On the additive artificial intelligence-based discovery of nanoparticle neurodegenerative disease drug delivery systems

Shan He[1,2], Julen Segura Abarrategi[1], Harbil Bediaga[2,3], Sonia Arrasate[1] and Humberto González-Díaz[*1,4,5]

Full Research Paper

Address:
[1]Department of Organic and Inorganic Chemistry, University of Basque Country UPV/EHU, 48940 Leioa, Spain, [2]IKERDATA S.L., ZITEK, UPV/EHU, Rectorate Building, nº6, 48940 Leioa, Greater Bilbao, Basque Country, Spain, [3]Painting Department, Fine Arts Faculty, University of the Basque Country UPV/EHU, 48940, Leioa, Biscay, Basque Country, Spain, [4]Instituto Biofisika (UPV/EHU-CSIC), 48940 Leioa, Spain and [5]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Biscay, Spain

Email:
Humberto González-Díaz[*] - humberto.gonzalezdiaz@ehu.eus

* Corresponding author

Abstract

Neurodegenerative diseases are characterized by slowly progressing neuronal cell death. Conventional drug treatment strategies often fail because of poor solubility, low bioavailability, and the inability of the drugs to effectively cross the blood–brain barrier. Therefore, the development of new neurodegenerative disease drugs (NDDs) requires immediate attention. Nanoparticle (NP) systems are of increasing interest for transporting NDDs to the central nervous system. However, discovering effective nanoparticle neuronal disease drug delivery systems (N2D3Ss) is challenging because of the vast number of combinations of NP and NDD compounds, as well as the various assays involved. Artificial intelligence/machine learning (AI/ML) algorithms have the potential to accelerate this process by predicting the most promising NDD and NP candidates for assaying. Nevertheless, the relatively limited amount of reported data on N2D3S activity compared to assayed NDDs makes AI/ML analysis challenging. In this work, the IFPTML technique, which combines information fusion (IF), perturbation theory (PT), and machine learning (ML), was employed to address this challenge. Initially, we conducted the fusion into a unified dataset comprising 4403 NDD assays from ChEMBL and 260 NP cytotoxicity assays from journal articles. Through a resampling process, three new working datasets were generated, each containing 500,000 cases. We utilized linear discriminant analysis (LDA) along with artificial neural network (ANN) algorithms, such as multilayer perceptron (MLP) and deep learning networks (DLN), to construct linear and non-linear IFPTML models. The IFPTML-LDA models exhibited sensitivity (Sn) and specificity (Sp) values in the range of 70% to 73% (>375,000 training cases) and 70% to 80% (>125,000 validation cases), respectively. In contrast, the IFPTML-MLP and IFPTML-

DLN achieved Sn and Sp values in the range of 85% to 86% for both training and validation series. Additionally, IFPTML-ANN models showed an area under the receiver operating curve (AUROC) of approximately 0.93 to 0.95. These results indicate that the IFPTML models could serve as valuable tools in the design of drug delivery systems for neurosciences.

Introduction

Over time, there has been a significant shift in global dietary habits and lifestyle standards. Poor dietary choices, irregular eating patterns, extended working hours, and sedentary behaviors have contributed to a trend towards an unhealthy lifestyle [1]. This shift has resulted in a rise in chronic degenerative diseases among the elderly population. These diseases encompass a diverse range of conditions characterized by the gradual deterioration of bodily structures and functions [2,3]. Although the exact causes leading to these diseases remain unidentified, there is evidence that oxidative damage plays a crucial role in the progressive neuronal cell death, particularly through the generation of reactive oxygen and nitrogen species [4,5]. In this regard, Alzheimer's and Parkinson's diseases are the most severe and untreatable conditions. Conventional drug treatment methods, such as acetylcholinesterase inhibitor drugs, often encounter obstacles due to their inadequate solubility, limited bioavailability, and inability to effectively penetrate the blood–brain barrier (BBB) [6]. Therefore, there is an urgent need to focus on the advancement of novel neurodegenerative disease drugs (NDDs) [7,8]. The major obstacle encountered by NDDs is the selectivity of the BBB, which limits the number of therapeutic substances able to reach the brain in order to induce a positive effect. Recently, many efforts have been made to develop systems that facilitate the passage of NDDs through the BBB.

Interestingly, nanoparticle (NP) systems are gaining increasing interest among the possible nanomedicine strategies for NDD transport to the central nervous system (CNS) [9,10]. For simplicity, we are going to call them nanoparticle neuronal diseases drug delivery systems (N2D3Ss). N2D3Ss have the ability to protect NDDs from chemical and enzymatic degradation, direct the active compound towards the target site with a substantial reduction of toxicity for the adjacent tissues, and help the NDDs to pass physiological barriers, increasing bioavailability without resorting to high dosages [5,11]. Therefore, researchers are studying and developing new treatment approaches that use N2D3Ss for diagnosis and treatment [12-15].

