Chapter 3
Machine Learning Applications in Nanomedicine and Nanotoxicology: An Overview

Gerardo M. Casañola-Martin
North Dakota State University, Fargo, USA

Hai Pham-The
https://orcid.org/0000-0003-4531-7223
Hanoi University of Pharmacy, Hanoi, Vietnam

ABSTRACT

The development of machine learning algorithms together with the availability of computational tools nowadays have given an increase in the application of artificial intelligence methodologies in different fields. However, the use of these machine learning approaches in nanomedicine remains still under-explored in certain areas, despite the development in hardware and software tools. In this review, the recent advances in the conjunction of machine learning with nanomedicine are shown. Examples dealing with biomedical properties of nanoparticles, characterization of nanomaterials, text mining, and image analysis are also presented. Finally, some future perspectives in the integration of nanomedicine with cloud computing, deep learning and other techniques are discussed.

INTRODUCTION

Nanomaterials have arisen as one of the promising fields in material sciences and technologies in the current century. The most prominent that should be mentioned include among others the development of new fuel cells (Liu, Ling, Su, & Lee, 2004), electronic devices (Novoselov et al., 2012), coatings (Ragesh, Anand Ganesh, Nair, & Nair, 2014), diagnostic imaging (Wu et al., 2002), and drug delivery (Muller & Keck, 2004). The field of nanomedicine has gained a great importance in the last years. In the case of nanomedicine successful applications of computational approaches, most of them related with QSAR (quantitative-structure activity relationships) studies.

DOI: 10.4018/978-1-6684-6291-1.ch003

In this sense in this review we cover the main applications of machine learning methods in medicine, where the most relevant publications are discussed for this field. This has led to the development of different approaches, based in the predictions of the effects of nanoparticles, the enzyme inhibition of carbon nanotubes and others in the use of artificial intelligence for the study of nanoimages to predict the response to certain effects, or the improvement of personalized cancer treatment based in nanometer-scale drug delivery systems.

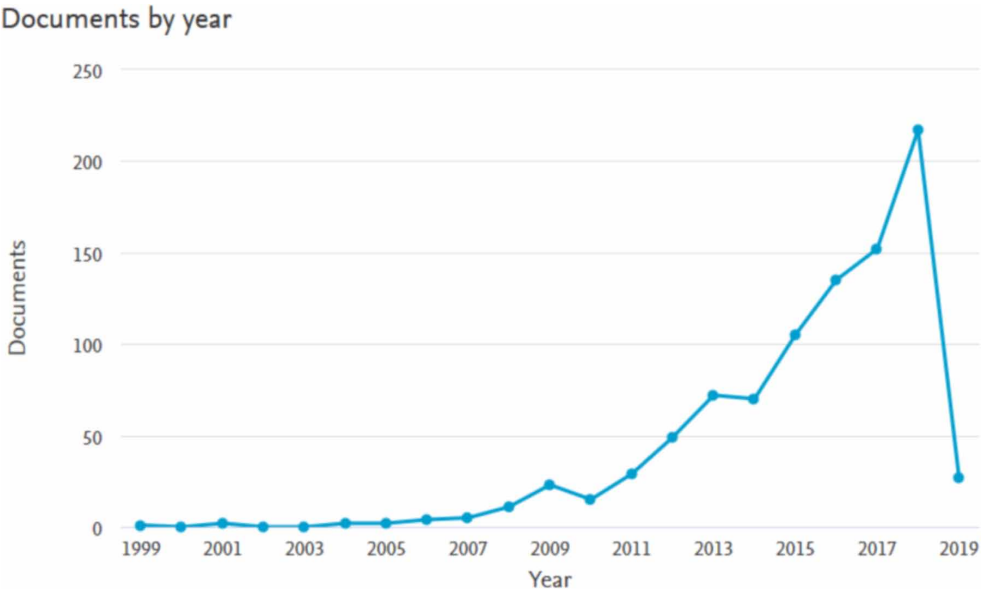
The State of the Art of Machine Learning and Nanomedicine

In the era of data, machine learning (ML) algorithms have played a crucial role by helping to improve the predictions of biological, physical, chemical and toxicological effects. As the size of the information increase these artificial intelligence methods are of valuable important to analyze this huge amount of data.

