

Meta-analysis of Cellular Toxicity for Graphene via Data-Mining the Literature and Machine Learning

Ying Ma

Southeast University

Jianli Wang

Southeast University

Jingying Wu

Southeast University

Chuxuan Tong

The University of Queensland

Ting Zhang (✉ zhangting@seu.edu.cn)

Southeast University

Research

Keywords: graphene, cytotoxicity, meta-analysis, random forests, machine learning

Posted Date: February 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-248207/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Science of The Total Environment on November 1st, 2021. See the published version at <https://doi.org/10.1016/j.scitotenv.2021.148532>.

Abstract

Background: Due to graphene is currently incorporated into various consumer product and numerous new applications, determining the relationships between physicochemical properties of graphene and their toxicity is a prominent concern for environmental and health risk analysis. Data from the literatures suggested that graphene exposure may resulted in cytotoxicity, however, the toxicity data of graphene is still insufficient to point out its side because of the complexity and heterogeneity of available data on potential risks of graphene.

Methods and Results: Here, we developed a meta-analysis approach for assembling published evidence on cytotoxicity based on 792 related publications, 986 cell survival rate samples, 762 IC₅₀ samples, and 100 LDH release samples. In this study, among corresponding attributes, we proved that the cytotoxicity of graphene assessed in the form of cell viability, IC₅₀ and LDH can be primarily predicted from exposure dose and detection method, diameter and surface modification, detection method and organ source, respectively. Furthermore, this paper provides guidance regarding three optional data sets for above-mentioned three endpoints that are chiefly related to cellular toxicity for future studies and cross-validation studies based on machine learning tools including Random Forests (RFs), Support Vector Machine (SVM), LASSO regression, and Elastic Net were conducted for result verification.

Conclusions: In summary, our study indicates that following rigorous methodological experimental and extract approaches accompanied with suitable machine learning tools, in parallel to continuous addition to reliable data set developed using our meta-analysis approach, will offer higher predictive power and accuracy, and also help to provide effective information on designing safe graphene.

1 Introduction

Graphene is a monocrystal and monolayer consisted of carbon atoms, sp² hybridized orbital closely arranged into two-dimensional honeycomb lattice nanometer materials¹⁻³. It is divided into oxidized graphene, reduced oxidized graphene, graphene nanoribbon, and graphene quantum dot^{2,4}. Graphene owns unique electronic and mechanical properties, saturated magnetization, excellent charge carrier mobility, high thermal conductivity, large specific surface area, extraordinary electrocatalytic activity, and good biocompatibility⁵, which makes it widely applied in electrochemical device, energy storage, nanomedicine, biotechnology, photothermal therapy, and tissue engineering⁶. However, with the widespread use of graphene and other nanomaterials, people are exposed to different degrees of graphene in the synthesis, manufacture, use, treatment of graphene. Therefore, the biosafety study of graphene materials is essential and necessary. Although the distinctive advantages and potential toxicity of graphene for bioapplications have been confirmed repeatedly, there is still no moderate consensus on what features are most important and necessary among these numerous attributes for experiment conduction with variable outcomes. Furthermore, the vast majority is performed using several graphene materials in limited experimental models and a globalized model relating properties to graphene cytotoxicity is crucial.

So far, a systematic evaluation of the toxicity of graphene based on in vitro cell experiments is lacking. On one hand, graphene for toxicity studies has different sources, mainly including commercialization and preparation by different laboratories, thus the toxicity data obtained by them is lacking of comparability. On the other hand, a variety of toxicity studies involves exposure time, exposed cell types and detection methods are also different, which results in uncertainties and variability in graphene toxicity studies and therefore it's difficult to generate the large amounts of information from individual studies. It is worth noting that the successful application of meta-analytic or literature data mining in nanotoxicological studies provide the possibility to summarize different studies. The advantages of meta-analysis include combining the results of different studies on the same topic⁷, summarizing current evidence on the specific issue objectively⁸, resolving conflicts between studies⁹, and yielding conclusive results when individual studies are inconclusive¹⁰. Previous studies conducted on Meta-Analysis of the relationships between physicochemical properties of nanoparticles and their toxicity endpoints. To be specific, Gernand¹¹ et al, developed a suite of machine-learning based methods for meta-analysis of nanotoxicity researches of 136 carbon nanotubes from 17 publications examined this association between pulmonary toxicity and 41 CNT attributes. The work identified that

CNT attributes including metallic impurities, CNT length and diameter and aggregate size contributed to pulmonary toxicity. Similar studies have been achieved exemplified by Eunkeu Oh¹² to predict the toxicity of quantum dot using random forest regression models and evaluate primary attribute including shell, ligand and surface modifications, diameter, assay type and exposure time which contribute to toxicity. In addition, Hagar I. Labouta¹³ et al provided guided prediction information with key attributes regarding NP cytotoxicity risk analysis using machine learning tools added with published cytotoxicity data. Rather than the obvious toxicity profiled such as viability compromised cell membrane integrity, meta-analysis allows to reveal hidden relationships that were not targeted in the primitive publications¹⁴.

Another challenging question is that of the physicochemical properties graphene, the types of cell models and experimental factors contributed with cytotoxicity. Among physicochemical properties, size has been proved particular significance in affecting graphene toxicity regardless of conflicting conclusions^{6, 15, 16}. Furthermore, the oxidation state plays an important role in graphene toxicity^{15, 17}. Also, surface modification and surface charge¹⁸ have been observed to contribute to graphene toxicity in a few papers. So do organs¹⁹, types and morphology²⁰ of cells. Most researches indicated that with the increase of graphene exposure dose²¹ and time²², the viability of the treated cells decreased continuously, manifesting a toxicity change from low to high. Nevertheless, low-dose graphene may also have a hormesis effect²³. There is also a marked concern in understanding different detection methods on the cytotoxicity of graphene^{24, 25}. Specially, meta-analysis approach or data-mining was conducted by extracting data from the literature, it is possible to reveal the relationship among characteristics of NP, the influencing factors and cytotoxic effects. It can be concluded that the strategies proposed to determine the properties of NPs leading to observed cellular effects.

It is worth mentioning that the machine learning tools were used for this evaluation as they perfectly match with the study and type of data. Based on the machine learning tools including Random Forests (RFs), Support Vector Machine (SVM), LASSO regression, and Elastic Net, these tools have been applied in various fields extensively, have important statistical features and estimated the correlation of these characteristics. Accordingly, the meta-analysis utilized the above-mentioned machine learning tools to identify relevant attributes, evaluate the consistency of graphene toxicity data according to published literature, and develop powerful prediction models (optional data collection) of graphene cell toxicity, which bears an active meaning and great value for reference over future graphene toxicity researches.

In this study, a meta-analysis approach has been developed to generalize reported literature data to provide a modeling framework of potential graphene toxicity. Considering the uncontrollable and variability characteristics of data in various researches aimed at improving predictability, we followed a series of strict inclusion criteria of a study by “*PICOS*” search strategy. Totally 2896 corresponding primitive research reports were examined, from which 792 ones were selected carefully. Of research indicators with relative widespread concern, 10 attributes (that are diameter, surface modification, oxidation state, exposure dose, exposure time, detection method, organ source, cell morph, cell source and cell line, totally) were ultimately selected in this study of per contribution to different cytotoxic endpoints. The paper provides a comprehensive collection of data on a large spectrum of graphene practically.

2 Method

Standards on inclusion and exclusion have been developed. following the PICOS (Patient, intervention, comparison, outcome, study design) model of evidence-based medicine, which is a useful tool for asking focused clinical questions was endowed with specific connotations as P- cell, I- Graphene and its derivatives, C- none, O- cytotoxicity, S- in vivo study based on the objective of exploring cellular toxicity for graphene in the study then the corresponding search term was set. We adjusted search terms in the Medical Subject Headings as well as use logical operators to group synonyms. Using Boolean operators so as to form the search strategy and further retrieved 792 literatures. We added synonyms or abbreviations that appropriate to the search terms (That is, (Graphite[MeSH Terms]) OR graphite[Title/Abstract]) OR graphene[Title/Abstract]) AND (in vitro[MeSH Terms]) OR in vitro) OR cell experiments) OR cell experiment) OR cell test) OR cell tests) OR in vitro experiment) OR in vitro test) OR in vitro tests)) AND (toxicity[MeSH Terms]) OR toxicity) OR cytotoxicity) OR toxic) OR toxicology) OR cellular toxicity), ultimately). Also, there were no restrictions on year of publication, with databases searched back to their inception. In

the original research, after excluding repeated literatures using EndNote X9 (Bld12062, Thompson Reuter, CA), depending on the title and abstract assessment, some improper literatures were excluded. The full texts of the remaining papers were downloaded subsequently and evaluated carefully for predetermined parameter define and meanwhile data were extracted from each publication meeting the inclusion criteria. Prior to the extraction experiment, relevant graphene attributes and experimental conditions that may influence experiment results were defined and divided into different sets. Moreover, the data were extracted from included article tools or displayed graphically (plotting and histogram diagram) by a program called 'GetData graph digitizer'. In order to compile machine-learning demanded data, data were coordinated in units of the graphene diameter (nm), and exposure dose ($\mu\text{g/ml}$), as well as exposure time (h) at al. Descriptive indexes were additionally added for the graphene, cells and the assay method at.al. The collected data were then processed to normalize numeric attributes and to delete incomplete data. It is noteworthy that we afterwards performed model establishment and development with standardized data after data extraction was independently conducted using a normalized approach.

