.
(18) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015,** *521*, 436.
(19) Norgeot, B.; Glicksberg, B. S.; Butte, A. J. A Call for Deep-Learning Healthcare. *Nature Medicine* **2019,** *25* (1), 14-15.
(20) Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A Guide to Deep Learning in Healthcare. *Nature Medicine* **2019,** *25* (1), 24-29.
(21) Copp, S. M.; Bogdanov, P.; Debord, M.; Singh, A.; Gwinn, E. Base Motif Recognition and Design of DNA Templates for Fluorescent Silver Clusters by Machine Learning. *Advanced Materials* **2016,** *28* (16), 3043-3043.
(22) Labouta, H. I.; Asgarian, N.; Rinker, K.; Cramb, D. T. Meta-Analysis of Nanoparticle Cytotoxicity Via Data-Mining the Literature. *ACS Nano* **2019,** *13* (2), 1583-1594.
(23) He, M.; Zhang, J.; Wang, H.; Kong, Y.; Xiao, Y.; Xu, W. Material and Optical Properties of Fluorescent Carbon Quantum Dots Fabricated from Lemon Juice Via Hydrothermal Reaction. *Nanoscale Research Letters* **2018,** *13* (1), 175.
(24) Zhang, Z.; Zhang, D.; Shi, C.; Liu, W.; Chen, L.; Miao, Y.; Diwu, J.; Li, J.; Wang, S. 3,4-Hydroxypyridinone-Modified Carbon Quantum Dot as a Highly Sensitive and Selective Fluorescent Probe for the Rapid Detection of Uranyl Ions. *Environmental Science: Nano* **2019,** *6* (5), 1457-1465.
(25) Misra, S. K.; Srivastava, I.; Tripathi, I.; Daza, E.; Ostadhossein, F.; Pan, D. Macromolecularly "Caged" Carbon Nanoparticles for Intracellular Trafficking Via Switchable Photoluminescence. *Journal of the American Chemical Society* **2017,** *139* (5), 1746-1749.
(26) Srivastava, I.; Misra, S. K.; Tripathi, I.; Schwartz-Duval, A.; Pan, D. In Situ Time-Dependent and Progressive Oxidation of Reduced State Functionalities at the Nanoscale of Carbon Nanoparticles for Polarity-Driven Multiscale near-Infrared Imaging. *Advanced Biosystems* **2018,** *2* (3), 1800009.
(27) Basu, N.; Mandal, D. Time-Resolved Photoluminescence of Ph-Sensitive Carbon Dots. *Carbon* **2019,** *144*, 500-508.
(28) Cao, X.; Wang, J.; Deng, W.; Chen, J.; Wang, Y.; Zhou, J.; Du, P.; Xu, W.; Wang, Q.; Wang, Q.; Yu, Q.; Spector, M.; Yu, J.; Xu, X. Photoluminescent Cationic Carbon Dots as Efficient Non-Viral Delivery of Plasmid Sox9 and Chondrogenesis of Fibroblasts. *Scientific Reports* **2018,** *8* (1), 7057.
(29) Misra, S. K.; Chang, H.-H.; Mukherjee, P.; Tiwari, S.; Ohoka, A.; Pan, D. Regulating Biocompatibility of Carbon Spheres Via Defined Nanoscale Chemistry and a Careful Selection of Surface Functionalities. *Scientific Reports* **2015,** *5*, 14986.
(30) Thoo, L.; Fahmi, M. Z.; Zulkipli, I. N.; Keasberry, N.; Idris, A. Interaction and Cellular Uptake of Surface-Modified Carbon Dot Nanoparticles by J774.1 Macrophages. *Central-European journal of immunology* **2017,** *42* (3), 324-330.
(31) Song, E.; Gaudin, A.; King, A. R.; Seo, Y.-E.; Suh, H.-W.; Deng, Y.; Cui, J.; Tietjen, G. T.; Huttner, A.; Saltzman, W. M. Surface Chemistry Governs Cellular Tropism of Nanoparticles in the Brain. *Nature Communications* **2017,** *8*, 15322.
(32) Alafeef, M.; Dighe, K.; Pan, D. Label-Free Pathogen Detection Based on Yttrium-Doped Carbon Nanoparticles up to Single-Cell Resolution. *ACS Applied Materials & Interfaces* **2019,** *11* (46), 42943-42955.
(33) Baker, S. N.; Baker, G. A. Luminescent Carbon Nanodots: Emergent Nanolights. *Angewandte Chemie International Edition* **2010,** *49* (38), 6726-6744.
(34) Misra, S. K.; Chang, H.-H.; Mukherjee, P.; Tiwari, S.; Ohoka, A.; Pan, D. Regulating Biocompatibility of Carbon Spheres Via Defined Nanoscale Chemistry and a Careful Selection of Surface Functionalities. *Scientific Reports* **2015,** *5* (1), 14986.