One of the main practical use is for virtual throughput screening of new chemical entities, based on the information gathered from previous one. A common procedure to perform this kind of study is by the Quantitative Structure-Activity Relationships (QSAR) methods, a statistical method, in a general way which try to correlate the biological effect with the features describing the compounds, molecular descriptors gathered directly from the chemical structure through mathematical equations of the connectivity graph. In the case of nanomaterials other descriptors can also be obtained like shape, size, composition, surface modifications, and propensity to agglomerate, interaction with different molecules, and aqueous solubility that could influence the properties of the nanomaterials affecting the biological responses.

The impact of this field was analyzed by retrieving in the Scopus database a search using the search criteria “Machine Learning” AND “Nanomedicine”. As can be observed in Figure 1 the time period for Machine Learning-Nanomedicine search comprises the time period from 1999-present, with 921 documents this is compressible because this is novel multidisciplinary field. As can be noted a constant increased in this field is observed. From this could be said that this thematic constitutes a hot topic nowadays due to the increased interest in drug discovery and design.

Figure 1. Yearly publications based on ‘Machine Learning’-‘Nanomedicine’ keywords

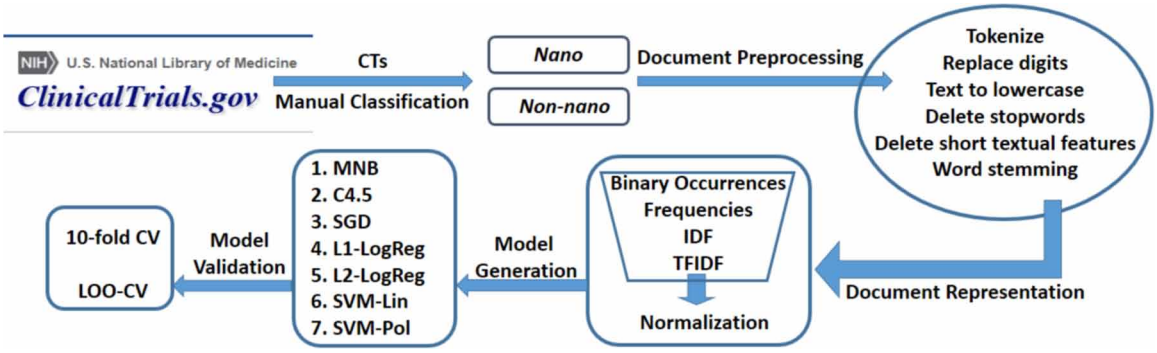


Innovative Applications and Methods

One of the big challenges for nanomedicine is the management of high volumes of data generated through all the years of experimental research. The improvement of the tools for information retrieval is a crucial step in order to organize the big datasets and compile the necessary and relevant information to be processed for its use in studies. In this sense, a work developed by de la Iglesia et al. make use of a machine learning method combined with Natural Language Processing (NLP) methods (de la Iglesia et al., 2014). In the mentioned report is done the classification of clinical trials (CT) in the ClinicalTrials.gov database that test nanotechnology products (nanodrugs and nanodevices) from those one that make testing of conventional drugs.

In the proposed method, the authors developed a workflow (see Figure 2) that comprises the extraction and manual annotation of the CT into two categories: nano and non-nano, with 500 CTs selected as positives (nano) and 500 CTs as negatives (non-nano). In a second step is performed the preprocessing of the data eliminating the irrelevant information, the standardization and filtering, by using Natural Language Toolkits to convert the information into textual features using a vector-based representation. In the following state several machine learning algorithms implemented in the Weka workbench are used and the best predictions model showed an F-measure value equal to 0.955. In summary this study helps to retrieve and organize the big data in CTs database, making useful for researchers.

