Biomaterials Science

ROYAL SOCIETY OF CHEMISTRY

MINIREVIEW

Cite this: *Biomater. Sci.*, 2021, **9**, 1598

Machine learning-integrated omics for the risk and safety assessment of nanomaterials

Farooq Ahmad, *[a] Asif Mahmood [b] and Tahir Muhmood [c]

With the advancement in nanotechnology, we are experiencing transformation in world order with deep insemination of nanoproducts from basic necessities to advanced electronics, health care products and medicines. Therefore, nanoproducts, however, can have negative side effects and must be strictly monitored to avoid negative outcomes. Future toxicity and safety challenges regarding nanomaterial incorporation into consumer products, including rapid addition of nanomaterials with diverse functionalities and attributes, highlight the limitations of traditional safety evaluation tools. Currently, artificial intelligence and machine learning algorithms are envisioned for enhancing and improving the nano-bio-interaction simulation and modeling, and they extend to the post-marketing surveillance of nanomaterials in the real world. Thus, hyphenation of machine learning with biology and nanomaterials could provide exclusive insights into the perturbations of delicate biological functions after integration with nanomaterials. In this review, we discuss the potential of combining integrative omics with machine learning in profiling nanomaterial safety and risk assessment and provide guidance for regulatory authorities as well.

Received 30th September 2020,
Accepted 21st December 2020

DOI: 10.1039/d0bm01672a

rsc.li/biomaterials-science

1. Introduction: the challenges of keeping nanomaterials safe

Nanomaterial safety is a foremost bottleneck in the dissemination of novel nano-enabled products to the market. Unanticipated harm is a main foundation of attrition and concern in nanomaterial dissemination into consumer products including medicines and nanotherapeutics, and participates in needless morbidity and mortality. Adverse nanomaterial reactions are unexpected outcomes exhibited after

[a]*College of Engineering and Applied Sciences, Nanjing National Laboratory of Microstructures, Jiangsu Key Laboratory of Artificial Functional Materials, Nanjing University, Nanjing, Jiangsu, 210093, China. E-mail: ahmadchemist@hotmail.com*
[b]*Beijing Key Laboratory of Photoelectronic/Electrophotonic Conversion Materials, School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing, 100081, China*
[c]*State Key Lab of Metal Matrix Composites, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China*

Farooq Ahmad

Farooq Ahmad received his PhD in Chemical Engineering and Technology (2016) from Zhejiang University of Technology, China, with research focused on nano–bio interaction-mediated nanotoxicity. He is currently a postdoctoral fellow at the College of Engineering and Applied Sciences, Nanjing University, China. His current research interests are artificial intelligence for synthesis, design, characterization, safety, and biomedical evaluation of nanomaterials by integrated metabolomics approach.

Asif Mahmood

Asif Mahmood received his PhD degree from the National Center for Nanoscience and Technology (NCNST), China. Currently, he is a postdoctoral fellow at the School of Chemistry and Chemical Engineering, Beijing Institute of Technology. His research interests include the machine learning and computational analysis of materials.

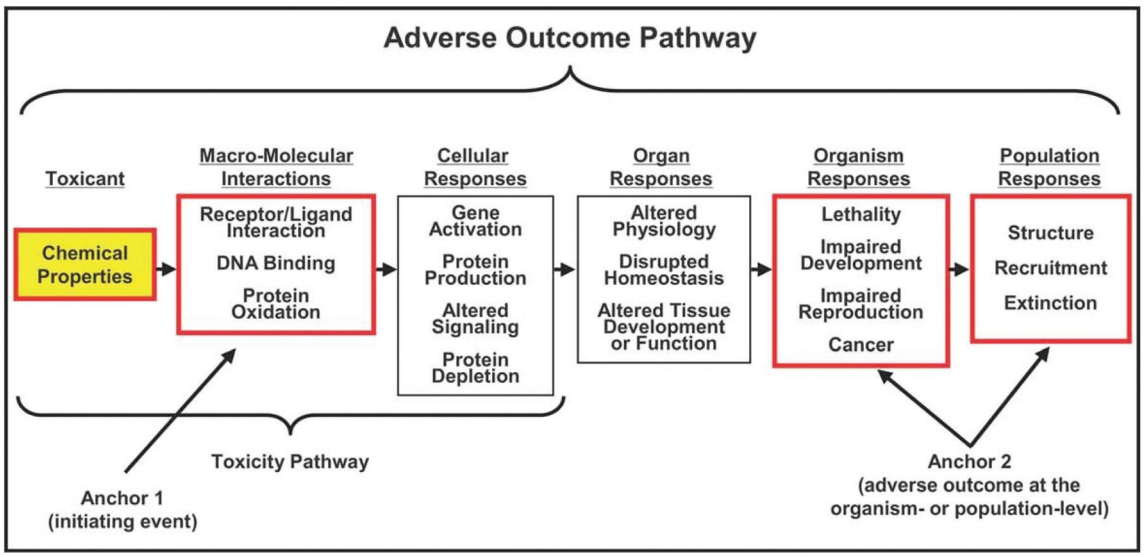

Fig. 1 Illustration of the adverse outcome pathway (AOP) from ref. 9, with permission from Wiley [2009].[9]

exposure to nanomaterials (depending on the physicochemical combinations). There must be a complementary system to report nanomaterial safety prior to its incorporation into products of daily use. Once the nano-enabled product is marketed, the fate of nanomaterials should be monitored through adverse effect (AE) reports/adverse outcome pathways (Fig. 1) to ensure that nanomaterial's (nanomedicine) safety information is up to date, and this is called nanomaterial vigilance (NV). As in the case of nanomedicine, none of such processes are error proofs because of the limitations of clinical trials, as it is impossible to assess all possible synergistic outcomes on a population scale to sense rare AEs. Therefore, there is need to build databases of spontaneously collected AE (from published data/data provided by the regulatory bodies) of nanomaterials to flag leads and accomplish assenting after-effect analyses. Unfortunately, these intelligences again raise concerns linked with under-reporting, and it is troublesome for rare events to occur due to nano–bio, nano-environmental and nano–nano interactions.[1–3] Furthermore, there is a great difference of biology among laboratory animal models and humans, which leads to piercing a hole in the nanomaterial safety and risk assessment.[4–6] Due to aforementioned concerns, researchers have started to use cheminformatics and computational tools to overcome these limitations and complement its NV.[7]

A glimpse of precise nanotoxicological end point can be obtained by integrative omics technologies. Information generated by omics data is more proximal to organisms and not comprehensive, which presents challenges in indicating nanotoxicology end points.[8] Thus, in most cases, no single tool can comprehend the sophistication of molecular episodes that steer AEs mediated by nanomaterials. Ideally, different technologies are promising to aid in the AEs of nanomaterials and to design a holistic depiction of nanomaterial safety and risk assessment. However, hyphenation of multi-omics data for nanomaterial risk and safety assessment is acquainted with the challenges of new informatics and interpretation. These problems can be resolved by the incorporation of latest analytical and statistical tools *via* merging disparate data sets and standardized control metrics as well. Herein, we portray the exploitation of multi-level omics tools in nanomaterial safety and risk assessment (SRA).