Regression models of CV, IC_{50} and LDH release were established by the Random Forests tool based on literature compiled data processed for analysis. Further, the most appropriate attribute sets via all-combination search of cell viability were determined to prepare for RF model development. In this research, decision trees and Random Forests learning algorithm were selected specially given that they can put forth their strength to process complex and heterogeneous graphene cytotoxicity data which had been compiled and predict the value of target variables (in this, cell viability, IC_{50} and LDH release) while assessing model performance and attribute importance meanwhile. RF model tend to conduct process as follows, including at first, sample extraction from the original data, and then selecting random attribute sets of each tree to create a decision tree, finally repeating the above process until the prediction. In addition, the significance of per attribute was assessed by a random recombination framework as well as RF model development. Due to that the data set approach does not take into account the interdependence and complementarity of information that may be generated in several areas, such combination of attribute data sets may not necessarily be the optimal ones, though in which top-ranked attributes were identified distinctly. Therefore, an all-combination attribute search was conducted under RF model development. Since there are only 10 attributes ultimately after data screening, you can easily confirm the most suitable viable set. The RF model prediction process for graphene cytotoxicity was in association with the development one. This document adopted the R^2 as determinant coefficient ($1 - \text{MAE}/\text{Var}(y)$) for evaluating accuracy of OOB prediction, where y demonstrates the observed terminal point (cell viability, IC_{50} or LDH release value) and MAE implies the mean absolute error of the model prediction). Finally, cross-validated approach was further conducted for evaluating the accuracy of models using these machine learning tools.

3 Results And Discussion

3.1 General workflow of data acquisition and processing

Meta-analysis of graphene cytotoxicity data carried out following the workflow displayed in Figure 1. The topic was determined by listing search terms and further retrieved 792 literatures. Those available data is valuable to different degree thus orderly screening and selecting was conduct (SI Table 1). The second screening was carried out Subsequently after the first screening was completed, which aimed at data collection was conducted as the remaining articles was carefully checked again based on refining exclusion principle (SI Table 1). After data collection and normalization, attributes extracted from the literature describing graphene cytotoxicity including physicochemical properties,, cell- and experiment-related conditions at.al were selected and listed (Table 1). Subsequently, cell viability, IC_{50} , and LDH as three indicators were characted and processed, based on which importance of macroscopic properties was presented in the form of heatmap (Figure 3). As per attribute included subvariant ones in Figure 4, the subvariant matched each set once again, whether subvariant variable subject to single attribute can have significant effects on these indexes or not was analyzed.

3.2 Cell viability model

Cell viability is an important parameter for toxicity studies of nanomaterial and tend to estimate cell metabolic level indicating growth potential²⁶. To data, the majority of studies have merely focused on material itself, such as the viability detection

methods²⁷ and oxidation state^{15,17}. From 57 publications, we obtained 986 cell viability-related data samples, each with attributes describing the material properties and experimental conditions by applying strict inclusion and exclusion criteria.

Use of cell viability models allows us to evaluate and predict effect of graphene on cell viability. Thus, RFs and Linear regression models (SI Figure 1, SI Figure 2) on cell viability were developed based on 10 features (selected among totally 18 attributes as part ones missing mass data removed) including specified nanoparticle-, cell-, and screening method-related ones identified from set by compiled literature (Table 1). Besides, parallel coordinate compiled for cell viability in collection of the appointed 10 attributed is shown from Figure 2. Figure 3 demonstrated that exposure dose ($R^2=0.331$), detection method ($R^2=0.154$) and organ source ($R^2=0.101$) had most significant effects on CV. As was confirmed in previous studies, the exposure dose contributed to the toxicity of GO most significantly²⁸⁻³⁰. In our study, $R^2 \geq 0.02$ is indicative of good feature importance on CV. We find that experimental condition attributes, such as exposure dose, diameter (especially 100-1000 nm), exposure time (12h, 24h and 48h) and detection method (e.g. neutral red assay and resazurin) have significant effect on CV (Figure 4). In previous researched, resazurin was a redox index for cell permeability determine. This helps to track the number of living cells and provides a stable quantitative assessment of fluorescence intensity and cell number.^{31,32}. Many assays support this view as they detected cell viability in embryonic and cancer cells through quantitative resazurin-based assay given that resazurin is more sensitive and can be used for multiple tests of all dyes in spite of that the oxidation and reduction reactions of many dyes to live cell membranes reflect metabolic activities.³³. Additionally, the neutral red assay permits evaluating cell viability under acidic pH or hypoxic conditions more accurately than other cell viability assay—which were reported in a few studies³⁴. Furthermore, the most commonly used colorimetric assays e.g. MTT, which readily penetrates viable eukaryotic cells and MST and WST-1 etc. have limitations including inability to explain variation in cellular metabolism throughout the life cycle³⁵. However, MTT is still a important detection method for correlating cell viability (Figure 4). The diversity of 14 cell assays with different anatomical / biological performances may imply various toxic reactions. Regardless of that the dependence on different intrinsic properties of graphene and experimental conditions is a complex definition, the relationship between these above-mentioned attributes and toxicity degree with certain scopes over specific experimental conditions (e.g. exposure time and dose). In our study, cell morph especially epithelial, neuronal and monocyte-macrophage had feature importance with CV. For example, Hu et al.³⁶ cultured A549 cells with different concentrations of fetal bovine serum, and detected the effects of GO respectively, from which cytotoxicity of GO was obviously observed. Also, Monocyte-Macrophage following up with epithelial cells is of chief importance to cell ability of all Cell morph³⁷. As results show that organ source including bone, adipose and blood and surface modification—PEI have top importance ranking for toxicity (Figure 4). As is known to all, PEI with high vaccination rate promotes the adsorption of GOMN adsorbents for Cr³⁵. What is more, PEI may affect positive and negative charges then enable cells more sensitive.

Given as attributes were included in the model one by one by the correlation ranged from high to low (SI Figure 3), especially after took the oxidation state into model (Figure 5) the result was obtained, with R^2 decreased distinctly, indicated that the correlation between attributes and cell viability cannot adequately represented the importance of attributes. Thus, all combination research was adopted for obtaining the most important attribute and their combination to cell viability (Table 2). Accordingly, the order in which specific attributes were added coincides with their significance for the correlation of graphene toxicity and each attribute's contribution to ending points was calculated to prove per effect on CV via the all-combine research. Hereinafter, scrambled sequence of each attribute were considered to be arranged by employing python with optional prediction accuracy (Table 2) and for the cases, the cell viability RF regression model depending on the above-mentioned 9 attributes, showed good performance of $R^2 \approx 0.820$. When the above three attributes (2, 6 and 7, Table 1) were excluded from the set, the RF model for cell viability with the remaining 6 attributes (that is 1,4,5, and 8-10, Table 1) illustrated only 1.8% decrease in R^2 (from 0.820 to 0.805), which is in line with the above percentage-based individual attribute contribution index. Given the trade off between the desire for increased generalization capability and strict model accuracy, the 6-attribute set was of chief significance in future prediction studies with regard to CV.

The above-mentioned top six features were adopted into model prediction which with significant $R^2=0.805$ (data listed in Figure 6) and scatter representing the correlation of CV.

3.3 IC₅₀ model

In order to overview the available IC₅₀ data in huge sum for major researches, figure 7 was especially drawn to report toxicity effect over the scope of exposure concentrations.

IC₅₀ value is the most sensitive measurement value, and logarithmic IC₅₀ has become a common parameter presenting cellular toxicity analysis^{38,39}. Total of 169 distinct IC₅₀ values (thought as an intuitive acute and chronic toxicity index derived from graphene cell viability data integrated over experiments) were identified from 49 reviewed research articles (Figure 6). In view of the above mentioned facts, IC₅₀ values analysis was based on the same attribute set employed for cell viability (totally 10 attributes in Table 1). Top-ranked attributes of IC₅₀ were diameter ($R^2=0.534$), surface modification ($R^2=0.174$), oxidation state ($R^2=0.151$). As is shown in Fig. 3, graphene diameter (especially in the range of $10^{2.5}$ - 10^3), proved to affect cell morphology and planting density⁴⁰, further graphene interaction with cell membranes, so then shift graphene toxicity. From our results of IC₅₀ poly saccharides as a kind of more negatively charged polymer adhesion on graphene, increased the volume of graphene further changed the surface charge of graphene and its derivatives¹⁷, which could explain the highest IC₅₀ value of polysaccharide-modified graphene as surface modification is a relevant attribute in the IC₅₀ model. However, lower IC₅₀ value of unmodified graphene was seen as cells are directly exposed to the graphene environment, facilitating the graphene cellular uptake¹². It is noted that oxygen-containing functional groups play an important role in hydrophilicity, stability and membrane affinity of graphene, thus different oxidation states contribute different degrees to IC₅₀ value.

Among the experimental conditions, detection method like LDH, NRU and resazurin at al showed high contribution (Figure 4), which may be trace back to the mechanism of toxicity of graphene on cells such as oxidative stress^{21,25}. Also, epithelial cells and peripheral monocyte in cell morph are identified as relevant correlative sub-attributes for IC₅₀, account for that the subjects cells have better sensitivity to graphene and its derivatives. However, conditional dependence on the sub-attributes of exposure dose, cell line, organ source, exposure time and cell source were hardly observed (Figure 3).

The IC₅₀ RF model and predictive data (Figure 7) demonstrated better model prediction performance and less scatter corresponding to the correlation of CV based on attributes describing graphene physicochemical properties and experimental conditions.