(35) Srivastava, I.; Sar, D.; Mukherjee, P.; Schwartz-Duval, A. S.; Huang, Z.; Jaramillo, C.; Civantos, A.; Tripathi, I.; Allain, J. P.; Bhargava, R.; Pan, D. Enzyme-Catalysed Biodegradation of Carbon Dots Follows Sequential Oxidation in a Time Dependent Manner. *Nanoscale* **2019,** *11* (17), 8226-8236.
(36) Pandit, S.; Banerjee, T.; Srivastava, I.; Nie, S.; Pan, D. Machine Learning-Assisted Array-Based Biomolecular Sensing Using Surface-Functionalized Carbon Dots. *ACS Sensors* **2019,** *4* (10), 2730-2737.
(37) Vercauteren, D.; Vandenbroucke, R. E.; Jones, A. T.; Rejman, J.; Demeester, J.; De Smedt, S. C.; Sanders, N. N.; Braeckmans, K. The Use of Inhibitors to Study Endocytic Pathways of Gene Carriers: Optimization and Pitfalls. *Molecular Therapy* **2010,** *18* (3), 561-569.
(38) Srivastava, I.; Misra, S. K.; Ostadhossein, F.; Daza, E.; Singh, J.; Pan, D. Surface Chemistry of Carbon Nanoparticles Functionally Select Their Uptake in Various Stages of Cancer Cells. *Nano Research* **2017,** *10* (10), 3269-3284.
(39) Wang, L. H.; Rothberg, K. G.; Anderson, R. G. Mis-Assembly of Clathrin Lattices on Endosomes Reveals a Regulatory Switch for Coated Pit Formation. *J Cell Biol* **1993,** *123* (5), 1107-1117.
(40) Kirchhausen, T.; Macia, E.; Pelish, H. E. Use of Dynasore, the Small Molecule Inhibitor of Dynamin, in the Regulation of Endocytosis. *Methods Enzymol* **2008,** *438*, 77-93.
(41) Qian, Z. M.; Li, H.; Sun, H.; Ho, K. Targeted Drug Delivery Via the Transferrin Receptor-Mediated Endocytosis Pathway. *Pharmacological Reviews* **2002,** *54* (4), 561-587.
(42) Saha, K.; Kim, S. T.; Yan, B.; Miranda, O. R.; Alfonso, F. S.; Shlosman, D.; Rotello, V. M. Surface Functionality of Nanoparticles Determines Cellular Uptake Mechanisms in Mammalian Cells. *Small (Weinheim an der Bergstrasse, Germany)* **2013,** *9* (2), 300-305.
(43) Gliem, M.; Heupel, W. M.; Spindler, V.; Harms, G. S.; Waschke, J. Actin Reorganization Contributes to Loss of Cell Adhesion in Pemphigus Vulgaris. *Am J Physiol Cell Physiol* **2010,** *299* (3), C606-613.
(44) Papakonstanti, E. A.; Stournaras, C. Cell Responses Regulated by Early Reorganization of Actin Cytoskeleton. *FEBS Letters* **2008,** *582* (14), 2120-2127.
(45) Allen, C.; Yu, Y.; Eisenberg, A.; Maysinger, D. Cellular Internalization of Pcl20-B-Peo44 Block Copolymer Micelles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1999,** *1421* (1), 32-38.
(46) Kampert, T.; Misra, S. K.; Srivastava, I.; Tripathi, I.; Pan, D. Phenotypically Screened Carbon Nanoparticles for Enhanced Combinatorial Therapy in Triple Negative Breast Cancer. *Cellular and Molecular Bioengineering* **2017,** *10* (5), 371-386.
(47) Pan, D.; Cai, X.; Yalaz, C.; Senpan, A.; Omanakuttan, K.; Wickline, S. A.; Wang, L. V.; Lanza, G. M. Photoacoustic Sentinel Lymph Node Imaging with Self-Assembled Copper Neodecanoate Nanoparticles. *ACS nano* **2012,** *6* (2), 1260-1267.
(48) Sigismund, S.; Woelk, T.; Puri, C.; Maspero, E.; Tacchetti, C.; Transidico, P.; Di Fiore, P. P.; Polo, S. Clathrin-Independent Endocytosis of Ubiquitinated Cargos. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (8), 2760-2765.
(49) Raghu, H.; Sodadasu, P. K.; Malla, R. R.; Gondi, C. S.; Estes, N.; Rao, J. S. Localization of Upar and Mmp-9 in Lipid Rafts Is Critical for Migration, Invasion and Angiogenesis in Human Breast Cancer Cells. *BMC Cancer* **2010,** *10*, 647.