Also, over the last few years, artificial intelligence/machine learning (AI/ML) models have been applied successfully to solve problems in different disciplines, especially in the interface of chemistry and ND research [16-19]. In this regard, we consider AI/ML to be helpful in the development of N2D3Ss to select the most efficient combination of NP and drug, taking into account properties regarding chemical absorption, distribution, metabolism, excretion, and toxicity (ADMET), and the biological activity regarding NDs [20]. Nevertheless, there is relatively limited experimental data on NPs reported in the scientific literature in comparison to drugs, which increases the difficulty of designing systems based on AI/ML techniques.

An additional essential downside of developing N2D3Ss with AI/ML techniques is the great complexity of the data to be explored. As a result, N2D3S development by the additive approach requires an AI/ML technique to achieve multioutput and multilabel classification [21-24]. In addition, the AI/ML technique includes a pre-processing step to perform information fusion (IF) of the preclinical NDD assay and NP cytotoxicity datasets. Nevertheless, most of the AI/ML methods reported to date only consider the structural/molecular descriptors of the NDDs or NPs as input. Therefore, these methods exclude completely non-structural parameters, specifically experimental conditions of the assays, in order to list NDD or NP labels. Consequently, the resulting model cannot predict multioutput properties and/or labels such as different organisms or cell lines [25-37]. Sizochenko et al. reported a new methodology for NP safety estimation in different organisms [38]. Predicting NP safety instead of biological activity has been the objective of other studies as well [37,39].

As a new strategy to tackle this problem, González-Díaz et al. have developed IFPTML, a multioutput, and input-coded multilabel ML method, which stands for information fusion (IF) + perturbation theory (PT) + machine learning (ML) algorithm [40]. In recent investigations, the IFPTML model has shown to be a powerful tool in molecular sciences and NDD research for the analysis of big datasets that include both structural and non-structural parameters. Application examples are drug screening, protein targeting, the prediction of coated-NP drug release systems [41-49], multitarget networks of neuroprotective compounds for a theoretical study of new asymmetric 1,2-rasagiline carbamates [50], a TOPS-MODE model of multiplexing neuroprotective effects of drugs, an experimental/theoretical study of new 1,3-rasagiline derivatives potentially useful in neurodegenerative diseases [51], as well as QSAR and complex networks in pharmaceutical design, microbiology, parasitology, toxicology, cancer, and neurosciences [52]. Furthermore, this new model also has been used for very similar systems to this research work such as NP systems, taking into account NP

structure and coating agents, synthesis conditions of NPs and loaded drugs, cancer co-therapy drugs, or assay conditions [53-57]. Here we developed IFPTML models for the proposal of N2D3Ss containing NDD and NP components.

Results and Discussion

In order to build the IFPTML models we carried out the steps shown in Figure 1, which shows the general workflow of all computational procedures in this study. For a better under-
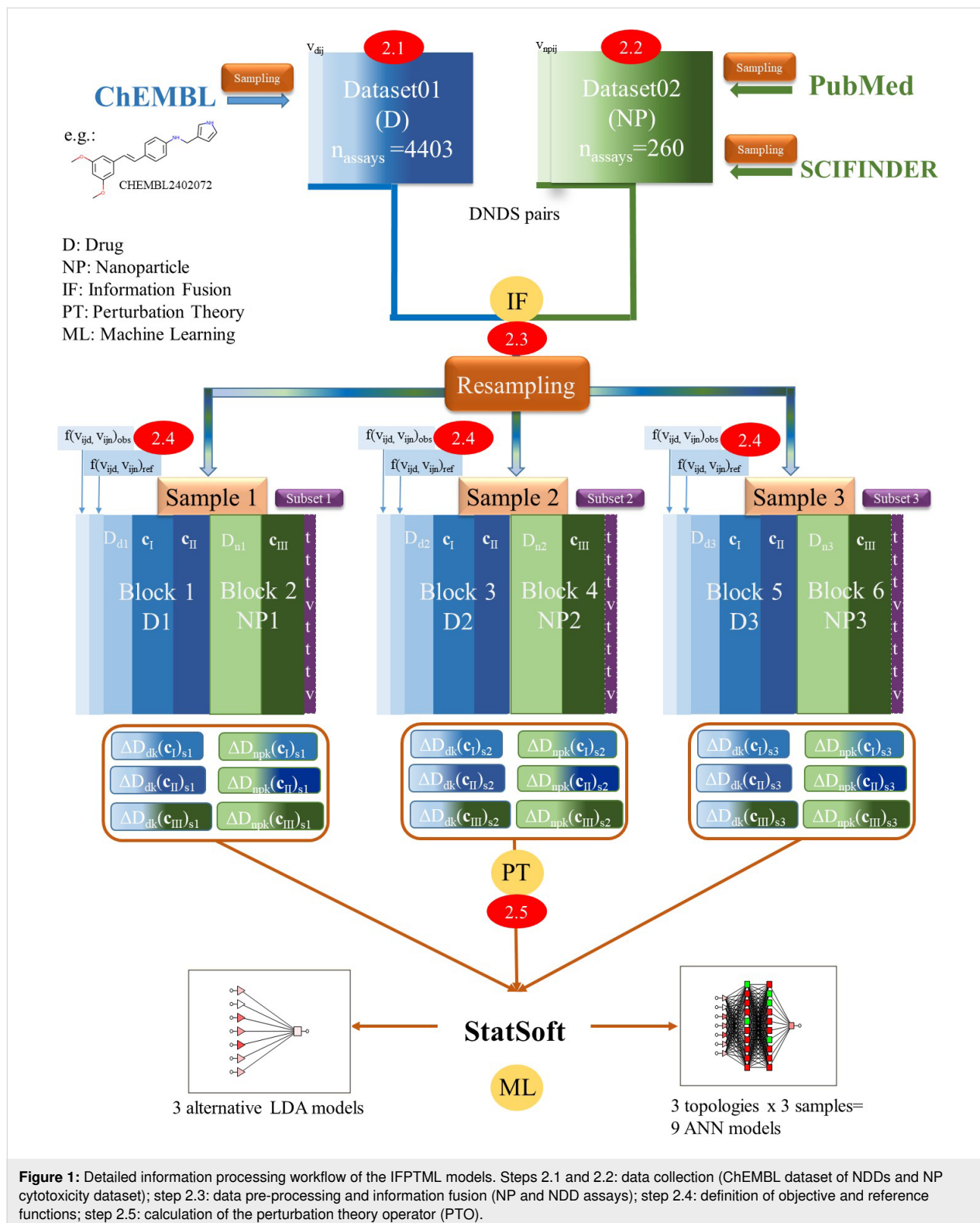