Emerging machine learning tools have already kicked off in unleashing the supremacy of big data generated in diverse departments, including omics (big)data. Deep learning algorithms may accomplish the amount of achievement in far too many operations, namely, gaming, computer vision and sound interpretation. There is a lot of praise for introducing AI

Tahir Muhmood

Tahir Muhmood is a post-doctoral research associate at the State Key Laboratory of Metal Matrix Composites, School of Material Sciences and Engineering, Shanghai Jiao Tong University, Shanghai China. He received his PhD from Nanjing University of Science and Technology. He received six "Academic Research Awards" and two "Academic Success Awards". He also wins the Elite Award from CS-Beijing Council, P.R. China. One of his research articles got the VIP title (Very Important Title) from WILEY–VCH Verlag GmbH & Co.KGaA, Weinheim, for publishing the innovative research in the field of carbon-based nano-catalysts. His research expertise includes advanced nanomaterial synthesis, nano-catalysts, photocatalysts, and metallic nanostructure controlling for energy and opto-electronic properties.

methods to toxicity end points of nanomaterials in both academia and industries.[10,11] Furthermore, an unsupervised deep learning algorithm (autoencoder) was magnificently cast off to train a model hinged on multi-omics data to forecast the survival of cancer patients.[12–14] Herein, we traversed the up-to-date literature ascendancy artificial intelligence (AI) tools such as machine learning (ML) and deep learning (DL), on unique repository of data for pre-clinical safety evaluation and post-marketing surveillance for nanomaterials called NV (nano-vigilance). We also invite the readers to switch to the preceding initial articles for even more information on ML[15–17] and DL[18,19] approaches in nano-safety.[20–23] Furthermore, integration of AI and ML with automations has further accelerated the development in omics technologies, including the fast and robust characterization of materials.[8,24] In the case of nanomaterial synthesis, characterization, safety and risk assessment, high-throughput synthesis and characterization techniques along with ML tools have not been widely adopted.[25] Given the fact, these aiding technologies will extensively be used in the coming years for safer nanomaterial design along with post-marketing vigilance.

In this mini review, we propose a framework of adopting ML with omics techniques for high-throughput synthesis, characterization and safety evaluation (pre-marketing and post marketing). Our aim is to provide a framework for an ML-led omics strategy for nanomaterial safety and risk assessment, and major challenges with possible solutions.

1.1. Pre-clinical nanomaterial safety and risk assessment

AI techniques have been exploited widely for the evaluation of preclinical nanomaterial safety and risk assessment. AEs of nanomaterials on living objects including humans, animals, plants, algae, and the environment present at different levels of food chain can be resolved by fixing careful designing of nanomaterials.[26] Therefore, premarketing (preclinical) assessments are necessary for avoiding the toxic nanomaterial (nanomedicine) incorporation into consumer products. Higher nanotoxicity in the environmental matrices and during clinical trials is still the major cause limiting the use of nanomaterials in consumer products, especially in medicine. Therefore, robust toxicity evaluations are compulsory for assessing nanomaterial safety and minimizing the price and dissemination time of nano-enabled products to the market. Conventionally, animal studies have long been used for safety and risk assessment,[26–28] but are impeded by price, time, and ethical concerns. Various computational techniques have proved to be excellent tools in evaluating the safety of nanomaterials. These AI-assisted ML techniques forecast AEs by monitoring the multiple descriptors of nanomaterials by utilizing target-oriented models and quantitative structure–activity relationships (QSARs).

2. Machine learning (ML)

Machine learning generally involves the exploitation of QSARs, regression, support vector machines (SVM) and ensemble learning techniques to process the data to extract useful information for nanomaterial safety evaluation. QSAR is a tool for quantitative establishment of relationship among chemical or structural attributes and potential pharmacological activity.[29] QSAR techniques have been extensively exploited for modeling various nanomaterial safety and risk assessment by exploring the various toxicity end points.[30] In particular, a QSAR model has the ability to analyze the associations between nanomaterial properties (size, shape, chemical nature, *etc.*) with different biological activities.[31,32] An excellent model has high predictive power and easily interpreting potential. Recently, various ML tools such as regression (random forest), SVMs, Naïve Bayesian, k-nearest neighbors (KNN), ensemble learning and deep learning have been used for QSAR modeling. As we have mentioned earlier, QSARs can be exploited in predicting the targeted safety and risk assessment of nanomaterials. In addition, the targeted nanotoxicity predictive models can only be performed with proprietary tools. Most of them utilized ML algorithms specifically designed for predicting toxicity end points. However, TargeTox[33] and PrOCTOR[34] are the open access tools for evaluating the toxicity.

To determine the chemical attributes of drug discovery, early QSAR tools focused on multivariate linear regression.[35] These methods are susceptible to high complexity of data, and the similarity of features leads to collinearity and reduced interpretability. Current regression-based methods use function discovery to answer those issues. Though regression-based methods have shown usefulness in QSAR prediction, linearity conventions and dimensionality complications are intrinsic in regression, and are challenging in QSAR modeling. These complications can be overcome by widely accessible SVMs and ensemble learning methods (*e.g.*, random forests), due to their high predictability, precision, reliability and simplicity of interaction.[36] SVMs use a hyperplane in an n dimensional space (n = number of descriptors) that distinctively categorizes the data. For instance, if there are two contributing descriptors, the hyperplane will be a line. With three contributing descriptors, the hyperplane will be a 2D plane.[37] Support vectors are data spots utilized to shape the SVM, and expand the margin of categorization, and data spots sited close to the hyperplane influence the orientation and disposition of this hyperplane. SVMs execute classification by exploiting kernel algorithm to plot vectors into a sophisticated dimension characteristic space.[38] Ensemble approaches incorporate a variety of ML variables into a versatile statistical model. In doing so, they have increased the predictive efficiency relative to a standard model. It also reduces the biasness and over-fitting of the data. Random forest is an ensemble learning method to eliminate the class imbalances and prevent overfitting in a QSAR framework by constructing a multitude of random decision forests.[39]

2.1. Deep learning (DL)

DL is a subcategory of machine learning that uses a hidden and complex network of ANNs (artificial neural networks) to extract useful information from raw data with robust predic-

tion potential, as compared to and without human interferences. Therefore, automatic generation of models in speech recognition, extracting image features, and complex decision making without human/expert input have notably transformed the fields of science and technology.[40,41] Basically, there are three models that are most commonly used: deep neural networks (DNNs), convolution neural networks (CNNs) and recurrent neural networks (RNNs).

DNNs constitute fully linked artificial neural networks, which include multilayer perceptron (MLP),[42] auto-encoder[43] and restricted Boltzmann machine (RBM).[44,45] MLP consists of input and output layers and multiple hidden layers between them as shown in Fig. 2. MLP exploit a large amount of data for training by adjusting the weights amid two neurons *via* back propagating algorithm for establishing the accurate network. Therefore, training an MLP is typically executed by a supervised method for a large amount of data. MLP is widely used when features are not related to time or space in omics research.