3.4 LDH release model

The purpose of LDH release model development is to estimate the influence of graphene physicochemical properties on cytotoxicity particularly. From 13 publications, we finally obtained 100 LDH-related data samples, each with attributes in terms of graphene properties and experimental conditions by applying strict inclusion and exclusion criteria. Then, LDH release models were established on this basis.

Cytotoxicity was usually evaluated by the release of lactate dehydrogenase⁴¹ (LDH) as LDH is one of the enzymes in living cells and normal cells penetrate the cell membrane is almost impossible. The membrane permeability will change and LDH in cytoplasm will be released into the culture medium, further proved LDH is a sensitive final index to evaluate the toxicity of graphene. In terms of LDH release were detection method, organ source, exposure dose, demonstrating $R^2=0.352$, 0.351 and 0.095 , respectively (Figure 3). As is known to all, detection method of LDH release is often specific and limited. Of the organ source, breast (e.g. MCF10A and Hs578Bst) ranked the first (Figure 4), this assumption however has not been supported by the adequate literatures. Furthermore, some studies proved the LDH release of graphene for breast cells yet this issue is open and needs more consideration and more evidence to clarify.

We have hypothesized that graphene cellular toxicity could be predicted based on material parameters and cell-related attributes. Of the graphene-related ones, organ detection method surface modification exposure dose diameter and oxidation state ranked the feature importance of LDH release from high to low, but cell-related attributes (that is, cell-source and cell-line in this study) has done little to LDH release (Figure 3).

For each dataset, we picked 20% of the data as the test set for RF model predicted versus observed LDH release (Figure 6). The predicted LDH model was based on seven attributes (exposure time and cell line excluded), presenting a significantly great level with $R^2 \approx 0.986$.

3.5 The proposed machine learning tools

In the study, some literature graphene family data attributes mentioned above own incomplete data. In view of it, 10 complete attributes associated with end points (cell viability, extended IC_{50} , and LDH release quantity) are finally adopted. Based on Random Forests (RFs), Support Vector Machine (SVM), LASSO regression, and Elastic Net, which have been applied in various fields extensively, statistically important features, and the correlation of these characteristics were estimated. Moreover, we repeatedly use samples with connected cytotoxicity assays aimed at checking statistical models and validating predictions. As mentioned above, the machine learning tools can perfectly match with the study and type of data for this assessment.

Random Forests (RFs)⁴² are integration methods, generally employs the decision tree as a base for its study. They tend to integrate the decision tree by average out with approximate unbiased model of noise to reduce the variance. Specifically, random disturbance and forest use the data sample input disturbance attributes at the same time. RFs are adopted widely in bioinformatics research including RNA methylation, protein interaction prediction and so on⁴³. Besides, RF is characterized as adding data sample perturbations and input property perturbations during training to handle a variety of data types. In the existing medical image analysis, they are mainly used in image processing of medical images, diagnosis of assisted medical treatment, and exploration of the pathogenesis of certain diseases. Kesler⁴⁴ employed random forests to extract and classify factors related to cognitive impairment. While they play role in Chiang⁴⁵ TLE research, that determine which of the hippocampus input the TLE to the left or right, Kacar's⁴⁶ multiple sclerosis research and Koley etc.⁴⁷ proposed framework that based on RFs to implement automatic diagnostic of four types of brain tumors. Besides, Serag etc.⁴⁸ adopted automatic SEG Mentation algorithm for human brain MRI image.

Lasso regression⁴⁹ can make some characteristic index decreases, and even make some small absolute value of coefficient of direct to 0, so as to enhance the generalization ability of the model. To achieve the feature data, especially linear relationship is sparse, or to find out the main characteristics in a bunch of features, then L1 regularization (Lasso regression) is preferred. As is known, genome scanning can easily produce tens of thousands of variables, while there is little increase in samples of the medical research, leading to the number of samples smaller than it of variable conditions, for example the diabetes development prediction model.⁵⁰ In these cases, the Lasso regression model could improve the accuracy of prediction, and use variable screening to simplify the model. Differently, Elastic Network⁵¹, which has been widely used in the studies of protein structure-function relationship, in the case of a lot of features linked to each other is very useful⁵². Lasso probably only random consider one of these characteristics, while elastic network more inclined to select more than one. Accordingly, LASSO regression tends to choose any one of independent variables to join after the screening model if these variables are extremely associated, . As for the Elastic Net model ($0 < \alpha < 1$), the constraints between the square and round shape, so that its characteristic is in an argument or a group of randomly selected trade-offs among the independent variables.

SVM⁵³, a tool used to find the hyperplane that maximizes the margin between the samples in various classes⁵⁴. It can guarantee the extremal solutions, namely the global optimal solution rather than the local minimum. It also determines the SVM method has good generalization ability to the unknown sample. Are due to these advantages, SVM model has been widely used in various fields, specially pattern recognition⁵⁵ including face detection and recognition, handwritten-digits recognition, text classification, speaker speech recognition, image recognition and retrieval, etc.^{56, 57} Moreover, it shows better generalization ability. Therefore, the empirical risk can be calculated based on the hinge loss function of SVM. Also, in this robust and sparse classifier regularization item was added for better structural risk⁵⁸.

In the study, the developed four machine learning tools (i.e. RFs, SVM, LASSO and ENet) have been vastly applied in various fields especially toxicology. The tools use as features quantitative toxicological profiles including individual contribution of each attribute related indicators, and each accumulation get forecasts. Moreover, regression models developed for cell viability

as the toxicity index by all-combined search including both most suitable attributes, based on a framework, in which model and attributes were selected automatically. Thus, a comprehensive analysis of large indicator datasets, whether it is systematic or not, was permitted. Also, the most corrective attributes are assessed for end points (i.e. cell viability, IC₅₀ and LDH). For the sake of overcoming the challenge of integrating cytotoxicity data originating from various studies and focus on perfect predictive performance, strict criteria for inclusion was used in published literature pool, in which data was rigorously limited to those involving widely adopted cell viability, IC₅₀ and LDH release quantity.

With regard to applicability, we draw particular attention to random forests, which are the most accurate of the current algorithms. With our research, it can effectively run in large databases. In our research, it has an effective way to evaluate the missing data and ensure its accuracy. In the absence of a large number of data (such as diameter), it is mainly random forest, which variables are important to estimate the classification. As far as applicability concerned, we emphasize on Random Forests as it has the untouchable accuracy of current algorithms and runs efficiently on a large portion of data along with our study even if massive missing data such as diameter in our study exist. Chiefly, Random forests help estimate that what variables are important in the classification.

In the study, 10 types of characteristic parameters affecting IC₅₀, LDH and cell viability were chosen. Based on random forests, SVM, LASSO regression, Linear Regression and elastic net, statistically important features and the correlation of these characteristics were estimated. In addition to random forests models, we also analyzed the remained machine learning tools mentioned above. In the process of choice, the data was divided into the test set and validation set, with the effect of the MAE to test method validation set. Mean Absolute Error (MAE)⁵⁹ tested the effect of the proposed algorithm by calculating the average absolute difference between the predicted and real values of the project score. The smaller the value is, the better the proposed algorithm performs. Therefore, MAE value was selected as the important evaluation standard in this meta-analysis. Figure 8 shows a boxplot of the prediction performance for each machine learning tool. Note that Figure 8 is based only on a certain number of assays (that have an annotation) and that 20 sets of data. Figure 8 showed that the SVM model has the greatest MAE value among all models, followed by lasso and ElasticNet models in the test set. Also, all MAE values demonstrating prediction performances of lasso (alpha was selected as 0.05, 0.1 and 0.2 respectively) and ElasticNet (ratio=0.85, 0.5, 0.1) models in different cross-validation studies did not vary appreciably. For the unified SVM, the parameter values determined from training was as follows: C = 800, 1000, 10000, given that C is used to control the cost of misclassification in the SVM. Moreover, when C = 10000, 20 sets of MAE value data are relatively concentrated.

4 Conclusions

A meta-analysis approach has been employed in this study including 792 publications mine, along with concomitantly 986 graphene cytotoxicity data samples describing 10 qualitative and quantitative attributes such as graphene, cell properties and further the methodological settings, accordingly generate a comparatively comprehensive compiled data set for graphene in practice. Predictive graphene cytotoxicity models for cell viability, IC₅₀ and LDH were developed using Random Forests learning algorithm. The prediction models demonstrated an accuracy of R² =0.805, R² = 0.903, and R² =0.986 for cell viability, IC₅₀ and LDH, respectively. Moreover, RFs models assessed normalized data for graphene cellular toxicity adequately and precisely. Importance of macroscopic properties for correlating toxicity indexes were evaluated using RF models, as is demonstrated that random permutation is used to achieve the optimal combination of these corresponding attributes as an attribute combination of measurable significance in correlating toxicity. We also note that three data sets composed of the most significant attributes of cell viability, IC₅₀ and LDH respectively are instructive for future experiment design and conduction. For example, diameter, cell morph, cell source, detection method, exposure time, exposure dose are composed of the optional attribute set for graphene toxicity study in which CV was used as the index of toxicity. This provides more targeted study design and analysis. The way is also applicable to IC₅₀ and LDH indicator. All possible arrangements were taken to avoid unnecessary steps aimed at similar researches being both significant cost saving and more scientific and rational.