Figure 2. Graphical workflow of the work developed by de la Iglesia et al. (2014)



In a recent review from Jones et al. are described the applications of machine learning in nanomedicine focused mainly in the prediction of biomedical responses (Jones, Ghandehari, & Facelli, 2016). In the first section, the authors show the QSAR studies related with nanoparticles and its responses related with nanoparticle adherence, nanoparticle size, cellular uptake, molecular release and so on (Epa et al., 2012; Fourches et al., 2010; Puzyn, Leszczynska, & Leszczynski, 2009; Toropova & Toropov, 2017; Winkler et al., 2014). The authors noted that the Artificial Neural Networks (ANN) are the main algorithms used through most of the works. From their point of view this choice is due to the complexity of the data given by the great heterogeneity in the nanomaterials (Tantra et al., 2015). This led to a large number of different features that also the difficulty to develop prediction studies.

In the same way, Winkler et al. (Papa, Doucet, Sangion, & Doucet-Panaye, 2016), modelled the bioactivity of gold nanoparticles in cell association studies. In this study, several machine learning algorithms

are proved, and all showed accuracy values (Ac) > 0.88 for the training set, and Ac > 0.73 for the prediction set. The same researcher published a review of computational modeling of nanomaterials, where are described most of the research related with the use of machine learning models in the prediction of biological effects (Winkler, 2016). The reports included here are based mainly in the strongest research teams and researchers working in the field. This review article helps to complement the previous ones already commented, to give a picture of the state of the art of the thematic.

However, some interesting approaches in modeling of nanomaterial properties are not included, like the case of the works of Michael et al. that predicting the energy gap of graphene nanoflakes by ML methods (Fernandez, Abreu, Shi, & Barnard, 2016). In this work the topological information gathered from the structure was used as features. The E_G values (response variable) were well predicted in the ML regression models with $R^2 > 0.80$. In a previous work these authors use the ML algorithms to establish a structure-property approach for some nanomaterial properties such as: ionization potential, electronic band gap, energy of Fermi level, and electron affinity Fernandez, Shi, & Barnard, 2016). The remarkable aspect in these studies is the encoding of the molecular structures to obtain the features (molecular descriptors) to perform the QSAR studies. These two studies and those commented below are summarized in Table 1.

Table 1. Summary of recent studies on QSPR-ML nanomedicine

Property Under Study[a]	ML Algorithm[b]	Accuracy	Reference
Graphene nanoflakes- E_G	SVM	$R^2_{train} = 0.862$ $R^2_{ext} = 0.812$	Fernandez 2016
Graphenes nanoflakes- E_A	SVM	$Q^2_{train} = 0.77$ $R^2_{ext} = 0.90$	(Fernandez, Shi, et al., 2016)
Graphenes nanoflakes- E_F	kNN	$Q^2_{train} = 0.99$ $R^2_{ext} = 0.89$	(Fernandez, Shi, et al., 2016)
Graphenes nanoflakes- E_G	SVM	$Q^2_{train} = 0.66$ $R^2_{ext} = 0.89$	(Fernandez, Shi, et al., 2016)
Graphenes nanoflakes- E_I	ANN	$Q^2_{train} = 0.78$ $R^2_{ext} = 0.90$	(Fernandez, Shi, et al., 2016)
Carbon Nanotube -ATPase Inhibitors	ANN	Sp = 99.6 Sn = 99.3	(González-Durruthy et al., 2019)
Carbon Nanotubes- Mitochondrial Oxygen Flux Dynamics	RF	$R^2_{train} = 0.855$ $R^2_{ext} = 0.856$	(González-Durruthy et al., 2017)
Computer Tomography images	kNN	97%	(Rani & Jawhar, 2019)

[a]Properties: E_G (Energy band gap), E_A (Electron affinity), E_F (Energy of the Fermi level), E_I (Ionization potential)