A recurrent neural network (RNN) is a class of ANN where connections between units form a directed cycle (Fig. 2). This creates a type of neural network that can model the dynamic temporal behavior. Unlike feed-forward neural networks, RNNs can use their internal memory to process arbitrary sequences of inputs.[46] RNNs have been progressively useful in countless arenas including natural language, speech and image processing and recognition. In omics research, RNNs have innumerable utilization including the discovery of the exon/intron margins of a gene and predicting RNA sequence-specific bias. RNNs are called so because the input of the hidden layer comprises not only the output of the input layer but also the output of the hidden layer at the former most moment.[47] A simple RNN algorithm can be scaled up into a multifaceted intricate network. A convolutional neural network (CNN) is a deep learning multilayer feed forward neural network architecture built *via* stacking many hidden convolutional layers consisting of activation layers and pooled layers arranged one above the other in sequence and a full connection layer (Fig. 2). CNNs have found applications in speech, face and general object, recognition, motion analysis and natural language processing. In addition, CNNs play a vital role in omics research, as well as gene expression prediction, protein classification and gene structure prediction.[48]

In short, these neural networks advantageously perform robust analysis of largely complex data for drug discovery by building QSAR models to predict activity profiles and drug target interactions, binding constants for different enzymes,[3] toxicity profiles, ADMET and physicochemical properties, characteristics of drug delivery systems, *etc.*, as well as performing virtual screening.[49] However, we should not underestimate the potential of deep learning technology over the traditional QSAR/QSPR analysis of small data sets with a limited number of descriptors. It is anticipated that neural networks will be more widely used in drug discovery in the future and applied in non-traditional areas such as bridging the gap between animal and human studies, drug delivery systems, biocompatible materials, and nanomedicine.[46,50]

2.2. Applications of DL in nano safety assessment

We need to establish reliable ML tools to find nanomaterial descriptors, which can correlate the *in vitro* studies for predicting the biological outcome/toxicity endpoints to *in vivo* studies. For example, an ML-assisted high-throughput method

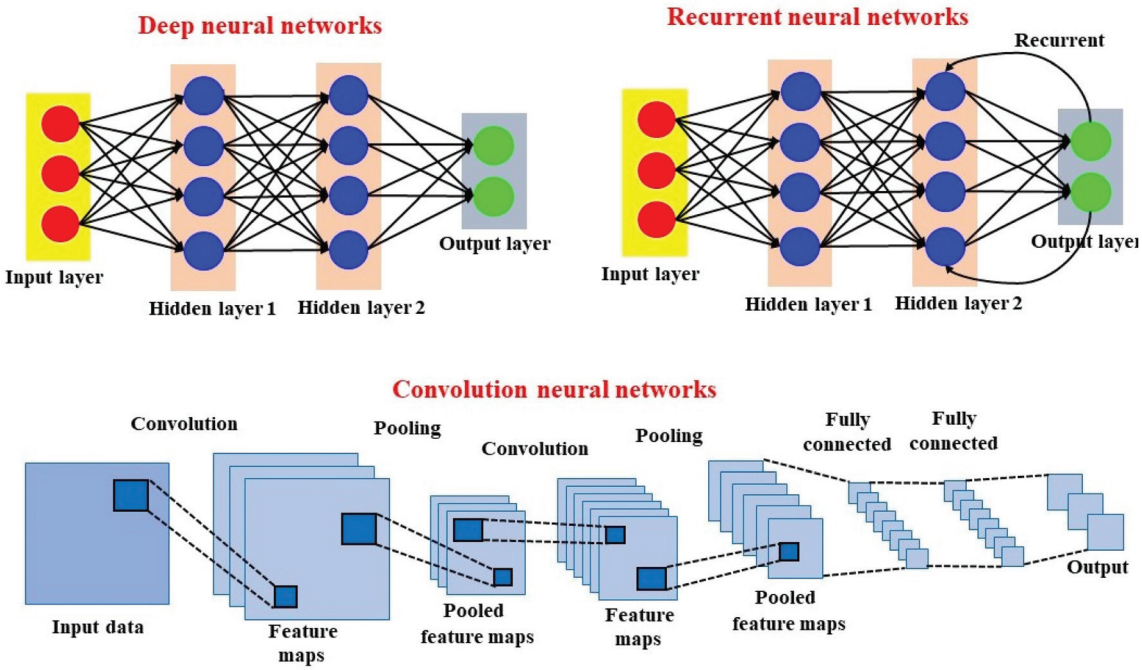

Fig. 2 Demonstrates deep neural networks (DNNs), convolution neural networks (CNNs) and recurrent neural networks (RNNs).

has been recently established to understand the mechanism and trends of the protein corona formation onto the nanomaterial and predict the nanoparticle behavior in the human body.[51,52] To formulate a predictive model from small data sets, genetic programming-based decision tree (GPTree) is a robust tool for automatic selection of descriptors and interpreting the properties of NP.[53] Recently, Sizochenko *et al.* have developed a multi-nano-read-across modeling system with self-organizing maps to predict the toxicity of metal oxides (184 NPs) and silica (15 NPs) nanoparticles data sets after incubation with bacteria, protozoa, algae, and mammalian cell lines. They integrated interspecies correlation analysis with a self-organizing plot to recognize the parameters mediating the nanotoxicity.[54] Another study arranged 260 metals and metal oxides, and 31 different compositions of silica NPs from data mining, indicating how descriptors of these NPs instigate the adverse biological outcomes.[55] Wang *et al.* developed a library for high-throughput synthesis of biocompatible/bioactive NPs for intelligently reducing the cost of nanomaterial discovery.[56] Lazarovits *et al*. used ML to uncover the fate and mechanisms of blood protein adsorption (*in vivo*) onto NPs after intravenous injection constantly changing blood flow. With the incorporation of ML technique, they designed the NPs with ~70% less liver and spleen accumulation.[57] A highly predictive model based on the neural network demonstrated the toxicity of iron oxide (based on the particle size and concentration, incubation time, and surface charge) NPs in kidney cells.[58] The incorporation of ML methods into "nano safety and risk assessment" requires establishment and cooperation among different labs and computational researchers, and policy makers. To fulfil those aims, EU launched various projects and actions through MODENA, MARINA, NANOSOLUTIONS and EU Horizon 2020 projects, Nano-SolveIT (https://nanosolveit.eu/),[59] SABYDOMA (https://www.bnn.at/projects/sabydoma), and NanoCommons (https://www.nanocommons.eu/). These all projects are assumed to provide tools for the development of training ML models for high-throughput synthesis and characterization of nanomaterials.

3. Postmarketing surveillance of nanomaterials

After the dreadful revelation in 1962 that a mild sleeping pill (thalidomide) is the cause of malformed limbs in thousands of new born babies without having any side effects on their mothers,[60] the World Health Organization (WHO) launched an international drug monitoring program (WHO Program for International Drug Monitoring). Moreover, breast implants from a French company, Poly Implant Protheses (PIP), showed a significant rate of failures due to the use of low-molecular weight silicone gel implants, leading to breast implant illness.[61] All these above-mentioned events highlighted the irregularities of the current regulatory framework in terms of preventing harmful products to the consumer. Considering the serious concerns arising from the extensive use of nanomaterials in consumer products, the European Union (EU) formulated the state-of-the-art computational methods (QSAR for modeling and predicting) for evaluating the safety and risk assessment of nanomaterials (toxicokinetic and toxicodynamic, dosimetry and biological fate) to provide guidance to the REACH regulation.[21] Furthermore, EU have formulated a system of regulating new medical devices incorporating nanomaterials for "high or medium" potential of AEs on humans and the environment.[62,63]

4. Integrated omics for nanotoxicity evaluation

System toxicology elaborates the nanomaterial action from the perspective of system biology, assessing the AEs of nanomaterial on the complete biological system rather than single targets, enzymes, metabolites/transcription factors, *etc*. This technique unveils the unknown nanomaterial AE resulting from the intricate interplay of various toxicological end points (targets and pathways). Applications of system nanotoxicology (safety and risk assessment) to adverse nanomaterial (nanomedicine) actions varies from its application in drug discovery, as it primarily provides the potential for off-targeting mediated adverse effects (from the clinical point of view).