Overall, this paper implied that machine learning of literature data-compiling could guide and predict chief attributes (e.g. graphene properties and experimental conditions) related to graphene cytotoxicity that should be characterized and attached

importance to. In view of that large data sets are available, literature data collection was expected to help to formulate advanced regulatory policies for health and environment. The collection of documents will help to formulate sound health policies. Our paper offers an excellent example that the value of the published information and proper modeling approaches should be considered rigorously to pair graphene cytotoxicity with the relevant attributes, which will reveal further potential relationships and help focus future studies on cellular toxicity for graphene more effectively and accurately.

A meta-analysis approach has been employed in this study including 792 publications mine, along with concomitantly 986 graphene cytotoxicity data samples describing 10 qualitative and quantitative attributes such as graphene, cell properties and further the methodological settings, accordingly generate a comparatively comprehensive compiled data set for graphene in practice. Predictive graphene cytotoxicity models for cell viability, IC_{50} and LDH were developed using Random Forests learning algorithm. The prediction models demonstrated an accuracy of $R^2 = 0.805$, $R^2 = 0.903$, and $R^2 = 0.986$ for cell viability, IC_{50} and LDH, respectively. Moreover, RFs models assessed normalized data for graphene cellular toxicity adequately and precisely. Importance of macroscopic properties for correlating toxicity indexes were evaluated using RF models, as is demonstrated that random permutation is used to achieve the optimal combination of these corresponding attributes as an attribute combination of measurable significance in correlating toxicity. We also note that three data sets composed of the most significant attributes of cell viability, IC_{50} and LDH respectively are instructive for future experiment design and conduction. For example, diameter, cell morph, cell source, detection method, exposure time, exposure dose are composed of the optional attribute set for graphene toxicity study in which CV was used as the index of toxicity. This provides more targeted study design and analysis. The way is also applicable to IC_{50} and LDH indicator. All possible arrangements were taken to avoid unnecessary steps aimed at similar researches being both significant cost saving and more scientific and rational.

Overall, this paper implied that machine learning of literature data-compiling could guide and predict chief attributes (e.g. graphene properties and experimental conditions) related to graphene cytotoxicity that should be characterized and attached importance to. In view of that large data sets are available, literature data collection was expected to help to formulate advanced regulatory policies for health and environment. The collection of documents will help to formulate sound health policies. Our paper offers an excellent example that the value of the published information and proper modeling approaches should be considered rigorously to pair graphene cytotoxicity with the relevant attributes, which will reveal further potential relationships and help focus future studies on cellular toxicity for graphene more effectively and accurately.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data and materials are included in the manuscript, tables, figures and supplements.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was Supported by the National Natural Science Foundation of China (No. 81673218), the Natural Science Foundation of Jiangsu Province (No. BK20201268), National College Students Innovation and Entrepreneurship Training Program (No. 201910286183X) and the Fundamental Research Funds for the Central Universities (No. 2242020K40192).

Authors' contributions

Ting Zhang designed the experiment. Jianli Wang and Jingying Wu extracted and disposed the data from the publications. The main experiments, data analysis and figure generations were performed by Ying Ma, and with the help from Chuxuan Tong. Ying Ma and Jianli Wang co-wrote the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

Not applicable.

Author's information

Affiliations

¹Key Laboratory of Environmental Medicine Engineering, Ministry of Education, School of Public Health, Southeast University, Nanjing, 210009, China.

Ying Ma, Jianli Wang, Jingying Wu, Ting Zhang

²School of Information Technology and Electrical Engineering, The University of Queensland Brisbane, QLD 4072, Australia.

Chuxuan Tong

References

1. Duch, M. C.; Budinger, G. R.; Liang, Y. T.; Soberanes, S.; Urich, D.; Chiarella, S. E.; Campochiaro, L. A.; Gonzalez, A.; Chandel, N. S.; Hersam, M. C.; Mutlu, G. M., Minimizing oxidation and stable nanoscale dispersion improves the biocompatibility of graphene in the lung. *Nano Lett* **2011**, *11* (12), 5201-7.
2. Bianco, A., Graphene: safe or toxic? The two faces of the medal. *Angew Chem Int Ed Engl* **2013**, *52* (19), 4986-97.
3. Liao, K. H.; Lin, Y. S.; Macosko, C. W.; Haynes, C. L., Cytotoxicity of Graphene Oxide and Graphene in Human Erythrocytes and Skin Fibroblasts. *Acs Applied Materials & Interfaces* **2011**, *3* (7), 2607-2615.
4. Sanchez, V. C.; Jachak, A.; Hurt, R. H.; Kane, A. B., Biological Interactions of Graphene-Family Nanomaterials: An Interdisciplinary Review. *Chemical Research in Toxicology* **2012**, *25* (1), 15-34.
5. Jastrzebska, A. M.; Kurtycz, P.; Olszyna, A. R., Recent advances in graphene family materials toxicity investigations. *J Nanopart Res* **2012**, *14* (12), 1320.
6. Akhavan, O.; Ghaderi, E.; Akhavan, A., Size-dependent genotoxicity of graphene nanoplatelets in human stem cells. *Biomaterials* **2012**, *33* (32), 8017-25.
7. Delgado-Rodríguez, M.; Sillero-Arenas, M., Systematic review and meta-analysis. *Med Intensiva* **2018**, *42* (7), 444-453.
8. Muhammad Bilal, E. O., Rong Liu, Joyce C. Breger, Igor L. Medintz, and Yoram Cohen, Bayesian Network Resource for Meta-Analysis Cellular Toxicity of Quantum Dots. *small* **2019**.
9. Meta-analysis_ a critical appraisal of the methodology, benefits and drawbacks. *British Journal of Hospital Medicine* **2019**.
10. Nikolakopoulou, A.; Mavridis, D.; Furukawa, T. A.; Cipriani, A.; Tricco, A. C.; Straus, S. E.; Siontis, G. C. M.; Egger, M.; Salanti, G., Living network meta-analysis compared with pairwise meta-analysis in comparative effectiveness research: empirical study. *Bmj* **2018**, *360*, k585.
11. Gernand, J. M.; Casman, E. A., Nanoparticle Characteristic Interaction Effects on Pulmonary Toxicity: A Random Forest Modeling Framework to Compare Risks of Nanomaterial Variants. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg* **2016**, *2* (2).
12. Oh, E.; Liu, R.; Nel, A.; Gemill, K. B.; Bilal, M.; Cohen, Y.; Medintz, I. L., Meta-analysis of cellular toxicity for cadmium-containing quantum dots. *Nat Nanotechnol* **2016**, *11* (5), 479-86.

13. 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.
14. Siqing Li a, Y. L., Zehao Huang, Yan Kou, Ailing Hua, Efficacy and safety of nano-silver dressings combined with recombinant human epidermal growth factor for deep second-degree burns: A meta-analysis. *Elsevier* **2019**.
15. Jia, P. P.; Sun, T.; Junaid, M.; Yang, L.; Ma, Y. B.; Cui, Z. S.; Wei, D. P.; Shi, H. F.; Pei, D. S., Nanotoxicity of different sizes of graphene (G) and graphene oxide (GO) in vitro and in vivo. *Environ Pollut* **2019**, *247*, 595-606.
16. Juan Ma, R. L., Xiang Wang, Qian Liu, Yunan Chen, Russell P. Valle,§ Yi Y. Zuo,§ Tian Xia, and Sijin Liu, Crucial Role of Lateral Size for Graphene Oxide in Activating Macrophages and Stimulating Pro-inflammatory Responses in Cells and Animals. *ACS NANO* **2015**.
17. Jang, S. C.; Kang, S. M.; Lee, J. Y.; Oh, S. Y.; Vilian, A. E.; Lee, I.; Han, Y. K.; Park, J. H.; Cho, W. S.; Roh, C.; Huh, Y. S., Nano-graphene oxide composite for in vivo imaging. *Int J Nanomedicine* **2018**, *13*, 221-234.
18. Seo, J. E.; Wu, Q.; Bryant, M.; Ren, L.; Shi, Q.; Robison, T. W.; Mei, N.; Manjanatha, M. G.; Guo, X., Performance of high-throughput CometChip assay using primary human hepatocytes: a comparison of DNA damage responses with in vitro human hepatoma cell lines. *Arch Toxicol* **2020**, *94* (6), 2207-2224.
19. Ren, W.; Chang, H.; Teng, Y., Sulfonated graphene-induced hormesis is mediated through oxidative stress in the roots of maize seedlings. *Sci Total Environ* **2016**, *572*, 926-934.
20. Khan, M.; Khan, M.; Al-Marri, A. H.; Al-Warthan, A.; Alkhatlan, H. Z.; Siddiqui, M. R.; Nayak, V. L.; Kamal, A.; Adil, S. F., Apoptosis inducing ability of silver decorated highly reduced graphene oxide nanocomposites in A549 lung cancer. *Int J Nanomedicine* **2016**, *11*, 873-83.
21. Usman, M. S.; Hussein, M. Z.; Fakurazi, S.; Masarudin, M. J.; Ahmad Saad, F. F., A bimodal theranostic nanodelivery system based on [graphene oxide-chlorogenic acid-gadolinium/gold] nanoparticles. *PLoS One* **2018**, *13* (7), e0200760.
22. Bitounis, D.; Parviz, D.; Cao, X.; Amadei, C. A.; Vecitis, C. D.; Sunderland, E. M.; Thrall, B. D.; Fang, M.; Strano, M. S.; Demokritou, P., Synthesis and Physicochemical Transformations of Size-Sorted Graphene Oxide during Simulated Digestion and Its Toxicological Assessment against an In Vitro Model of the Human Intestinal Epithelium. *Small* **2020**, *16* (21), e1907640.
23. Kang, Y.; Liu, J.; Wu, J.; Yin, Q.; Liang, H.; Chen, A.; Shao, L., Graphene oxide and reduced graphene oxide induced neural pheochromocytoma-derived PC12 cell lines apoptosis and cell cycle alterations via the ERK signaling pathways. *Int J Nanomedicine* **2017**, *12*, 5501-5510.
24. Tobias Lammel, P. B., Maria-Luisa Fernández-Cruz and José M Navas, Internalization and cytotoxicity of graphene oxide and carboxyl graphene nanoplatelets in the human hepatocellular carcinoma cell line Hep G2. *Particle and Fibre Toxicology* **2013**.
25. Lammel, T.; Boisseaux, P.; Fernandez-Cruz, M. L.; Navas, J. M., Internalization and cytotoxicity of graphene oxide and carboxyl graphene nanoplatelets in the human hepatocellular carcinoma cell line Hep G2. *Particle and Fibre Toxicology* **2013**, *10*.
26. Akter, S.; Addepalli, R.; Netzel, M. E.; Tinggi, U.; Fletcher, M. T.; Sultanbawa, Y.; Osborne, S. A., Antioxidant-Rich Extracts of Terminalia ferdinandiana Interfere with Estimation of Cell Viability. *Antioxidants (Basel)* **2019**, *8* (6).
27. Fang, I. J.; Trewyn, B. G., Application of Mesoporous Silica Nanoparticles in Intracellular Delivery of Molecules and Proteins. *Methods in Enzymology* **2012**, *508* (1), 41.
28. Liu, J.-H.; Wang, T.; Wang, H.; Gu, Y.; Xu, Y.; Tang, H.; Jia, G.; Liu, Y., Biocompatibility of graphene oxide intravenously administrated in mice—effects of dose, size and exposure protocols. *Toxicology Research* **2015**, *4* (1), 83-91.
29. Yakobson, B. I. B.; Ding, F. F., Correction to Observational Geology of Graphene, at the Nanoscale. *Acs Nano* **2012**, *5* (3), 1569-1574.
30. Zhang, S.; Yang, K.; Feng, L.; Liu, Z., In vitro and in vivo behaviors of dextran functionalized graphene. *Carbon* **2011**, *49* (12), 4040-4049.