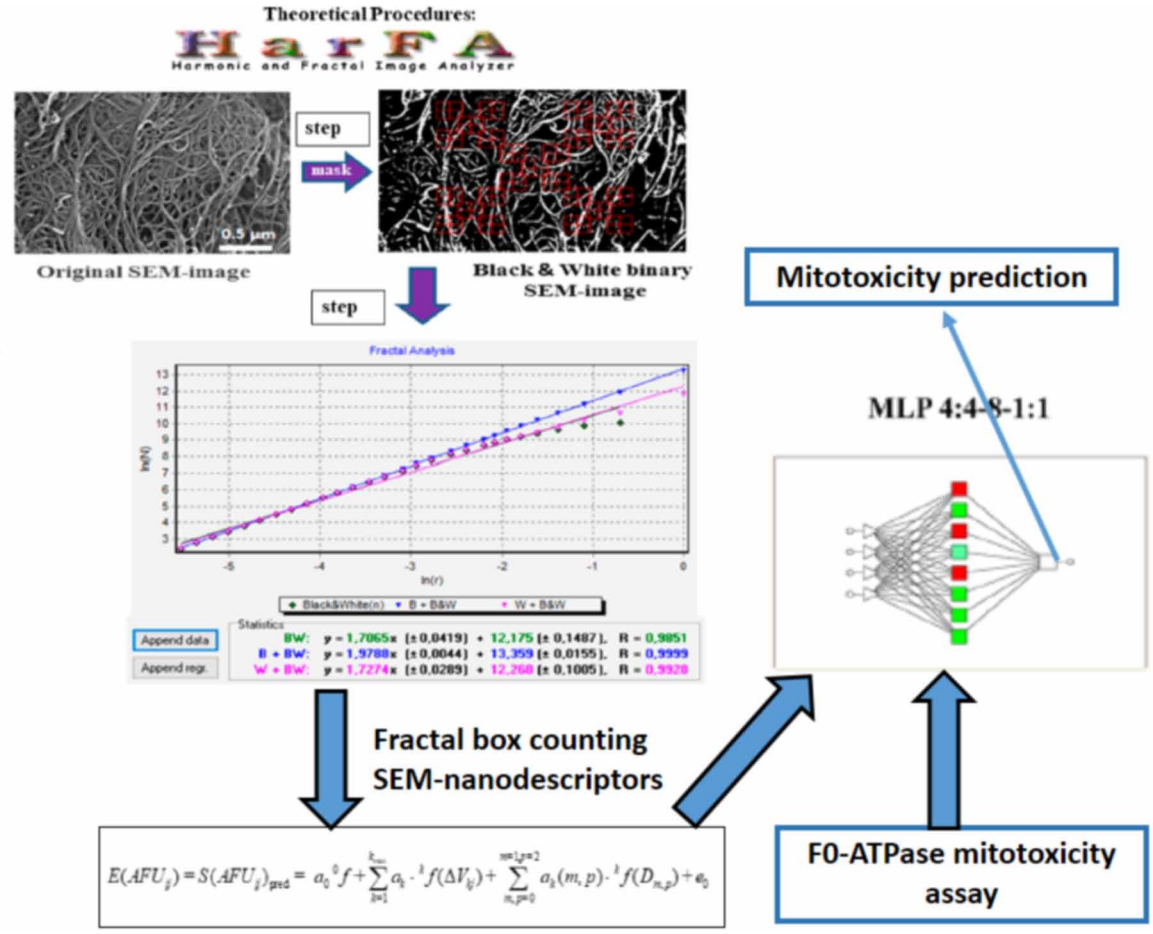
[b]Machine Learning Algorithms: ANN (Artificial Neural Networks), kNN (K-Nearest Neighbor), SVM (Support Vector Machine), RF (Random Forest)

Recently, some novel approaches have been developed in the efforts to apply the QSAR methodologies to challenging problems. As an example, a study performed by González-Durruty et al. (González-Durruthy et al., 2019) is shown. The authors studied the mitochondrial F0F1-ATP ATPase a target that offer a higher selectivity for selecting positive response (inhibitors) from inactive ones(non-inhibitors). In this sense, fractal SEM (scanning electron microscopy) nanodescriptors were calculated by using the

original SEM-image in a box counting algorithm linked to the calculation of the fractal dimension of the object by a linear fit of the data (Odziomek et al., 2017). Besides, subset of descriptors based in the different conditions (moving averages) an approach widely used in QSAR studies (González-Durruthy et al., 2017). These features were then fed into Artificial Neural Networks (ANN) for development of the classification models and adequate statistical parameters (Specificity >98.9% and Sensitivity >99.0%). The general theoretical workflow for this study is showed in Figure 3.