Exposure of cells or species to substances may cause a sequence of functional pathway consequences resulting in changes in the concentrations, associations, and reinforcement channels of different sorts of biomolecules. A single omics tool, *e.g.* transcriptomics, can recognize one form of biomolecule and will therefore catch only variations in a specific subset of the biological cascade. Thus, while the application of single omics analysis can contribute to the detection of biomarkers for some contaminants, and lack a systemic reasoning of toxicity pathways of adverse effect.[64] The incorporation of several omics data sets ensures a major increase in the identification of this pathway reaction to a contaminant, by-knowledge as such and, in particular, by a system biology/understanding. The goal of the integrated omics model is to detect molecular mechanisms, group samples, and predict consequences (Fig. 3). To deduce the intricate biological process connected to nanomaterial-mediated adverse outcomes, data collected from multiple etymologies should be integrated together followed by expert analysis. This system biology technique to understand omics perturbations resulting in various toxicological end points including hypertension, or response to a particular treatment, is challenging; nonetheless, it is one of the most pertinent problems confronting computational scientists today.

Multi-omics information is scattered over numerous repositories and is linked with the corresponding data sets. A multi-omics (genomics and epigenomics) was used for profiling DNA methylation, mRNA and microRNA expression in A549, BEAS-2B and THP-1 at lower exposure (48 h) concentrations of carbon nanomaterials.[65] Another study investigated the adverse effects of nanomaterials (SiO_2 and Mn_2O_3) through the

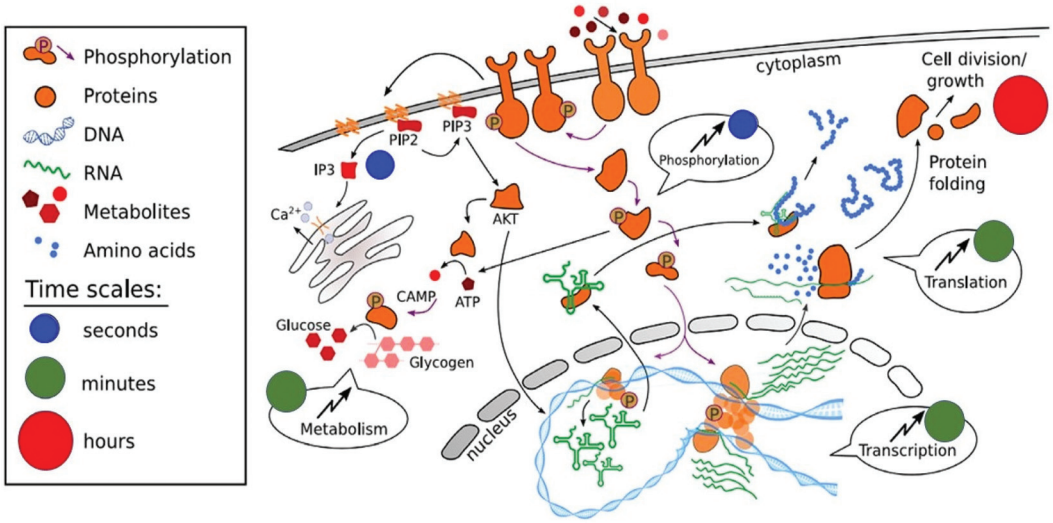

Fig. 3 Depiction of various biomolecules involved in response to an event extending to various omics layers. Reproduced and adapted with permission from ref. 64, Springer-Verlag GmbH Germany[2020].[64]

(multi-omics) untargeted proteomics, untargeted metabolomics, src homology2 (SH2) profiling at doses of 2.5 to 10 μg cm^{-2} after 24-hour exposure. Among which SiO_2 NPs induced strong adverse effects.[66] Besides the above-mentioned studies, one of the biggest problems is that multi-omics data are typically not cross-indexed amid repositories making it a laborious task. Because in most cases, omics data sets are not properly referenced/indexed on deposition in public repositories.[64]

5. Machine learning hyphenation with omics for nanomaterial safety and risk assessment

A conceptual framework has been proposed for pre- and post-marketing surveillance through integrating the machine learning with omics techniques for nanomaterial safety and risk assessment, as shown in Fig. 4. Omics data are processed by machine learning algorithms to characterize endotypes for autonomous comparison of safety and risk assessment of nanomaterials for preclinical safety assessment and post-marketing vigilance and decision making. Furthermore, the current sophistications in the latest microarray technology with subsequent improvements by next-generation sequencing and molecular analysis (*e.g.* gene expression) were rendered on a complete genome-inclusive scale, commonly stated as omics,[67–71] which also helps the dream of the hyphenation of ML with omics come true. However, the diversity and complexity of the high-dimensional nature of the omics data (thousands to millions of genomic features associated with an unlimited number of phenotypes) can be comprehended by utilization of right statistical data processing with deeper and broader understanding,[72] as any discrepancies of statistical method application yield false interpretation (erroneous results) of the data.[73–75] Therefore, it is urgent to incorporate new techniques for omics data analysis and for making useful conclusions. With the potential of stronger gigantic data analysis and feature learning, deep learning manages to execute comparatively fitting experimental results. Therefore, rational implementation of deep learning in omics demands public access to various data bases including encyclopedia of DNA elements[76] and the gene expression omnibus (GEO).[77]

Another challenge that machine learning encounters is that drawbacks pivot on the issue of class prediction, a challenge for constructing a molecular assortment that can forecast a phenotype or result of interest (*e.g.* disease detection and prognosis). Traditional omics discovery analysis can conduct molecular measurements to extract a molecular classification algorithm. The reason behind the above-mentioned challenges is that a characteristic omics 'discovery' study may execute molecular analysis (say gene expression) on a relatively huge number p ($p \approx 10^5$–10^7) of variables (genes, regulatory elements) and, due to cost or logistical reasons, on a fairly smaller number (n) of samples ($n \approx 100$–1000).[78–80] Therefore, machine learning can be used to overcome the gaps in omics science.