Figure 1: Detailed information processing workflow of the IFPTML models. Steps 2.1 and 2.2: data collection (ChEMBL dataset of NDDs and NP cytotoxicity dataset); step 2.3: data pre-processing and information fusion (NP and NDD assays); step 2.4: definition of objective and reference functions; step 2.5: calculation of the perturbation theory operator (PTO).

standing of all steps, we enumerated them with 2.1, 2.2., and so on.

Figure 2 shows the connections regarding methodology and used databases to our previous publications. For each PTML model development, data download/compilation, data curation, and so on were carried out separately by researchers. First, the database of antineurodegenerative drugs (ADs) was downloaded from ChEMBL by Alonso and coworkers. These researchers employed this database to create advanced predictive models known as multitarget or multiplexing QSAR. These models are designed to forecast both the potential neurotoxicity and neuroprotective effects of drugs across various experimen-

tal setups, including multiple assays, drug targets, and model organisms [41]. Later, Romero Durán et al. enriched the AD database and constructed multitarget networks of neuroprotective compounds to study new asymmetric 1,2-rasagiline carbamates. These authors developed a TOPS-MODE model to analyze the multiple neuroprotective effects of drugs and to conduct experimental/theoretical studies on new 1,3-rasagiline derivatives potentially useful in neurodegenerative diseases [50]. Additionally, Romero Durán et al. expanded the AD database to develop artificial neural network (ANN) algorithms. These models were designed to forecast how ADs interact with targets within the CNS interactome [58]. Speck-Planche et al. compiled manually a database of NPs from the literature. They
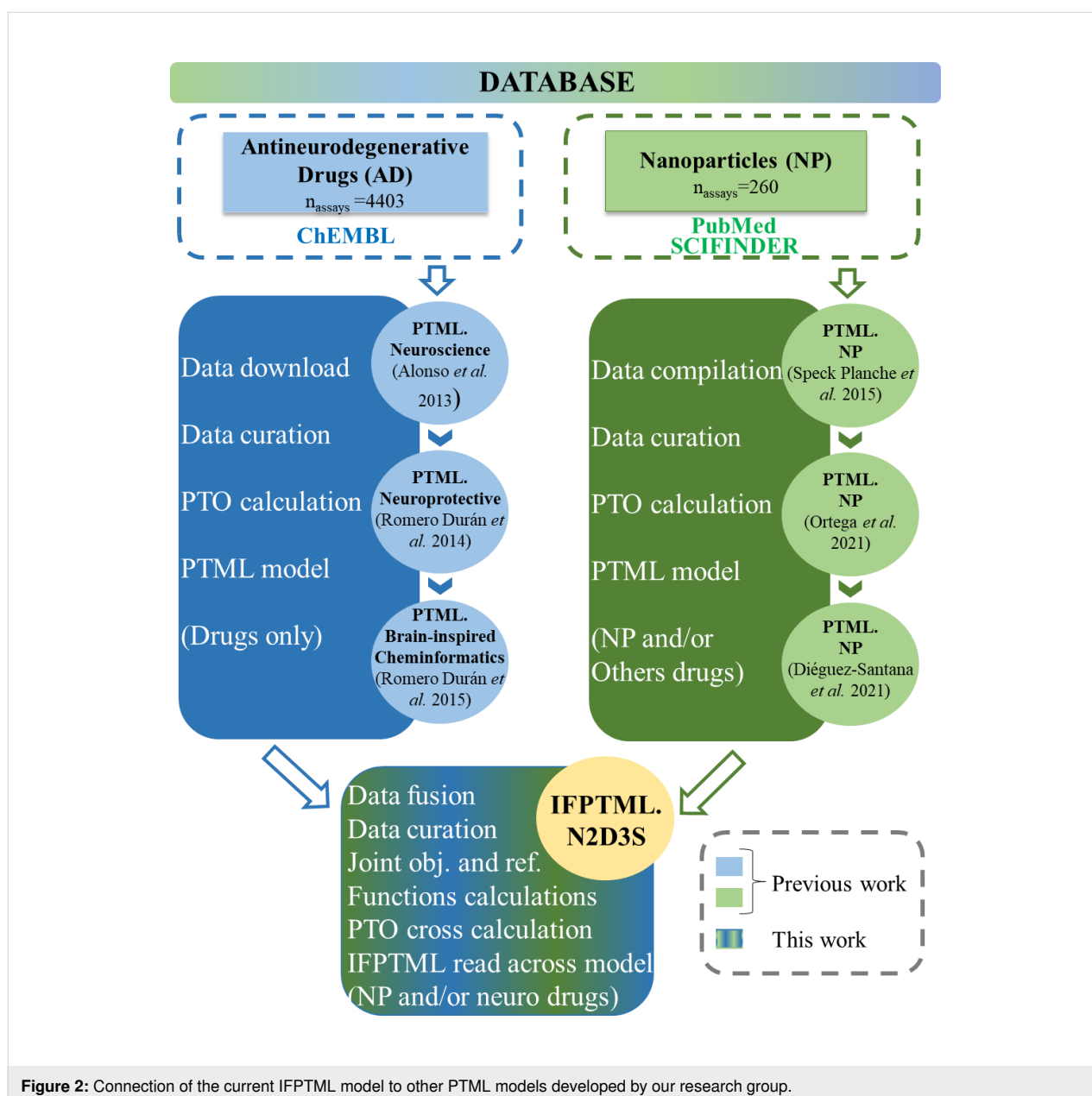