31. Xi; Ren; Luis; F.; Tapias; Bernhard; J.; Jank; Douglas; J., Ex vivo non-invasive assessment of cell viability and proliferation in bio-engineered whole organ constructs. *Biomaterials* **2015**.
32. Tapias, L. F.; Gilpin, S. E.; Ren, X.; Wei, L.; Fuchs, B. C.; Tanabe, K. K.; Lanuti, M.; Ott, H. C., Assessment of Proliferation and Cytotoxicity in a Biomimetic Three-Dimensional Model of Lung Cancer. *The Annals of Thoracic Surgery* **2015**.
33. John O'Brien, I. W., Terry Orton and François Pognan, Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *FEBS* **2000**,267.
34. Berridge, M. V.; Herst, P. M.; Tan, A. S., Tetrazolium dyes as tools in cell biology: new insights into their cellular reduction. *Biotechnol Annu Rev* **2005**,11, 127-152.
35. Gomez-Gutierrez, J. G.; Bhutiani, N.; McNally, M. W.; Chuong, P.; Yin, W.; Jones, M. A.; Zeiderman, M. R.; Grizzle, W. E.; McNally, L. R., The neutral red assay can be used to evaluate cell viability during autophagy or in an acidic microenvironment in vitro. *Biotech Histochem* **2020**, 1-9.
36. Hu, W.; Peng, C.; Lv, M.; Li, X.; Zhang, Y.; Chen, N.; Fan, C.; Huang, Q., Protein Corona-Mediated Mitigation of Cytotoxicity of Graphene Oxide. *Acs Nano* **2011**,5 (5), 3693-700.
37. Ma, J.; Liu, R.; Wang, X.; Liu, Q.; Chen, Y.; Valle, R. P.; Zuo, Y. Y.; Xia, T.; Liu, S., Crucial Role of Lateral Size for Graphene Oxide in Activating Macrophages and Stimulating Pro-inflammatory Responses in Cells and Animals. *Acs Nano* **2015**,9 (10), 10498-515.
38. Baguley, B. C.; Hicks, K. O.; Wilson, W. R., *Tumor cell cultures in drug development*. 2002.
39. Heuvel, J. P. V., 2.03 - Receptor Theory and the Ligand–Macromolecule Complex. *Comprehensive Toxicology* **2010**, 27-50.
40. Choi, Y. J.; Gurunathan, S.; Kim, J. H., Graphene Oxide-Silver Nanocomposite Enhances Cytotoxic and Apoptotic Potential of Salinomycin in Human Ovarian Cancer Stem Cells (OvCSCs): A Novel Approach for Cancer Therapy. *Int J Mol Sci* **2018**,19 (3).
41. Seddon, W. L.; Prosser, C. L., Non-enzymatic isolation and culture of channel catfish hepatocytes. *Comparative Biochemistry & Physiology Part A Molecular & Integrative Physiology* **1999**,123 (1), 9.
42. Schapire, R. E., Random Forests. *Machine Learning* **2001**.
43. Jia, J.; Liu, Z.; Xiao, X.; Liu, B.; Chou, K. C., iPPI-Esml: An ensemble classifier for identifying the interactions of proteins by incorporating their physicochemical properties and wavelet transforms into PseAAC. *J Theor Biol* **2015**,377, 47-56.
44. Kesler, S. R.; Noll, K.; Cahill, D. P.; Rao, G.; Wefel, J. S., - The effect of IDH1 mutation on the structural connectome in malignant astrocytoma. **2017**,- 131 (- 3), - 574.
45. Chiang, S.; Levin, H. S.; Wilde, E.; Haneef, Z., White matter structural connectivity changes correlate with epilepsy duration in temporal lobe epilepsy. *Epilepsy Research* **2016**,120, 37-46.
46. Kačar, K.; Rocca, M. A.; Copetti, M.; Sala, S.; Mesaroš, Š.; Stosić Opincal, T.; Caputo, D.; Absinta, M.; Drušević, J.; Kostić, V. S.; Comi, G.; Filippi, M., Overcoming the Clinical–MR Imaging Paradox of Multiple Sclerosis: MR Imaging Data Assessed with a Random Forest Approach. *American Journal of Neuroradiology* **2011**,32, 2098-2102.
47. Koley, S.; Sadhu, A. K.; Mitra, P.; Chakraborty, B.; Chakraborty, C., Delineation and diagnosis of brain tumors from post contrast T1-weighted MR images using rough granular computing and random forest. *Applied Soft Computing* **2016**,41, 453-465.
48. Serag, A. W., AG-Telford, EJ (Telford, Emma J, SEGMA: An Automatic SEGmentation Approach for Human Brain MRI Using Sliding Window and Random Forests. *FRONTIERS IN NEUROINFORMATICS* **2017**.
49. Tibshirani, R., Regression shrinkage and selection via the Lasso a retrospective. **2010**.
50. Xiaohui, R., Lasso regression model of variable selection function and its application in prediction model of the development of diabetes. *China Science & Technology Overview* **2017**.
51. Hastie, H. Z. a. T., Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B* **2005**.
52. Bahar, I.; Atilgan, A. R.; Erman, B., Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design* **1997**,2 (3), 173-181.
53. Saitta, L., Support-Vector Networks. *Machine Learning* **1995**.

54. Pasolli, E.; Truong, D. T.; Malik, F.; Waldron, L.; Segata, N., Machine Learning Meta-analysis of Large Metagenomic Datasets: Tools and Biological Insights. *PLoS Comput Biol* **2016**, *12* (7), e1004977.
55. Boser, B. E.; Guyon, I. M.; Vapnik, V. N., A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*, 1992; pp 144-152.
56. Edgar Osuna, R. F., An Improved Training Algorithm for Support Vector Machines.
57. Zhi-Qiang Zeng, H.-B. Y., Hua-Rong Xu, Yan-Qi Xie, Ji Gao, Fast Training of Support Vector Machines Using Sequential Minimal Optimization. *International Conference on Intelligent System and Knowledge Engineering* **2008**.
58. Hsieh, W. W., Learning and generalization. In *Machine Learning Methods in the Environmental Sciences*, 2009; pp 127-156.
59. Herlocker, J. L.; Konstan, J. A.; Borchers, A.; Riedl, J., An algorithmic framework for performing collaborative filtering. *ACM SIGIR FORUM* **1999**, *51* (2), 227-234.

Tables

Table1. Overview of literature Graphene family data attributes.

Overview of literature Graphene family data attributes.										
1. Graphene Diameter [nm]	Au	2%	Others	4%	Breast	14%	Others	15%	Neutral Red Assay	4%
0.8-6000	COOH	2%			Lung	13%			Others	11%
	Others	16%	5. Cell Source	3	Blood	11%	8. Detection Method	22		
2. Oxidation State	3		Human	76%	Liver	8%	CCK8	24%	9. Exposure Time [hrs]	
Primary	6%	4. Cell Morph	12	Mouse	19%	Embryo	7%	MTT	15%	0.5-120
Oxidation	72%	Epithelial	53%	Rat	5%	Cervix	7%	Alamar Blue	13%	
Reduction	22%	Stem-cell	14%			Bone	6%	MTS	7%	10. Exposure Dose [$\mu\text{g mL}^{-1}$]
		Fibroblast	10%	6. Cell Line	2	Adipose	6%	Trypan Blue	6%	0.00001-5000
3. Surface Modification	32	Monocyte	7%	Cell line	93%	Adrenal-gland	5%	Calcein AM assay	6%	
Unmodified	75%	Monocyte-Macrophage	5%	Primary	7%	Ascites	3%	Neutral Red Uptake Assay	5%	
PEG	3%	Neuronal	5%			Ovary	3%	WST-1	5%	
HI	2%	Endothelial	2%	7. Organ Source	27	Bone-marrow	20%	LDH	4%	

Table 2. Importance of attributes in RF model via random permutation.