There are various approaches present for the integrative analysis of multi-omics data.[81–83] We will emphasis predominantly on the application of those approaches for sensing an AE of particular pathway in the toxicological settings. Bersanelli *et al.* devised a simultaneous and sequential statistical integration approach, based on which they discriminate network-based *versus* network-free algorithms applicable to interconnected networks to model inconsistent interactions and utilize the Bayesian *vs.* non-Bayesian algorithms to compute the probability distribution on the measured omics

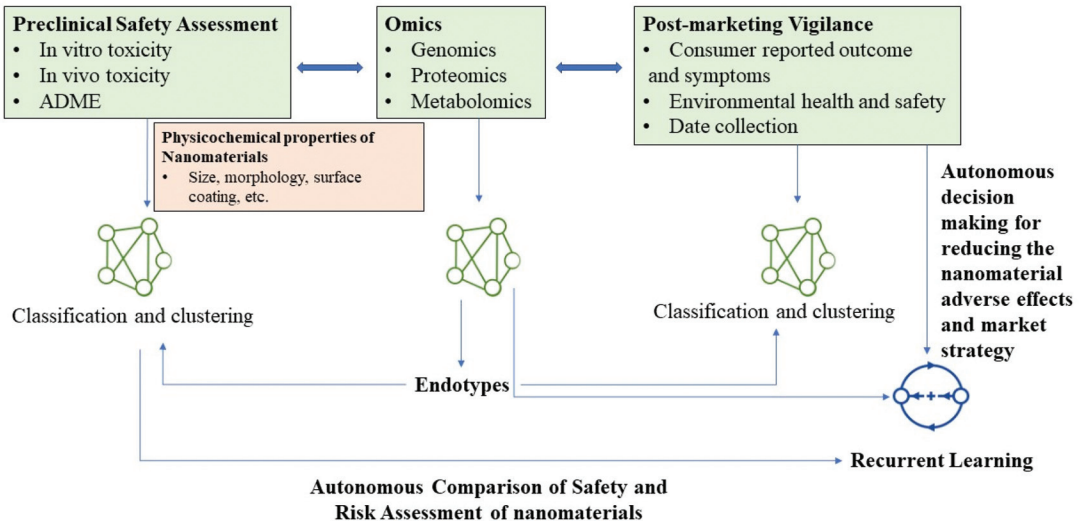

Fig. 4 A conceptual framework for the future use of machine learning in nanomaterial safety and risk assessment.

data.[84] Another group classifies the integration techniques established on the assimilation of additional biological evidence by means of a supervised or an unsupervised tactics. The first plans at forecasting a convinced response variable exploiting the omics attributes as predictors, or at modeling a regulatory network of the fundamental molecular system. The latter is primarily exploited to retrieve an exploratory synopsis of the data and to unravel latent associations amid omics layers.[81]

Currently, deep learning tools have been successfully applied to forecast and classify the purposeful segments in DNA patterns, together with replication domain, transcription factor binding site (TFBS), transcription initiation point, promoter, enhancer and gene deletion site. Recently, a pre-trained DNN and a hidden Markov hybrid model were developed for the identification of discreet replication segments based on the types with high accuracy and robustness.[85] Furthermore, classification of the genome sequences on their basis of TFBS with a median area under the curve value of 0.946 was achieved by developing a deep convolutional network (CNN).[86] A CNN-based algorithm was used to develop predictive models by the identification, classification and analysis of prokaryotic and eukaryotic promoter and non-promoter sequences with great accuracy (0.90 on TATA and 0.89 on non-TATA promoter sequences).[48] Similarly, a deep three-layer feed forward neural network was used effectively to distinguish functional enhancers and promoters in GM12878 lymphoblastic cells with maximum accuracy of 93.59%.[87] Furthermore, Wang *et al.* developed a highly accurate and sensitive CNNdel tool (based on shallow CNN) to detect genomic deletions by comparing the real data from 1000 Genomes Project.[88] Typically, DNA sequence data are widely included as primary training data for predicting and distinguishing core functions in DNA sequences.

Furthermore, DL can also be used for predicting RNA architecture including RBP binding sites, possible slicing sites and RNA types. For example, a deep belief neural network (DBNN) algorithm, developed by training from input data of RNA secondary and tertiary structure information, can be used to discover potential binding motifs and binding sites.[89] A novel slice junction classification algorithm (DeepSplice) was developed based on deep CNN with improved accuracy, computational efficiency and flexibility.[90] Similarly, DL tools can indeed classify protein structures, namely, protein secondary and tertiary structure prediction, protein structure consistency evaluation and protein interaction map prediction, *etc*. Based on the DL algorithm developed to identify the amino acid sequence and secondary/tertiary structure of protein, backbone torsion angles with other features are an iterative input. The replacement of SVM with DNN significantly improves the protein model predictability, indicated by the enhancement in Pearson's correlation coefficient up to 5%.[48] An ultra DNN (coupling two deep residual neural networks) to predict associates by assimilating together sequence conservation information and evolutionary coupling achieves the highest F1 score on free-modeling targets in the latest critical assessment of protein structure prediction (CASP).[91]

5.1. Nanotoxicological corpora

In addition to standardized data, which pose both standardization and visualization problems and may not be widely available, a critical tool is the biomedical and clinical organization. Nanotoxicological corpora require practicing toxicological neural networks that integrate previous expertise and assist with context-specific usage semantics of some terms. Throughout this field, Natural Language Processing (NLP) techniques are indispensable to harvest perceptions and principles and incorporate ML or start making decision explicitly from inserting representations of certain documents. A trained KNN algorithm with the exploring and considerate adverse nanomaterial reactions (ANRs) by integrative mining of toxicological records to national networks with a generalized ML

model provides a ground for decent generalization performances to forecast binary ANR outcomes.

Along with the advancement of word embeddings, sequence-to-sequence (seq2seq) learning and, now, attention mechanisms, NLP has prospered substantially from the advancement in DL from subsequent years to develop improved language frameworks.[92–95] The toxicological discipline for NLP has also been a difficult area of research, and these new approaches have been introduced efficiently to NV challenges. Language neural networks can be trained utilizing toxicological literature and thereafter implemented to the safety and risk assessment of nanomaterials, where the authors used PubMed to learn how to better describe theories observed in descriptive databases for the classification of ANEs (adverse nanomaterial effects).

NLP involves the name entity recognition (NER) followed by deep learning process for the identification of nanomaterial to AE interaction prediction. These both tasks (multitasking) can be executed by inherent structure of neural networks and shared weights. DL multitasking performance was improved by sharing the multitask comprising shared neural networks of different multi tasks.[96,97] Furthermore, NLP practices have also been exploited with data gathered through social media and online health communities.[98,99] Therefore, the data collected through these social media platforms yield high noise data.[100] Postmarketing NV has been directed by means of twitter data[101,102] with embedding techniques and biLSTM deep classifiers that beat provisional random field tactics, discussions forums,[103] or more domain-specific health social networking sites.[104]

6. Conclusion and future perspectives

Deep learning tool is undoubtedly fit for undertaking omics complications. The synergy of omics with DL is still at its infancy. Therefore, we outlined the relevant recent work and provided the guideline for this area. In addition, the publicly available data with the development of ML and DL techniques has brought a significant transformation in the field of NV. This article emphasizes on synergistic adoption of AI techniques with omics data for exploring the safety and risk assessment of pre- and post-marketing monitoring of nanomaterials. We have also evidenced an increasing integration of multitude of datasets—from molecular to clinical—leveraging DL to explore the nanotoxicity end points by complex interplay of various biological systems. All omics investigations are big data enterprises, requiring large amounts of digital storage, immense computing power, and the statistical expertise necessary to work with large data sets.[105]

How can deep learning tools transform the narrative of drug discovery and safety in pre- and post-clinical settings? Because, the naïve use of machine learning techniques to omics data will transcend to overfitting and represent false biological alterations linked with the phenotype of concern.[106] Furthermore, the time between sample collection and storage, and the type of surgery or protocol used to collect blood or buccal swabs may affect molecular measurements, yet these factors may not have been recorded or are not publicly available.[107] In short, omics data science is getting older, and universally accepted guidelines should be formulated to avoid spreading any misinformation and to steer the field in the right direction. The multi-omics field of nanotoxicity is relatively new, and demands increased the attention to leveraging the booming benefits in the technological prospects of omics-based platforms with state-of-the-art tactics for the discovery of unseen biological mechanisms under the influence of nanomedicine.