Figure 2: Connection of the current IFPTML model to other PTML models developed by our research group.

constructed a QSAR model to investigate multiple antibacterial profiles of NPs under diverse experimental conditions. Furthermore, Ortega-Tenezaca et al. enriched the NP dataset and developed a PTML model for the discovery of antibacterial NPs [59]. Diéguez et al. expanded the NP database and developed a PTML model in order to design antibacterial drug and NP systems [10].

In this study, we utilized the IFPTML model to investigate N2D3Ss, encompassing assays of ADs and preclinical assays for NPs. To achieve this, we conducted the IF of AD and NP databases, curated the data, combined the objective and reference functions, and calculated the PTO.

NDDs ChEMBL dataset

First, we collected the data of preclinical assays for NDDs from the ChEMBL dataset (see step 2.1. in Figure 1) [60-62]. This dataset contained 4403 preclinical assays for 2566 NDDs (unique drugs), that is, approximately 1.71 assays for each drug. The information downloaded from ChEMBL included discrete variables c_{dj} used to specify the conditions/labels of each assay. These variables are c_{d0}, the biological activity parameter, c_{d1}, the target protein involved in NDs, c_{d2}, the cell line for NDD assays, and c_{d3}, the model organism. Each one of these assays included one out of $n(c_{d0}) = 46$ possible biological activity parameters (e.g., EC_{50} or K_i (nM)). They also involved some of the $n(c_{d1}) = 21$ target proteins, $n(c_{d2}) = 7$ cell lines (SH-SY5Y, CHO-K1, HEK293, PC-12, CHO, HEK-293T, and HuT78), and $n(c_{d3}) = 7$ model organisms (*Homo sapiens, Rattus norvegicus, Mus musculus, Cavia porcellus, Canis lupus familiaris, Macacafas cicularis*, and *Caenorhabditis elegans*). The information downloaded from ChEMBL also included another set of discrete variables used to codify the nature/quality of data. These variables are c_{d4}, the type of target, c_{d5}, the type of assay, c_{d6}, the data curation, c_{d7}, the confidence score, and c_{d8}, the target mapping. Specifically, the target types are $n(c_{d4}) = 6$ (single protein, organism, tissue, non-molecular target, and ADMET), and the assay types are $n(c_{d5}) = 3$ (binding, functional, and ADMET). In addition, data curation has $n(c_{d6}) = 3$ different values (auto-curation, expert, and intermediate), the confidence scores are $n(c_{d7}) = 4$ (9: direct single protein target assigned, 1: target assigned is non-molecular, 0: default value, that is, target assignment has yet to be curated, and 8: homologous single protein target assigned) and the target mapping is $n(c_{d8}) = 3$ (protein, non-molecular target, and homologous protein). Furthermore, this database included the molecular descriptor $\mathbf{D}_{dk} = [D_{d1}, D_{d2}, D_{d3}]$ in order to define the chemical structure of the NDD compound. Specifically, we used two types of molecular descriptor for the *i*-th compound, namely D_{d1} = logarithm of the *n*-octanol/water partition coefficient (LOGP$_i$) and D_{d2} = topological polar surface area (PSA$_i$). The

detailed information of this dataset is given in Supporting Information File 1 (datasheet "ChEMBL").