The nanoscale imaging technique has emerged as another interesting and very useful approach as is showed in the study performed by Rani et al. (Rani & Jawhar, 2019). IN this case, the Computer Tomography (CT) images were used to develop a measuring tool to detect lung tumor areas. The Support Vector Machine and K-Nearest Neighbor were the selected ML algorithms for the image classification and feature engineering. The performance of this classifier (Ac = 97%) provide an efficient method for the segmentation of lung lesion and the diagnosis and shows an interesting application for the field of medical informatics.

Figure 3. Theoretical workflow on mitochondrial F0-ATPase mitotoxic inhibition. Reprinted (adapted) with permission from (González-Durruthy et al., 2012).



CONCLUSION

Nanomedicine represents a challenge for nanoinformatics because of the widely spread data generated by the scientists for decades. This big amount of data is very heterogeneous and requires the use of powerful computational resources. These requirements could be satisfied with the integration of the machine learning algorithms in high performance computing (HPC) or Cloud computing platforms. Besides, the use of consensus models could benefit the predictions for this kind of studies, avoiding the bias that a single classifier/regressor could have on a collected data.

The combination of these approaches, with those showed in this review, such as text mining, image acquisition, multi-scale modeling could improve the effectiveness for the management of the data and information on nanomedical research and hence the discovery of new knowledge about the biomedical properties of nanomaterials.

ACKNOWLEDGMENT

GMCM is grateful for computer access, financial and administrative support provided by North Dakota State University, Department of Coatings and Polymer Materials.

REFERENCES

De La Iglesia, D., García-Remesal, M., Anguita, A., Munoz-Marmol, M., Kulikowski, C., & Maojo, V. (2014). A machine learning approach to identify clinical trials involving nanodrugs and nanodevices from ClinicalTrials.gov. *PLoS One*, *9*(10), e110331. doi:10.1371/journal.pone.0110331 PMID:25347075

Epa, V. C., Burden, F. R., Tassa, C., Weissleder, R., Shaw, S., & Winkler, D. A. (2012). Modeling Biological Activities of Nanoparticles. *Nano Letters*, *12*(11), 5808–5812. doi:10.1021/nl303144k PMID:23039907

Fernandez, M., Abreu, J. I., Shi, H., & Barnard, A. S. (2016). Machine Learning Prediction of the Energy Gap of Graphene Nanoflakes Using Topological Autocorrelation Vectors. *ACS Combinatorial Science*, *18*(11), 661–664. doi:10.1021/acscombsci.6b00094 PMID:27598830

Fernandez, M., Shi, H., & Barnard, A. S. (2016). Geometrical features can predict electronic properties of graphene nanoflakes. *Carbon*, *103*, 142–150. doi:10.1016/j.carbon.2016.03.005

Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S. Y., Mumper, R. J., & Tropsha, A. (2010). Quantitative Nanostructure– Activity Relationship Modeling. *ACS Nano*, *4*(10), 5703–5712. doi:10.1021/nn1013484 PMID:20857979

González-Durruthy, M., Manske Nunes, S., Ventura-Lima, J., Gelesky, M. A., González-Díaz, H., Monserrat, J. M., ... Cordeiro, M. N. D. S. (2019). MitoTarget Modeling Using ANN-Classification Models Based on Fractal SEM Nano-Descriptors: Carbon Nanotubes as Mitochondrial F0F1-ATPase Inhibitors. *Journal of Chemical Information and Modeling*, *59*(1), 86–97. doi:10.1021/acs.jcim.8b00631 PMID:30408958

González-Durruthy, M., Monserrat, M. J., Rasulev, B., Casañola-Martín, M. G., Barreiro Sorrivas, M. J., Paraíso-Medina, S., ... Munteanu, R. C. (2017). Carbon Nanotubes' Effect on Mitochondrial Oxygen Flux Dynamics: Polarography Experimental Study and Machine Learning Models using Star Graph Trace Invariants of Raman Spectra. *Nanomaterials (Basel, Switzerland)*, *7*(11), 386. doi:10.3390/nano7110386 PMID:29137126