ML algorithms, including DL methods, have enabled the utilization of AI in the industrial settings and in day-to-day life by presenting new vistas of complex problem resolution in the blink of an eye. In particular, ML algorithms have a superlative ability to predict the unseen nodes of toxicity with ignored attributes of nanomaterials or nanomedicines within complex biological and environmental settings, which can be ultimately exploited by regulatory authorities in policy making to regulate the insemination of nanomaterials by exposing and averting possible adverse effects and improve the health care system after amalgamation with drug discovery and manage electronic health records.[108,109] Hence, various pharmaceutical companies have taken the initiative to indulge ML for drug discovery and development, but facing the lack of interpretability mechanism of the trained neural networks. Therefore, we do not know how neural networks achieved a particular outcome, and hence, we cannot build trust for continuous use in drug discovery. Given the current pace of ML advancements, we still have to witness how ML algorithms can change the game of preclinical drug safety.

A key point that we emphasized throughout this mini review, that is, the narrative of bringing AI to medicine, is just the beginning. There has been remarkably little prospective validation for tasks that machines could perform to help toxicologist, regulatory bodies and clinicians to predict the potential outcomes, leading to development and improvement of health systems through regulatory and policy implementation and decision making. Overall, the risk of implementing AI approaches for nano-vigilance is low with gigantic opportunities as it may have a positive impact on health care.

Conflicts of interest

Authors declare no competing interest.

Acknowledgements

This work was financially supported by the grant from the National Natural Science Foundation of China (No. 21950410533).

References

1 L. G. Soeteman-Hernandez, M. D. Apostolova, C. Bekker, S. Dekkers, R. C. Grafström, M. Groenewold, Y. Handzhiyski, P. Herbeck-Engel, K. Hoehener, V. Karagkiozaki, S. Kelly, A. Kraegeloh, S. Logothetidis, C. Micheletti, P. Nymark, A. Oomen, T. Oosterwijk, I. Rodríguez-Llopis, S. Sabella, A. Sanchez Jiménez, A. J. A. M. Sips, B. Suarez-Merino, I. Tavernaro, J. van Engelen, S. W. P. Wijnhoven and C. W. Noorlander, *Mater. Today Commun.*, 2019, **20**, 100548.

2 P. Mokaberi, F. Babayan-Mashhadi, Z. Amiri Tehrani Zadeh, M. R. Saberi and J. Chamani, *J. Biomol. Struct. Dyn.*, 2020, 1–20, DOI: 10.1080/07391102.2020.1766570.

3 A. Sharifi-Rad, J. Mehrzad, M. Darroudi, M. R. Saberi and J. Chamani, *J. Biomol. Struct. Dyn.*, 2020, 1–15, DOI: 10.1080/07391102.2020.1724568.

4 N. Shakibapour, F. Dehghani Sani, S. Beigoli, H. Sadeghian and J. Chamani, *J. Biomol. Struct. Dyn.*, 2019, **37**, 359–371.

5 P. Mokaberi, V. Reyhani, Z. Amiri-Tehranizadeh, M. R. Saberi, S. Beigoli, F. Samandar and J. Chamani, *New J. Chem.*, 2019, **43**, 8132–8145.

6 S. J. Morgan, C. S. Elangbam, S. Berens, E. Janovitz, A. Vitsky, T. Zabka and L. Conour, *Toxicol. Pathol.*, 2012, **41**, 508–518.

7 D. W. Hobson, *WIREs Nanomed. Nanobiotechnol.*, 2009, **1**, 189–202.

8 K. J. Karczewski and M. P. Snyder, *Nat. Rev. Genet.*, 2018, **19**, 299–310.

9 G. T. Ankley, R. S. Bennett, R. J. Erickson, D. J. Hoff, M. W. Hornung, R. D. Johnson, D. R. Mount, J. W. Nichols, C. L. Russom, P. K. Schmieder, J. A. Serrrano, J. E. Tietge and D. L. Villeneuve, *Environ. Toxicol. Chem.*, 2010, **29**, 730–741.

10 H. Chen, O. Engkvist, Y. Wang, M. Olivecrona and T. Blaschke, *Drug Discovery Today*, 2018, **23**, 1241–1250.

11 P. Mamoshina, A. Vieira, E. Putin and A. Zhavoronkov, *Mol. Pharm.*, 2016, **13**, 1445–1454.

12 A. Yala, C. Lehman, T. Schuster, T. Portnoi and R. Barzilay, *Radiology*, 2019, **292**, 60–66.

13 K. Chaudhary, O. B. Poirion, L. Lu and L. X. Garmire, *Clin. Cancer Res.*, 2018, **24**, 1248–1259.

14 G. Liu, C. Dong and L. Liu, *PLoS One*, 2016, **11**, e0165457.

15 A. Rajkomar, J. Dean and I. Kohane, *N. Engl. J. Med.*, 2019, **380**, 1347–1358.

16 Y. Baştanlar and M. Özuysal, in *miRNomics: MicroRNA Biology and Computational Analysis*, Springer, 2014, pp. 105–128.

17 G. Handelman, H. Kok, R. Chandra, A. Razavi, M. Lee and H. Asadi, *J. Intern. Med.*, 2018, **284**, 603–619.

18 R. Miotto, F. Wang, S. Wang, X. Jiang and J. T. Dudley, *Briefings Bioinf.*, 2018, **19**, 1236–1246.

19 C. Xiao, E. Choi and J. Sun, *J. Am. Med. Inf. Assoc.*, 2018, **25**, 1419–1428.

20 L. Lamon, D. Asturiol, A. Vilchez, R. Ruperez-Illescas, J. Cabellos, A. Richarz and A. Worth, *Comput. Toxicol.*, 2019, **9**, 143–151.

21 K. A. A. Worth, D. Asturiol Bofill, J. Bessems, K. Gerloff, R. Graepel, E. Joossens, L. Lamon, T. Palosaari and A. Richarz, *JRC TECHNICAL REPORTS*, Joint Research Centre, Italy, European Union, Luxembourg, Itally, 2017.

22 I. Furxhi, F. Murphy, M. Mullins, A. Arvanitis and C. A. Poland, *Nanomaterials*, 2020, **10**, 116.

23 C. Oksel, C. Y. Ma, J. J. Liu, T. Wilkins and X. Z. Wang, *Particuology*, 2015, **21**, 1–19.

24 T. E. H. Allen, A. J. Wedlake, E. Gelžinytė, C. Gong, J. M. Goodman, S. Gutsell and P. J. Russell, *Chem. Sci.*, 2020, **11**, 7335–7348.