Selected Attributes	R ²
Diameter', 'Cell Morph', 'Cell Source', 'Detection Method', 'Exposure Time', 'Exposure Dose'	0.80484849
Diameter', 'Cell Morph', 'Cell Source', 'Cell Line', 'Detection Method', 'Exposure Time', 'Exposure Dose'	0.80732814
Diameter', 'Cell Morph', 'Cell Source', 'Organ Source', 'Detection Method', 'Exposure Time', 'Exposure Dose'	0.80853033
Diameter', 'Oxidation State', 'Cell Morph', 'Cell Source', 'Organ Source', 'Detection Method', 'Exposure Time', 'Exposure Dose'	0.81246238
Diameter', 'Oxidation State', 'Cell Morph', 'Cell Source', 'Cell Line', 'Organ Source', 'Detection Method', 'Exposure Time', 'Exposure Dose'	0.82032346

Figures

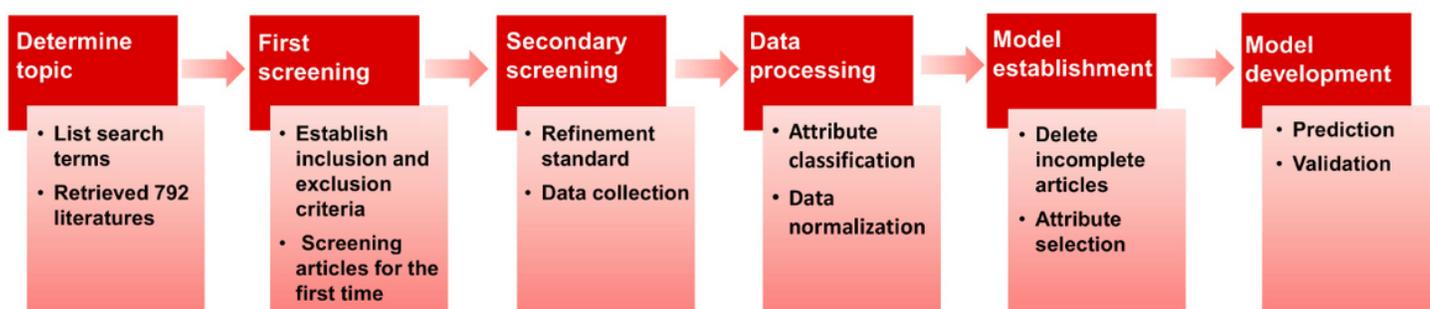


Figure 1

Meta-analysis workflow. Workflow for analysis of literature Graphene family toxicity data. Graphene family toxicity data, together with their physicochemical properties and experimental condition parameters, were collected via manual literature data-mining. The collected data were then processed to normalize numeric attributes and to delete incomplete data. Regression models were developed using the Random Forest technique for both cell viability with the most suitable attributes selected via exhaustive search.

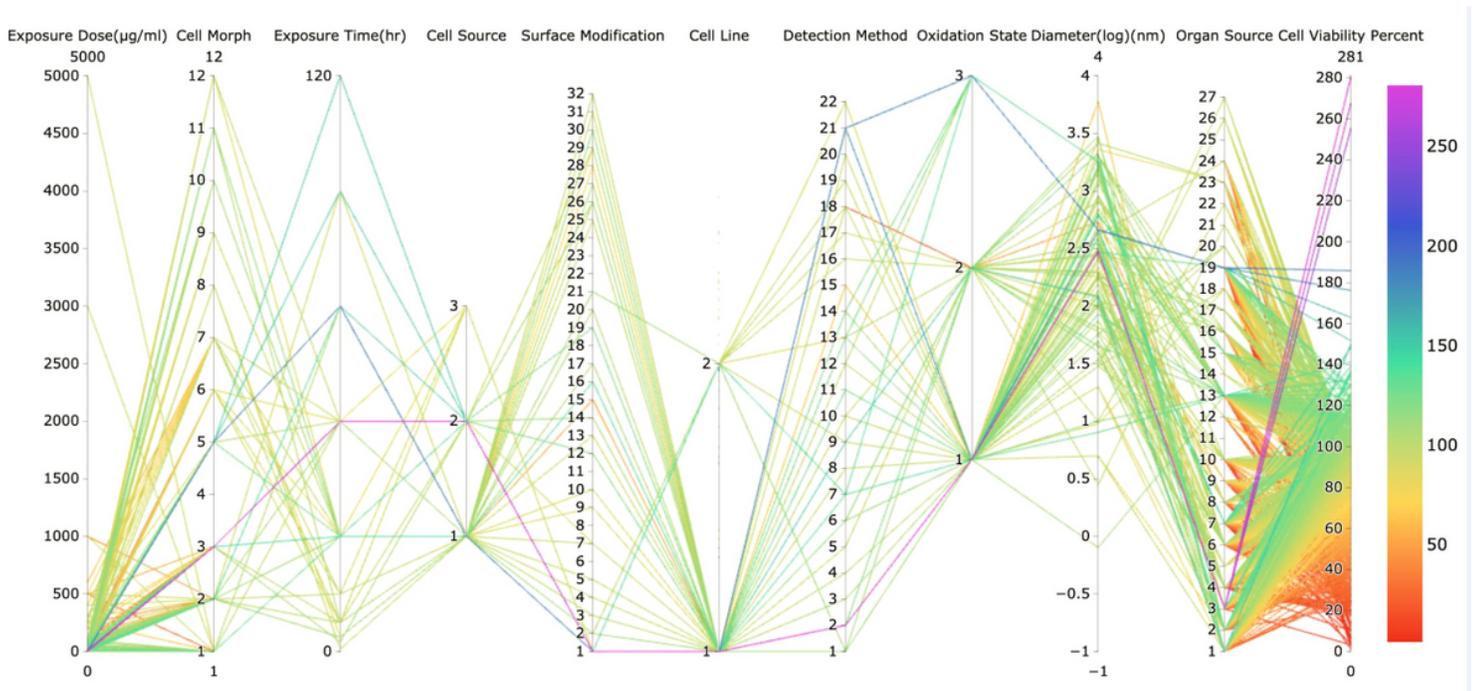


Figure 2

Parallel coordinate chart for cell viability. Note that the intercepts on the axes are joined by line segments to make a polyline, displaying the cytotoxicity of the normalized properties listed in the overview table.

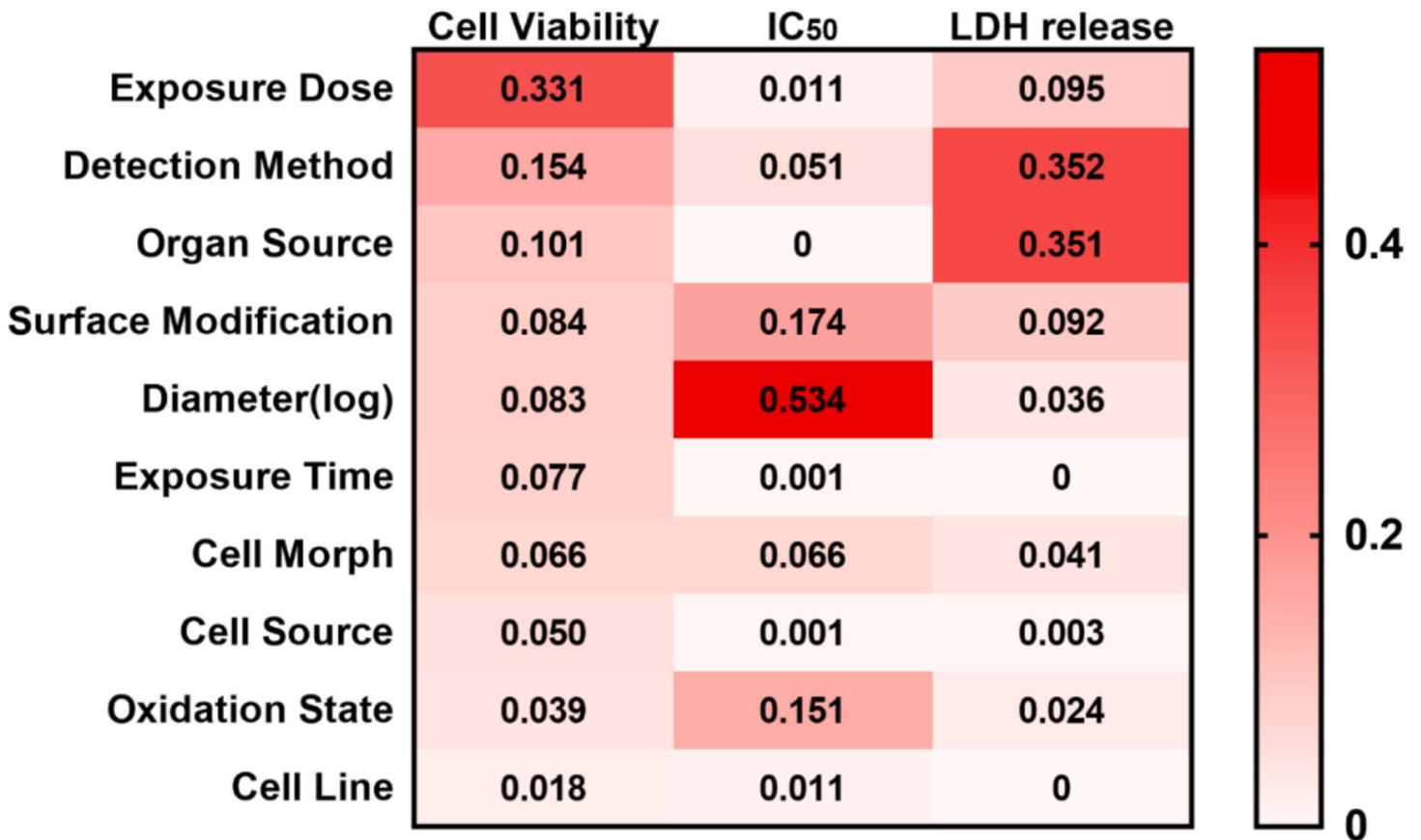


Figure 3

Importance of macroscopic properties for correlating cell viability, IC50 and LDH. An attribution which leads to significant increased R square is considered to be an important attribute for correlating graphene bioactivity (i.e., cell viability, IC50 or LDH).