Jones, D. E., Ghandehari, H., & Facelli, J. C. (2016). A review of the applications of data mining and machine learning for the prediction of biomedical properties of nanoparticles. *Computer Methods and Programs in Biomedicine*, *132*, 93–103. doi:10.1016/j.cmpb.2016.04.025 PMID:27282231

Liu, Z., Ling, X. Y., Su, X., & Lee, J. Y. (2004). Carbon-Supported Pt and PtRu Nanoparticles as Catalysts for a Direct Methanol Fuel Cell. *The Journal of Physical Chemistry B*, *108*(24), 8234–8240. doi:10.1021/jp049422b

Muller, R. H., & Keck, C. M. (2004). Challenges and solutions for the delivery of biotech drugs – a review of drug nanocrystal technology and lipid nanoparticles. *Journal of Biotechnology*, *113*(1), 151–170. doi:10.1016/j.jbiotec.2004.06.007 PMID:15380654

Novoselov, K. S., Fal′ko, V. I., Colombo, L., Gellert, P. R., Schwab, M. G., & Kim, K. (2012). A roadmap for graphene. *Nature*, *490*, 192. doi:10.1038/nature11458 PMID:23060189

Odziomek, K., Ushizima, D., Oberbek, P., Kurzydłowski, K. J., Puzyn, T., & Haranczyk, M. (2017). Scanning electron microscopy image representativeness: Morphological data on nanoparticles. *Journal of Microscopy*, *265*(1), 34–50. doi:10.1111/jmi.12461 PMID:27571322

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. *SAR and QSAR in Environmental Research*, *27*(7), 521–538. doi:10.1080/1062936X.2016.1197310 PMID:27329717

Puzyn, T., Leszczynska, D., & Leszczynski, J. (2009). Toward the Development of "Nano-QSARs": Advances and Challenges. *Small*, *5*(22), 2494–2509. doi:10.1002mll.200900179 PMID:19787675

Ragesh, P., Anand Ganesh, V., Nair, S. V., & Nair, A. S. (2014). A review on 'self-cleaning and multifunctional materials'. *Journal of Materials Chemistry. A, Materials for Energy and Sustainability*, *2*(36), 14773–14797. doi:10.1039/C4TA02542C

Rani, K. V., & Jawhar, S. J. (2019). Novel Technology for Lung Tumor Detection Using Nanoimage. *Journal of the Institution of Electronics and Telecommunication Engineers*, 1–15. doi:10.1080/03772063.2019.1565955

Tantra, R., Oksel, C., Puzyn, T., Wang, J., Robinson, K. N., Wang, X. Z., ... Wilkins, T. (2015). Nano(Q)SAR: Challenges, pitfalls and perspectives. *Nanotoxicology*, *9*(5), 636–642. doi:10.3109/17435390.2014.952698 PMID:25211549

Toropova, A. P., & Toropov, A. A. (2017). Nano-QSAR in cell biology: Model of cell viability as a mathematical function of available eclectic data. *Journal of Theoretical Biology*, *416*, 113–118. doi:10.1016/j.jtbi.2017.01.012 PMID:28087422

Winkler, D. A. (2016). Recent advances, and unresolved issues, in the application of computational modelling to the prediction of the biological effects of nanomaterials. *Toxicology and Applied Pharmacology*, *299*, 96–100. doi:10.1016/j.taap.2015.12.016 PMID:26723909

Winkler, D. A., Burden, F. R., Yan, B., Weissleder, R., Tassa, C., Shaw, S., & Epa, V. C. (2014). Modelling and predicting the biological effects of nanomaterials. *SAR and QSAR in Environmental Research*, *25*(2), 161–172. doi:10.1080/1062936X.2013.874367 PMID:24625316

Wu, X., Liu, H., Liu, J., Haley, K. N., Treadway, J. A., Larson, J. P., ... Bruchez, M. P. (2002). Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots. *Nature Biotechnology*, *21*(1), 41–46. doi:10.1038/nbt764 PMID:12459735

This research was previously published in the International Journal of Applied Nanotechnology Research (IJANR), 4(1); pages 1-7, copyright year 2019 by IGI Publishing (an imprint of IGI Global).