25 J. M. Cole, *Acc. Chem. Res.*, 2020, **53**, 599–610.

26 A. B. Raies and V. B. Bajic, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2016, **6**, 147–172.

27 I. J. Onakpoya, C. J. Heneghan and J. K. Aronson, *Crit. Rev. Toxicol.*, 2016, **46**, 477–489.

28 M. D. Segall and C. Barber, *Drug Discovery Today*, 2014, **19**, 688–693.

29 K. Roy, S. Kar and R. N. Das, *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*, Academic press, 2015.

30 G. Patlewicz and J. M. Fitzpatrick, *Chem. Res. Toxicol.*, 2016, **29**, 438–451.

31 Z. Y. Algamal, M. H. Lee, A. M. Al-Fakih and M. Aziz, *J. Chemom.*, 2015, **29**, 547–556.

32 G. Basei, D. Hristozov, L. Lamon, A. Zabeo, N. Jeliazkova, G. Tsiliki, A. Marcomini and A. Torsello, *NanoImpact*, 2019, **13**, 76–99.

33 A. Lysenko, A. Sharma, K. A. Boroevich and T. Tsunoda, *Life Sci. Alliance*, 2018, **1**, e201800098.

34 K. M. Gayvert, N. S. Madhukar and O. Elemento, *Cell Chem. Biol.*, 2016, **23**, 1294–1301.

35 J. M. Luco and F. H. Ferretti, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 392–401.

36 J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl and V. Svetnik, *J. Chem. Inf. Model.*, 2015, **55**, 263–274.

37 Y.-C. Lo, S. E. Rensi, W. Torng and R. B. Altman, *Drug Discovery Today*, 2018, **23**, 1538–1546.

38 J. Nalepa and M. Kawulok, *Artif. Intel. Rev.*, 2019, **52**, 857–900.

39 T. G. Dietterich, *The handbook of brain theory and neural networks*, 2002, vol. 2, pp. 110–125.

40 Y. LeCun, Y. Bengio and G. Hinton, *Nature*, 2015, **521**, 436–444.

41 M. Hägele, P. Seegerer, S. Lapuschkin, M. Bockmayr, W. Samek, F. Klauschen, K.-R. Müller and A. Binder, *Sci. Rep.*, 2020, **10**, 6423.

42 D. Svozil, V. Kvasnicka and J. í. Pospichal, *Chemom. Intell. Lab. Syst.*, 1997, **39**, 43–62.

43 G. E. Hinton and R. R. Salakhutdinov, *Science*, 2006, **313**, 504–507.

44 X. Wang, Y. Zhao and F. Pourpanah, *Int. J. Mach. Learn. Cybern.*, 2020, **11**, 747–750.

45 Y. Zhong, W. Chen, Z. Wang, Y. Chen, K. Wang, Y. Li, X. Yin, X. Shi, J. Yang and K. Li, *Comput. Networks*, 2020, **169**, 107049.

46 I. I. Baskin, D. Winkler and I. V. Tetko, *Expert Opin. Drug Discovery*, 2016, **11**, 785–795.

47 W. Y. Goh, C. P. Lim, K. K. Peh and K. Subari, *Neural Comput. Appl.*, 2002, **10**, 311–317.

48 R. K. Umarov and V. V. Solovyev, *PLoS One*, 2017, **12**, e0171410.

49 A. A. Buglak, A. V. Zherdev and B. B. Dzantiev, *Molecules*, 2019, **24**, 4537.

50 E. N. Muratov, J. Bajorath, R. P. Sheridan, I. V. Tetko, D. Filimonov, V. Poroikov, T. I. Oprea, I. I. Baskin, A. Varnek, A. Roitberg, O. Isayev, S. Curtalolo, D. Fourches, Y. Cohen, A. Aspuru-Guzik, D. A. Winkler, D. Agrafiotis, A. Cherkasov and A. Tropsha, *Chem. Soc. Rev.*, 2020, **49**, 3525–3564.

51 Y. Duan, R. Coreas, Y. Liu, D. Bitounis, Z. Zhang, D. Parviz, M. Strano, P. Demokritou and W. Zhong, *NanoImpact*, 2020, **17**, 100207.

52 Z. Ban, P. Yuan, F. Yu, T. Peng, Q. Zhou and X. Hu, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 10492–10499.

53 C. Oksel, D. A. Winkler, C. Y. Ma, T. Wilkins and X. Z. Wang, *Nanotoxicology*, 2016, **10**, 1001–1012.

54 N. Sizochenko, A. Mikolajczyk, K. Jagiello, T. Puzyn, J. Leszczynski and B. Rasulev, *Nanoscale*, 2018, **10**, 582–591.

55 R. Concu, V. V. Kleandrova, A. Speck-Planche and M. N. D. S. Cordeiro, *Nanotoxicology*, 2017, **11**, 891–906.

56 W. Wang, A. Sedykh, H. Sun, L. Zhao, D. P. Russo, H. Zhou, B. Yan and H. Zhu, *ACS Nano*, 2017, **11**, 12641–12649.

57 J. Lazarovits, S. Sindhwani, A. J. Tavares, Y. Zhang, F. Song, J. Audet, J. R. Krieger, A. M. Syed, B. Stordy and W. C. W. Chan, *ACS Nano*, 2019, **13**, 8023–8034.

58 F. Hataminia, Z. Noroozi and H. Mobaleghol Eslam, *Toxicol. in Vitro*, 2019, **59**, 197–203.

59 A. Afantitis, G. Melagraki, P. Isigonis, A. Tsoumanis, D. D. Varsou, E. Valsami-Jones, A. Papadiamantis, L.-J. A. Ellis, H. Sarimveis, P. Doganis, P. Karatzas, P. Tsiros, I. Liampa, V. Lobaskin, D. Greco, A. Serra, P. A. S. Kinaret, L. A. Saarimäki, R. Grafström, P. Kohonen, P. Nymark, E. Willighagen, T. Puzyn, A. Rybinska-Fryca, A. Lyubartsev, K. Alstrup Jensen, J. G. Brandenburg, S. Lofts, C. Svendsen, S. Harrison, D. Maier, K. Tamm, J. Jänes, L. Sikk, M. Dusinska, E. Longhin, E. Rundén-Pran, E. Mariussen, N. El Yamani, W. Unger, J. Radnik, A. Tropsha, Y. Cohen, J. Leszczynski, C. Ogilvie Hendren, M. Wiesner, D. Winkler, N. Suzuki, T. H. Yoon, J.-S. Choi, N. Sanabria, M. Gulumian and I. Lynch, *Comput. Struct. Biotechnol. J.*, 2020, **18**, 583–602.

60 J. E. Ridings, in *Teratogenicity Testing*, Springer, 2013, pp. 575–586.

61 C. Onnekink, R. M. Kappel, W. C. Boelens and G. J. M. Pruijn, *Sci. Rep.*, 2020, **10**, 9558.

62 E. Commission, Nando (New Approach Notified and Designated Organisations) Information System, https://ec.europa.eu/growth/tools-databases/nando/. August 2020.

63 E. Commission, Factsheet for healthcare professionals and health institutions, https://ec.europa.eu/docsroom/documents/35963/attachments/1/translations/en/renditions/native, (accessed August 2020, DOI: 10.2873/761339).