(50) Misra, S. K.; Ostadhossein, F.; Daza, E.; Johnson, E. V.; Pan, D. Hyperspectral Imaging Offers Visual and Quantitative Evidence of Drug Release from Zwitterionic-Phospholipid-Nanocarbon When Concurrently Tracked in 3d Intracellular Space. *Advanced Functional Materials* **2016,** *26* (44), 8031-8041.
(51) Vranic, S.; Boggetto, N.; Contremoulins, V.; Mornet, S.; Reinhardt, N.; Marano, F.; Baeza-Squiban, A.; Boland, S. Deciphering the Mechanisms of Cellular Uptake of Engineered Nanoparticles by Accurate Evaluation of Internalization Using Imaging Flow Cytometry. *Particle and fibre toxicology* **2013,** *10*, 2-2.
(52) Park, J.; Ha, M. K.; Yang, N.; Yoon, T. H. Flow Cytometry-Based Quantification of Cellular Au Nanoparticles. *Analytical Chemistry* **2017,** *89* (4), 2449-2456.
(53) Maiorano, G.; Sabella, S.; Sorce, B.; Brunetti, V.; Malvindi, M. A.; Cingolani, R.; Pompa, P. P. Effects of Cell Culture Media on the Dynamic Formation of Protein−Nanoparticle Complexes and Influence on the Cellular Response. *ACS Nano* **2010,** *4* (12), 7481-7491.
(54) Shannahan, J. H.; Brown, J. M.; Chen, R.; Ke, P. C.; Lai, X.; Mitra, S.; Witzmann, F. A. Comparison of Nanotube–Protein Corona Composition in Cell Culture Media. *Small* **2013,** *9* (12), 2171-2181.

(55) Pan, D.; Pramanik, M.; Senpan, A.; Ghosh, S.; Wickline, S. A.; Wang, L. V.; Lanza, G. M. Near Infrared Photoacoustic Detection of Sentinel Lymph Nodes with Gold Nanobeacons. *Biomaterials* **2010,** *31* (14), 4088-4093.
(56) Martin, T. M.; Harten, P.; Venkatapathy, R.; Das, S.; Young, D. M. A Hierarchical Clustering Methodology for the Estimation of Toxicity. *Toxicol Mech Methods* **2008,** *18* (2-3), 251-266.
(57) Oksel, C.; Ma, C. Y.; Wang, X. Z. Current Situation on the Availability of Nanostructure–Biological Activity Data. *SAR and QSAR in Environmental Research* **2015,** *26* (2), 79-94.
(58) Wang, W.; Sedykh, A.; Sun, H.; Zhao, L.; Russo, D. P.; Zhou, H.; Yan, B.; Zhu, H. Predicting Nano–Bio Interactions by Integrating Nanoparticle Libraries and Quantitative Nanostructure Activity Relationship Modeling. *ACS Nano* **2017,** *11* (12), 12641-12649.
(59) Triguero, I.; García-Gil, D.; Maillo, J.; Luengo, J.; García, S.; Herrera, F. Transforming Big Data into Smart Data: An Insight on the Use of the K-Nearest Neighbors Algorithm to Obtain Quality Data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2019,** *9* (2), e1289.
(60) Prasad, S.; Mann Bruce, L. Limitations of Principal Components Analysis for Hyperspectral Target Recognition. *Geoscience and Remote Sensing Letters, IEEE* **2008,** *5*, 625-629.
(61) Krogh, A. What Are Artificial Neural Networks? *Nature Biotechnology* **2008,** *26*, 195.
(62) Zhang, Q.; Yu, H.; Barbiero, M.; Wang, B.; Gu, M. Artificial Neural Networks Enabled by Nanophotonics. *Light: Science & Applications* **2019,** *8* (1), 42.
(63) Hadoush, H.; Alafeef, M.; Abdulhay, E. Automated Identification for Autism Severity Level: Eeg Analysis Using Empirical Mode Decomposition and Second Order Difference Plot. *Behavioural Brain Research* **2019,** *362*, 240-248.
(64) Abdulhay, E.; Alafeef, M.; Alzghoul, L.; Al Momani, M.; Al Abdi, R.; Arunkumar, N.; Munoz, R.; de Albuquerque, V. H. C. Computer-Aided Autism Diagnosis Via Second-Order Difference Plot Area Applied to Eeg Empirical Mode Decomposition. *Neural Computing and Applications* **2018**,
(65) Alafeef, M.; Fraiwan, M.; Alkhalaf, H.; Audat, Z. Shannon Entropy and Fuzzy C-Means Weighting for Ai-Based Diagnosis of Vertebral Column Diseases. *Journal of Ambient Intelligence and Humanized Computing* **2019**,
(66) Ye, W.; Chen, C.; Wang, Z.; Chu, I.-H.; Ong, S. P. Deep Neural Networks for Accurate Predictions of Crystal Stability. *Nature Communications* **2018,** *9* (1), 3800.