NP cytotoxicity dataset

Simultaneously, we downloaded the data of preclinical assays for the cytotoxicity of NPs from different sources (see step 2.2. in Figure 1). We selected 62 papers from the scientific literature databases Pubmed and SciFinder [63-65]. This dataset included 260 preclinical assays for 31 unique NPs. Therefore, the number of assays for each NP is about 8.39. Moreover, the data covered a huge range of properties of NPs such as morphology, physicochemical properties, coating agents, length, and time of assay. These properties were defined as discrete variables c_{nj} applied to identify the conditions/labels of each assay. Then, we enumerated all particular conditions of each assay as a general vector $\mathbf{c}_{nj} = [c_{n1}, c_{n2}, c_{n3},…, c_{nmax}]$. These variables are c_{n0}, the biological activity parameter, c_{n1}, the cell line, c_{n2}, the NP shape, c_{n3}, the measurement conditions, and c_{n4}, the coating agent. Each of these assays involved at last one out of $n(c_{n0}) = 5$ possible biological activity parameters (CC_{50}, EC_{50}, IC_{50}, LC_{50}, and TC_{50}). They also include $n(c_{n1}) = 53$ cell lines (e.g., A549 (H), RAW 264.7, and Neuro-2A (M)) and $n(c_{n2}) = 10$ NP shapes (spherical, irregular, slice-shaped, needles, rods, elliptical, pseudo-spherical, polyhedral, pyramidal, and strips). In addition, they contain $n(c_{n3}) = 8$ NP measurement conditions (dry, H_2O, DMEM, RPMI, 1% Trion X-100/H_2O, H_2O/TMAOH, egg/H_2O, and H_2O/HMT) and $n(c_{n4}) = 16$ coating agents (UC, PEG-Si(OMe)$_3$, PVA, sodium citrate, 11-mercaptoundecanoic acid, PVP, propylamonium fragment, undecylazide fragment, CTAB, *N,N,N*-trimethyl-3(1-propene) ammonium fragment, potato starch, *N*-acetylcysteine, CMC-90, 2,3-dimercaptopropanesulfonate, 3-mercaptopropanesulfonate, and thioglycolic acid). The full information of this dataset is shown in Supporting Information File 1 (datasheet "NP").

DNDS pair resampling
IF processing of biological parameters

First, we described and acquired the objective value in order to design the IFPTML model for N2D3S. We defined the target function by applying the vectors of descriptors for all cases $\mathbf{D}_k$ to use as the input variable in the ML model. The target function is commonly achieved by a mathematical conversion of the original theoretical or observed feature of the scheme under analysis [66-68]. In this IFPTML model, it includes two groups of observed values, specifically $v_{ij}(c_{d0})$ and $v_{nj}(c_{n0})$. In addition, it contains two types of input vectors, $\mathbf{D}_{dk}$ and $\mathbf{D}_{nk}$, for the preclinical NDD and NP assays, respectively. Moreover, in this dataset was a large number of different biological parameters c_{d0} and c_{n0}. For example, there are properties such as half the maximum inhibitory concentration (IC_{50} (nM)), half the

maximum effective concentration (EC_{50} (nM)), or the lethal concentration of a substance for an organism (LC_{50} (nM)). Another difficulty is that the majority of $v_{ij}(c_{d0})$ and $v_{nj}(c_{n0})$ values collected are numbers with decimals. Furthermore, in order to acquire the optimum N2D3S, we prioritize some properties and deprioritize others. In this context, we introduced a "desirability" parameter to tackle this problem

The desirability value was established as $d(c_{d0}) = 1$ or $d(c_{n0}) = 1$ when the value of $v_{ij}(c_{d0})$ or $v_{nj}(c_{n0})$ needs to be maximized, otherwise $d(c_{d0}) = -1$ or $d(c_{n0}) = -1$. The different NDD and NP properties/characteristics possess a large number of designations or labels c_{d0} and c_{n0}, respectively, and increase the unreability of the data, making it more laborious to build a regression model. For example, in context of a specific case, biological activity parameters c_{d0} with $d(c_{d0}) = 1$ are Bmax (fmol/mg), the total number of receptors expressed in the same units, activity (%), and Cp (nM). Whereas parameters with $d(c_{d0}) = -1$ are, for example, EC_{50} (nM), IC_{50} (nM), and Imax (%). To address this problem, we used a cutoff value to divide AD and NP assays into favorable and non-favorable assays. It is worth mentioning that using a cutoff is a common practice in drug discovery processes. As a result, acquiring the final target function, the pre-processing of all observed $v_{ij}(c_{d0})$ and $v_{nj}(c_{n0})$ values is crucial in order to remove or reduce imprecisions. Eventually, IF processing of the parameters $v_{ij}(c_{d0})$ and $v_{nj}(c_{n0})$ enabled us to obtain a target function of the N2D3Ss.