64 S. Canzler, J. Schor, W. Busch, K. Schubert, U. E. Rolle-Kampczyk, H. Seitz, H. Kamp, M. von Bergen, R. Buesen and J. Hackermüller, *Arch. Toxicol.*, 2020, **94**, 371–388.

65 G. Scala, P. Kinaret, V. Marwah, J. Sund, V. Fortino and D. Greco, *NanoImpact*, 2018, **11**, 99–108.

66 A. Bannuscher, I. Karkossa, S. Buhs, P. Nollau, K. Kettler, M. Balas, A. Dinischiotu, B. Hellack, M. Wiemann, A. Luch, M. von Bergen, A. Haase and K. Schubert, *Nanotoxicology*, 2020, **14**, 181–195.

67 S. P. Fodor, J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu and D. Solas, *Science*, 1991, **251**, 767–773.

68 M. Schena, D. Shalon, R. W. Davis and P. O. Brown, *Science*, 1995, **270**, 467–470.

69 C. M. Perou, S. S. Jeffrey, M. Van De Rijn, C. A. Rees, M. B. Eisen, D. T. Ross, A. Pergamenschikov, C. F. Williams, S. X. Zhu and J. C. Lee, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9212–9217.

70 D. A. Wheeler, M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, Y.-J. Chen, V. Makhijani and G. T. Roth, *Nature*, 2008, **452**, 872–876.

71 U. Nagalakshmi, Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein and M. Snyder, *Science*, 2008, **320**, 1344–1349.

72 A. E. Teschendorff, *Nat. Mater.*, 2019, **18**, 422–427.

73 J. T. Leek, R. B. Scharpf, H. C. Bravo, D. Simcha, B. Langmead, W. E. Johnson, D. Geman, K. Baggerly and R. A. Irizarry, *Nat. Rev. Genet.*, 2010, **11**, 733–739.

74 A. E. Teschendorff, J. Zhuang and M. Widschwendter, *Bioinformatics*, 2011, **27**, 1496–1505.

75 L. R. Jager and J. T. Leek, *Biostatistics*, 2014, **15**, 1–12.

76 E. P. Consortium, *Science*, 2004, **306**, 636–640.

77 T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman and M. Holko, *Nucleic Acids Res.*, 2012, **41**, D991–D995.

78 V. Vapnik, *The nature of statistical learning theory*, Springer science & business media, 2013.

79 J. Friedman, T. Hastie and R. Tibshirani, *J. Stat. Softw.*, 2010, **33**, 1.

80 S. Webb, *Nature*, 2018, **554**, 555–557.

81 S. Tarazona, L. Balzano-Nogueira and A. Conesa, in *Comprehensive Analytical Chemistry*, ed. J. Jaumot, C. Bedia and R. Tauler, Elsevier, 2018, vol. 82, pp. 505–532.

82 T. Eicher, G. Kinnebrew, A. Patt, K. Spencer, K. Ying, Q. Ma, R. Machiraju and E. A. Mathé, *Metabolites*, 2020, **10**, 202.

83 M. Chierici, N. Bussola, A. Marcolini, M. Francescatto, A. Zandonà, L. Trastulla, C. Agostinelli, G. Jurman and C. Furlanello, *Front. Oncol.*, 2020, **10**, DOI: 10.3389/fonc.2020.01065.

84 M. Bersanelli, E. Mosca, D. Remondini, E. Giampieri, C. Sala, G. Castellani and L. Milanesi, *BMC Bioinf.*, 2016, **17**, S15.

85 F. Liu, C. Ren, H. Li, P. Zhou, X. Bo and W. Shu, *Bioinformatics*, 2016, **32**, 641–649.

86 J. Lanchantin, R. Singh, Z. Lin and Y. Qi, 2016, arXiv preprint arXiv:1605.01133.

87 Y. Li, W. Shi and W. W. Wasserman, *BMC Bioinf.*, 2018, **19**, 1–14.

88 J. Wang, C. Ling and J. Gao, *A high-precision shallow Convolutional Neural Network based strategy for the detection of Genomic Deletions*, 2016.

89 S. Zhang, J. Zhou, H. Hu, H. Gong, L. Chen, C. Cheng and J. Zeng, *Nucleic Acids Res.*, 2016, **44**, e32–e32.

90 R. Fakoor, F. Ladhak, A. Nazi and M. Huber, *Using deep learning to enhance cancer diagnosis and classification*, 2013.

91 S. Wang, S. Sun, Z. Li, R. Zhang and J. Xu, *PLoS Comput. Biol.*, 2017, **13**, e1005324.

92 T. Mikolov, K. Chen, G. Corrado and J. Dean, 2013, arXiv preprint arXiv:1301.3781.

93 I. Sutskever, O. Vinyals and Q. V. Le, *Sequence to sequence learning with neural networks*, 2014.

94 D. Britz, A. Goldie, M.-T. Luong and Q. Le, 2017, arXiv preprint arXiv:1703.03906.

95 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, *Attention is all you need*, 2017.

96 S. Dev, S. Zhang, J. Voyles and A. S. Rao, *Automated classification of adverse events in pharmacovigilance*, 2017.

97 S. Zhang, S. Dev, J. Voyles and A. S. Rao, *Attention-based multi-task learning in pharmacovigilance*, 2018.

98 A. C. Tricco, W. Zarin, E. Lillie, S. Jeblee, R. Warren, P. A. Khan, R. Robson, G. Hirst and S. E. Straus, *BMC Med. Inf. Decis. Making*, 2018, **18**, 38.

99 A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya and G. Gonzalez, *J. Biomed. Inf.*, 2015, **54**, 202–212.

100 S. Comfort, S. Perera, Z. Hudson, D. Dorrell, S. Meireis, M. Nagarajan, C. Ramakrishnan and J. Fine, *Drug Saf.*, 2018, **41**, 579–590.

101 A. Cocos, A. G. Fiks and A. J. Masino, *J. Am. Med. Inf. Assoc.*, 2017, **24**, 813–821.

102 R. A. Calix, R. Gupta, M. Gupta and K. Jiang, *Deep gramulator: Improving precision in the classification of personal health-experience tweets with deep learning*, 2017.

103 J. Liu and G. Wang, *Int. J. Med. Inf.*, 2018, **117**, 33–43.

104 A. Nikfarjam, A. Sarker, K. O'connor, R. Ginn and G. Gonzalez, *J. Am. Med. Inf. Assoc.*, 2015, **22**, 671–681.

105 D. K. Arnett and S. A. Claas, *Circ. Res.*, 2018, **122**, 1409–1419.

106 R. Simon, M. D. Radmacher, K. Dobbin and L. M. McShane, *J. Natl. Cancer Inst.*, 2003, **95**, 14–18.

107 A. Biton, I. Bernard-Pierrot, Y. Lou, C. Krucker, E. Chapeaublanc, C. Rubio-Pérez, N. López-Bigas, A. Kamoun, Y. Neuzillet and P. Gestraud, *Cell Rep.*, 2014, **9**, 1235–1245.

108 B. Norgeot, B. S. Glicksberg and A. J. Butte, *Nat. Med.*, 2019, **25**, 14–15.

109 A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun and J. Dean, *Nat. Med.*, 2019, **25**, 24–29.