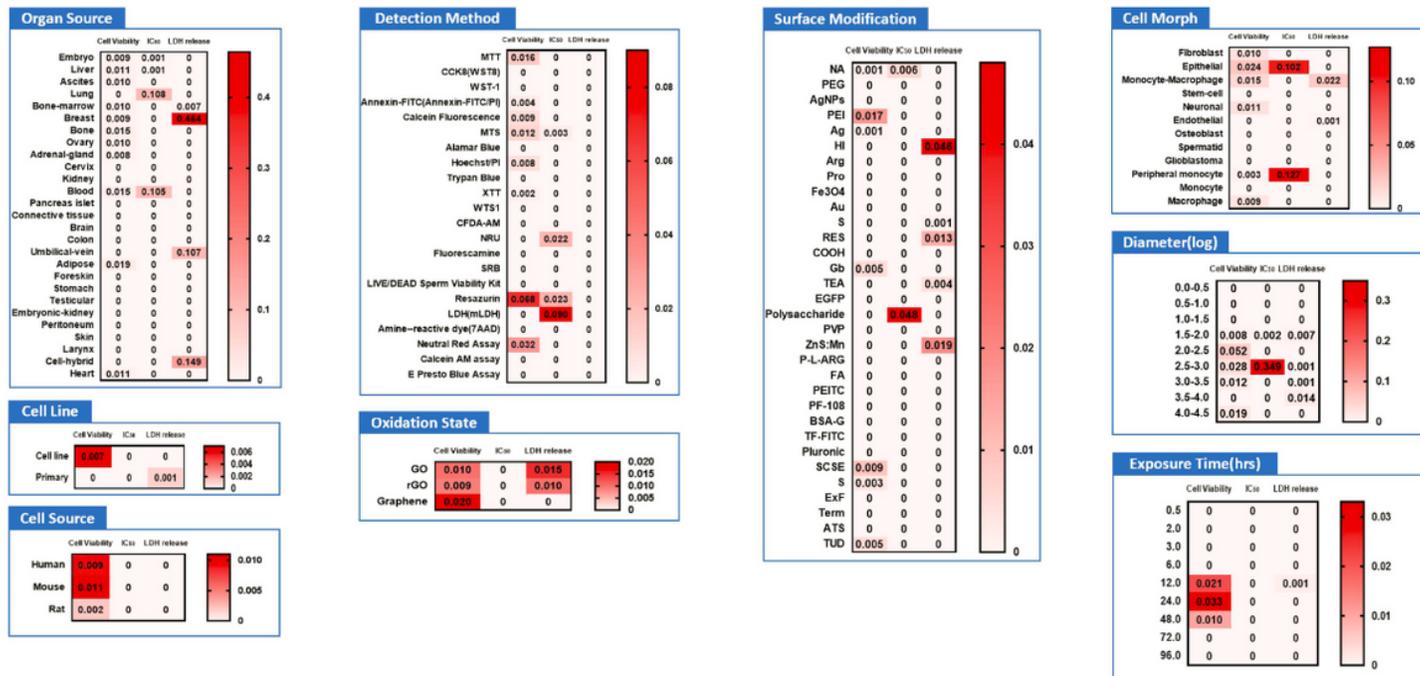


Figure 4

Conditional dependence of graphene bioactivity on subgroups of graphene properties and experimental conditions. The conditional dependence was illustrated via the distribution of samples concerning organ source, cell line, cell source, oxidation state, detection method, cell morph, diameter (Log10), exposure time, surface modification.

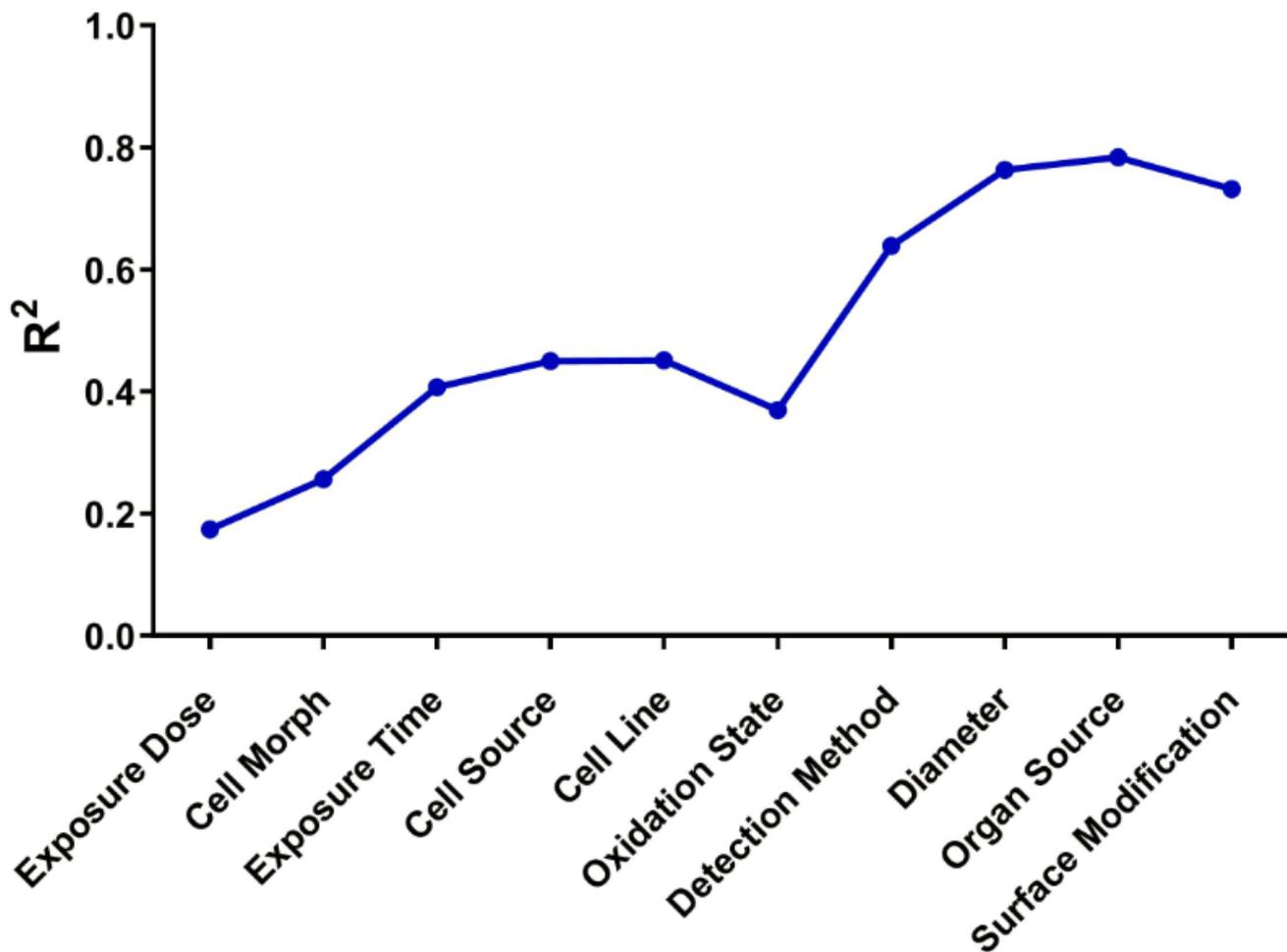


Figure 5

Importance of attributes from an exhaustive search for explaining cell viability. Attributes were incrementally added to the model according to the strength of the correlation.