We also used a cutoff to rescale the parameters of $v_{ij}(c_{d0})$ and $v_{nj}(c_{n0})$ to obtain the Boolean (dummy) functions $f(v_{ij}(c_{d0}))_{obs}$ and $f(v_{nj}(c_{n0}))_{obs}$. These values were obtained as $f(v_{ij}(c_{d0}))_{obs} = 1$ if $v_{ij}(c_{d0}) >$ cutoff and $d(c_{d0}) = 1$, or $v_{ij}(c_{d0}) <$ cutoff and desirability $d(c_{d0}) = -1$; otherwise $f(v_{ij}(c_{d0})) = 0$. Similarly, $f(v_{nj}(c_{n0}))_{obs} = 1$ if $v_{nj}(c_{n0}) >$ cutoff and $d(c_{n0}) = 1$, or $v_{nj}(c_{n0}) <$ cutoff and $d(c_{n0}) = -1$; else $f(v_{ij}(c_{d0}), v_{nj}(c_{n0})) = 0$. The values $f(v_{ij}(c_{d0}))_{obs} = 1$ and $f(v_{nj}(c_{n0}))_{obs} = 1$ mean to have a positive desired effect of both NDDs and NPs. As a result, the target function was described as $f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{obs} = f(v_{ij}(c_{d0}))_{obs} \cdot f(v_{nj}(c_{n0}))_{obs}$. Therefore, the outcome of the IF scaling $f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{obs}$ is determined by the i-th NDD compound and the n-th NP measurement conditions. The remaining cases, $f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{obs} = 0$, indicate that at least one of the abovementioned conditions fail.

Definition of objective and reference functions
IF phase for combining the references

After we obtained the target function, the next step is to describe the input variables of the IFPTML model. Input variable for this model is the reference function $f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{ref}$. The function $f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{ref}$ plays an impor-

tant role because this function characterizes the expected probability $f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{ref} = p(f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{ref} = 1)$ for achieving the required level of activity for a specific property acquired from well-known systems. IFPTML uses values from well-known systems or subset systems as reference. Afterwards, this model includes the effect of different deviations (perturbations) of the query function from the reference function. Accordingly, $f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{ref}$ can be considered a function related to observed (not predicted) outcomes. In the above section, we mentioned the step of IF scaling to transform the original $v_{ij}(c_{d0})$ and $v_{nj}(c_{n0})$ values into $f(v_{ij}(c_{d0}))_{obs}$ and $f(v_{nj}(c_{n0}))_{obs}$ functions. When we acquire $f(v_{ij}(c_{d0}))_{obs}$ and $f(v_{nj}(c_{n0}))_{obs}$ for all cases in our dataset, the next step is to quantify each of the positive outcomes $n(f(v_{ij}(c_{d0}))_{obs} = 1)$ and $n(f(v_{nj}(c_{n0}))_{obs} = 1)$. Subsequently, in order to obtain the reference or expected functions (Figure 3), we divide the previous values by the entire number of cases for the NDD and NP systems separately. We describe these functions as $f(v_{ij}(c_{d0}))_{ref} = p(f(v_{ij}(c_{d0}))_{obs} = 1) = n(f(v_{ij}(c_{d0}))_{obs} = 1)/n(c_{d0})_j$ and $f(v_{nj}(c_{n0}))_{ref} = p(f(v_{nj}(c_{n0}))_{obs} = 1) = n(f(v_{nj}(c_{n0}))_{obs} = 1)/n(c_{n0})_j$. In this context, we can calculate the reference function directly to recognize the probability products for both subsystems $f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{ref} = p(f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{obs} = 1) = p(f(v_{ij}(c_{d0}))_{obs} = 1) \cdot p(f(v_{nj}(c_{n0}))_{obs} = 1)$. It is worth mentioning that the usage of the reference function at this point is another representation of the IF (combination) of NDD and NP datasets.

PTO calculation
IFPTML N2D3S data analysis

As we mentioned in the previous section, we acquired the results of many cytotoxicity preclinical assays of different NPs [69,70]. Complementarily, we obtained the data of preclinical assays for NDDs from the ChEMBL database [60,71,72]. It included the calculation of the vectors $\mathbf{D}_{nk}$ and $\mathbf{D}_{dk}$ of structural descriptors for all NPs and NDDs. In addition, we constructed the vectors $\mathbf{c}_{nj}$ and $\mathbf{c}_{dj}$ in order to list each label and assay condition for all preclinical assays of NPs and NDDs. Subsequently, we obtained the values $\Delta D_{dk}(\mathbf{c}_{dj})$ and $\Delta D_{nk}(\mathbf{c}_{nj})$ of the respective moving average deviation PTOs.