(67) Sayes, C.; Ivanov, I. Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis* **2010,** *30* (11), 1723-1734.
(68) Winkler, D. A.; Burden, F. R.; Yan, B.; Weissleder, R.; Tassa, C.; Shaw, S.; Epa, V. C. Modelling and Predicting the Biological Effects of Nanomaterials. *SAR and QSAR in Environmental Research* **2014,** *25* (2), 161-172.
(69) Zhang, H.; Mardyani, S.; Chan, W. C.; Kumacheva, E. Design of Biocompatible Chitosan Microgels for Targeted Ph-Mediated Intracellular Release of Cancer Therapeutics. *Biomacromolecules* **2006,** *7* (5), 1568-1572.
(70) Schuurmann, G.; Ebert, R. U.; Chen, J.; Wang, B.; Kuhne, R. External Validation and Prediction Employing the Predictive Squared Correlation Coefficient Test Set Activity Mean Vs Training Set Activity Mean. *J Chem Inf Model* **2008,** *48* (11), 2140-2145.
(71) Golbraikh, A.; Tropsha, A. Predictive Qsar Modeling Based on Diversity Sampling of Experimental Datasets for the Training and Test Set Selection. *J Comput Aided Mol Des* **2002,** *16* (5-6), 357-369.
(72) Tabe-Bordbar, S.; Emad, A.; Zhao, S. D.; Sinha, S. A Closer Look at Cross-Validation for Assessing the Accuracy of Gene Regulatory Networks and Models. *Scientific Reports* **2018,** *8* (1), 6620.
(73) Ma, W.; Cheng, F.; Liu, Y. Deep-Learning-Enabled on-Demand Design of Chiral Metamaterials. *ACS Nano* **2018,** *12* (6), 6326-6334.
(74) Alafeef, M.; Fraiwan, M. On the Diagnosis of Idiopathic Parkinson's Disease Using Continuous Wavelet Transform Complex Plot. *Journal of Ambient Intelligence and Humanized Computing* **2018**,
(75) Valeh-e-Sheyda, P.; Yaripour, F.; Moradi, G.; Saber, M. Application of Artificial Neural Networks for Estimation of the Reaction Rate in Methanol Dehydration. *Industrial & Engineering Chemistry Research* **2010,** *49* (10), 4620-4626.
(76) Titano, J. J.; Badgeley, M.; Schefflein, J.; Pain, M.; Su, A.; Cai, M.; Swinburne, N.; Zech, J.; Kim, J.; Bederson, J.; Mocco, J.; Drayer, B.; Lehar, J.; Cho, S.; Costa, A.; Oermann, E. K. Automated Deep-Neural-Network Surveillance of Cranial Images for Acute Neurologic Events. *Nature Medicine* **2018,** *24* (9), 1337-1341.
(77) Gade, A. M.; Meadows, M. K.; Ellington, A. D.; Anslyn, E. V. Differential Array Sensing for Cancer Cell Classification and Novelty Detection. *Organic & Biomolecular Chemistry* **2017,** *15* (46), 9866-9874.
(78) Bajaj, A.; Rana, S.; Miranda, O. R.; Yawe, J. C.; Jerry, D. J.; Bunz, U. H. F.; Rotello, V. M. Cell Surface-Based Differentiation of Cell Types and Cancer States Using a Gold Nanoparticle-Gfp Based Sensing Array. *Chemical Science* **2010,** *1* (1), 134-138.
(79) Mei, X.; Lee, H.-C.; Diao, K.-y.; Huang, M.; Lin, B.; Liu, C.; Xie, Z.; Ma, Y.; Robson, P. M.; Chung, M.; Bernheim, A.; Mani, V.; Calcagno, C.; Li, K.; Li, S.; Shan, H.; Lv, J.; Zhao, T.; Xia, J.; Long, Q.; Steinberger, S.; Jacobi, A.; Deyer, T.; Luksza, M.; Liu, F.; Little, B. P.; Fayad, Z. A.; Yang, Y. Artificial Intelligence–Enabled Rapid Diagnosis of Patients with Covid-19. *Nature Medicine* **2020**,
(80) Srivastava, I.; Khan, M. S.; Dighe, K.; Alafeef, M.; Wang, Z.; Banerjee, T.; Ghonge, T.; Grove, L. M.; Bashir, R.; Pan, D. On-Chip Electrical Monitoring of Real-Time "Soft" and "Hard" Protein Corona Formation on Carbon Nanoparticles. *Small Methods n/a* (n/a), 2000099.

Table of Contents artwork