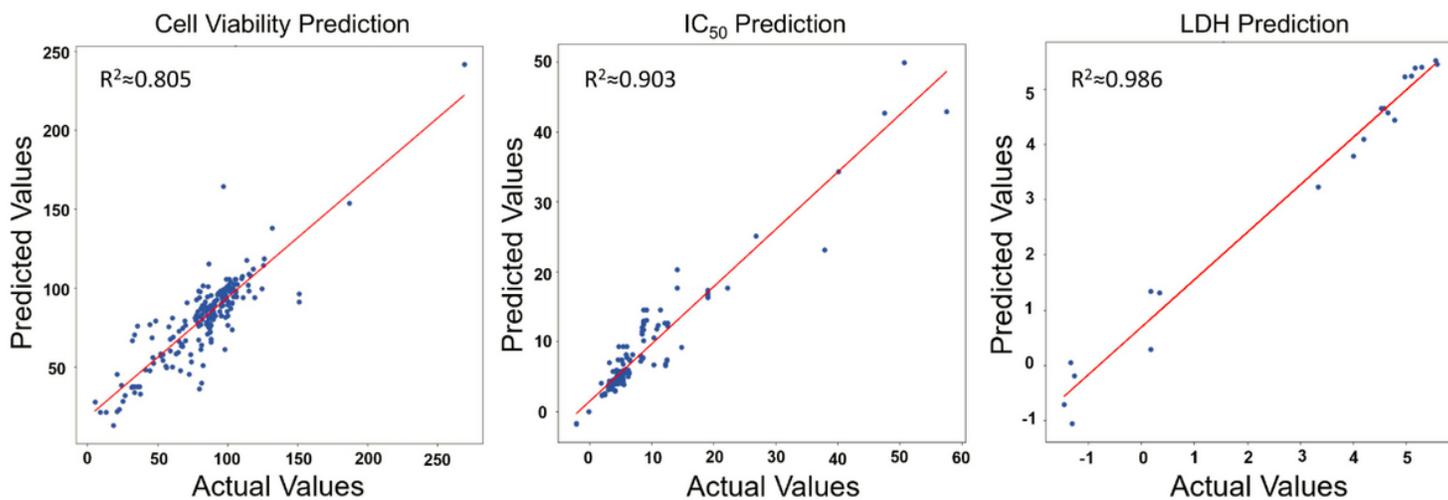


Figure 6

RF models. (a) RF model predicted versus observed cell viability, IC₅₀(Loge), and LDH release.

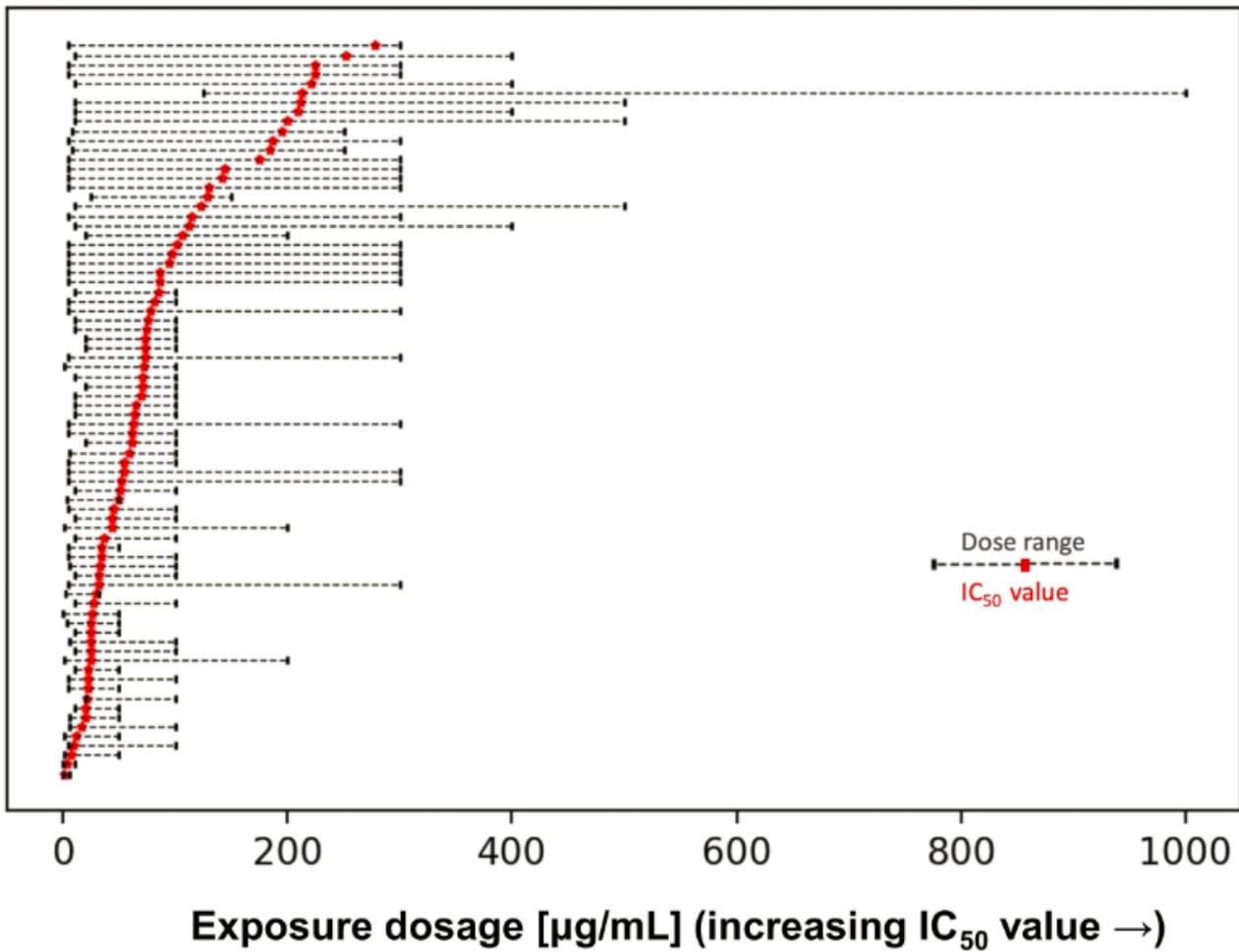


Figure 7

IC_{50} values (red) and dosage range used (black) for major researches. The IC_{50} values are sorted and plotted by increasing values to simplify the visual presentation. IC_{50} is defined here as the exposure concentration at which there is 50% cell death or inhibition of cell growth or other utilized toxicity metric.

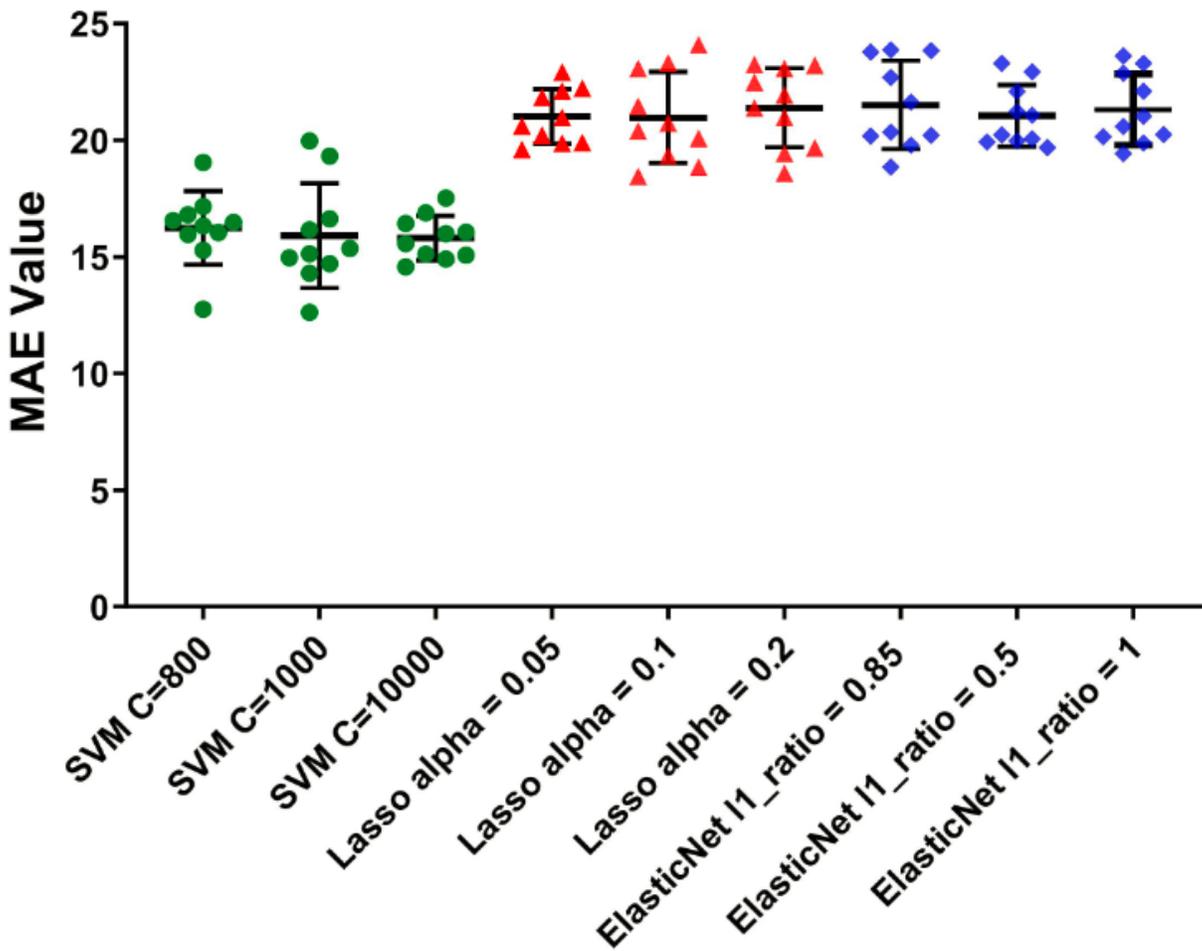


Figure 8

Model comparison. Prediction performances in different cross-validation studies were measured by Mean Absolute Error (MAE).

Supplementary Files

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

- [Additionalfile.docx](#)