The NDD vector lists each element $\boldsymbol{D}_{dk} = [D_{d1}, D_{d2}]$. Precisely, these elements are the NDD structural descriptors, which have enabled the development of various strategies to characterize and classify the structure of potential bioactive molecules [73]. These structural descriptors are $D_{d1} =$ logarithm of the n-octanol/water partition coefficient ($LOGP_i$) and $D_{d2} =$ topological polar surface area (PSA_i). In contrast, the cytotoxicity NP vector lists the elements as $\mathbf{D}_{nk} = [D_{n1}, D_{n2}, D_{n3}, D_{n4}, D_{n5}, D_{n6}, D_{n7}, D_{n8}, D_{n9}, D_{n10}, D_{n11}, D_{n12}, D_{n13}, D_{n14}, D_{n15}, D_{n16}, D_{n17}, D_{n18}, D_{n19}, D_{n20}]$. Specifically, they are $D_{n1} =$ NMUn
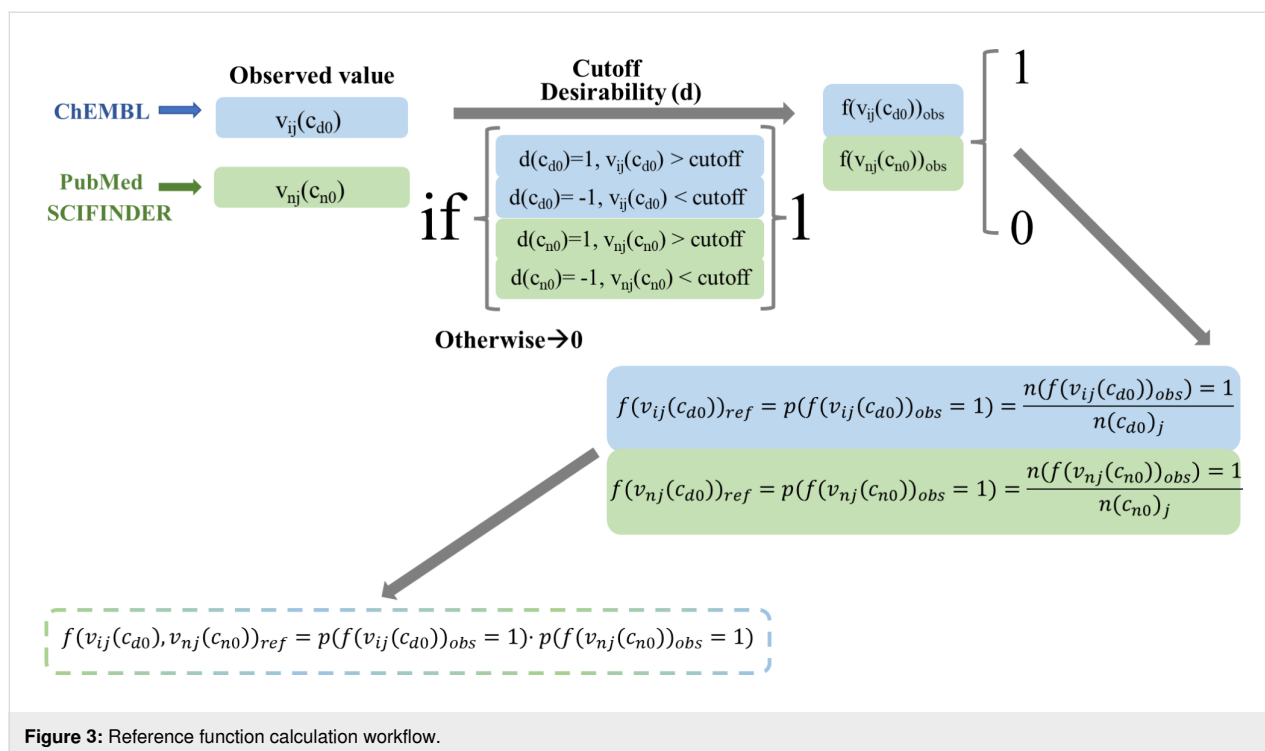
Figure 3: Reference function calculation workflow.

(number of monomer units), D_{n2} = Lnp (NP length), D_{n3} = Vnu (NP volume), D_{n4} = Enu (NP electronegativity), D_{n5} = Pnu (NP polarizability), D_{n6} = Uccoat (unsaturation count), D_{n7} = Uicoat (unsaturation index), D_{n8} = Hycoat (hydrophilic factor), D_{n9} = AMR coat (Ghose–Crippen molar refractivity), D_{n10} = TPSA(NO)coat (topological polar surface area using N,O polar contributions), D_{n11} = TPSA(Tot)coat (topological polar surface area using N,O,S,P polar contributions), D_{n12} = ALOGP-coat (Ghose–Crippen octanol/water partition coefficient), D_{n13} = ALOGP2coat (squared Ghose–Crippen octanol/water partition coefficient (logP^2)), D_{n14} = SAtotcoat (total surface area from P_VSA-like descriptors), D_{n15} = SAacccoat (surface area of acceptor atoms from P_VSA-like descriptors), D_{n16} = SAdoncoat (surface area of donor atoms from P_VSA-like descriptors), D_{n17} = Vxcoat (McGowan volume), D_{n18} = VvdwMGcoat (van der Waals volume from McGowan volume), D_{n19} = VvdwZAZcoat (van der Waals volume from the Zhao–Abraham–Zissimos equation), and D_{n20} = PDIcoat (packing density index).

PT data preprocessing

Apart from the vectors $\mathbf{D}_{dk}$ and $\mathbf{D}_{nk}$, the IFPTML study takes into account all vectors $\mathbf{c}_{dj}$ and $\mathbf{c}_{nj}$ as parts of the non-numerical experimental conditions and labels for both NDD and NP preclinical assays. We calculated the PTOs of the NDD and NP preclinical assays including this additional information. We used Equation 1 and Equation 2 in order to obtain the moving average (MA) PTOs of NDDs and NPs. The PT model begins

with the expected value of a well-known activity and adds the effect of different perturbations/variations to the system. Consequently, the model includes two different input variables, namely the reference or expected-value function $f(v_{ij})_{ref}$ and the PT operators $\Delta D_k(c_j)$. Specifically, they are applied for accounting structural and assay information on NDDs and NPs. In addition, the PTOs $\Delta D(D_{dk})$ and $\Delta D(D_{nk})$ label structural and/or physicochemical characteristics of NDDs and NPs on the variables $\Delta D(D_{dk})$ and $\Delta D(D_{nk})$, respectively. Furthermore, the PTOs $\Delta D(D_{dk})$ and $\Delta D(D_{nk})$ classify biological assay data of NDDs and NPs with the variables $\langle D(D_{dk})_{\mathbf{c}dj}\rangle$ and $\langle D(D_{nk})_{\mathbf{c}nj}\rangle$, respectively. $\langle D(D_{dk})\rangle$ and $\langle D(D_{nk})\rangle$ are the representations of the average operator for counting all cases with the equivalent subset of methodology conditions $\mathbf{c}_{dj}$ and $\mathbf{c}_{nj}$, respectively. Accordingly, they ought to provide exact values for a particular assay with minimum one altered element in methodology conditions of the vectors $\mathbf{c}_{dj}$ or $\mathbf{c}_{nj}$. In this regard, they can specify which assay we are referring to [53-57]. Another kind of PTOs involved in this model is the NDD–NP coating agent moving average balance (MAB) PTO $\Delta\Delta D(D_{ca1}, D_{ca2}, D_{dk})$ (Equation 3). The MAB PTO takes into consideration the likenesses between the information on NDDs and the NP coating agent. Furthermore, PTOs centered straightly on MA and/or linear and non-linear conversions of MA have been applied for NDD and NP development in previous research work [49,55,56]. The MAS is another way of expressing the combination of IF and PT cumulative procedures of NDD and NP datasets.

$$\Delta D\left(D_{\mathrm{d}k}\right) = D\left(D_{\mathrm{d}k}\right) - \left\langle D\left(D_{\mathrm{d}k}\right)_{\mathbf{c}_{\mathrm{d}j}}\right\rangle \qquad (1)$$

$$\Delta D\left(D_{\mathrm{n}k}\right) = D\left(D_{\mathrm{n}k}\right) - \left\langle D\left(D_{\mathrm{n}k}\right)_{\mathbf{c}_{\mathrm{n}j}}\right\rangle \qquad (2)$$

$$\Delta\Delta D\left(D_{\mathrm{ca1}}, D_{\mathrm{ca2}}, D_{\mathrm{d}k}\right) = \Delta D\left(D_{\mathrm{d}k}\right) - \left[\Delta D\left(D_{\mathrm{ca1}}\right) + \Delta D\left(D_{\mathrm{ca2}}\right)\right] \quad (3)$$

IF phase and proposal of training and validation series subsets

To develop the ML models, each of the sample cases are assigned to either the training (subset t) or validation (subset v) series. The process of assignment ought to be random, illustrative, and stratified [74]. Because of the nature of this combinatory system, our sampling also has to take into account the IF scaling procedure. Initially, we obtained the NDD activity dataset from the open database ChEMBL, which has been compiled from primary published literature. The preclinical NP cytotoxicity assays were acquired from journal articles. Afterwards, we prepared each case as the following labels $c_{\mathrm{d}0}$, $c_{\mathrm{d}1}$, $c_{\mathrm{d}2}$, $c_{\mathrm{d}3}$, $c_{\mathrm{d}4}$, $c_{\mathrm{d}5}$, $c_{\mathrm{d}6}$, $c_{\mathrm{d}7}$, $c_{\mathrm{d}8}$, $c_{\mathrm{n}0}$, $c_{\mathrm{n}1}$, $c_{\mathrm{n}2}$, $c_{\mathrm{n}3}$, and $c_{\mathrm{n}4}$. These cases were organized by ranking the labels alphabetically from A to Z (as we mentioned before, they are non-numeric variables in nature). The preference order of the labels on the procedure of ranking was $c_{\mathrm{d}0} \rightarrow c_{\mathrm{n}0} \rightarrow c_{\mathrm{d}1} \rightarrow c_{\mathrm{n}1} \rightarrow c_{\mathrm{d}2} \rightarrow c_{\mathrm{n}2} \rightarrow c_{\mathrm{d}3} \rightarrow c_{\mathrm{n}3}$. In other words, we organized the cases first by $c_{\mathrm{d}0}$, then by $c_{\mathrm{n}0}$, and so forth. This preference order considers the IF step by interchanging labels from AD and NP datasets. Afterwards, we assigned three quarters of the cases to subset t and the remaining quarter to subset v. This random assignment improves the likelihood that nearly all categories of individual labels are denoted by subsets t and v (stratified or proportional random sampling). In addition, this boosts the possibility that practically all cases for each label are in a distribution of 3/4 in subset t and 1/4 subset v, known as representative sampling. It is worth mentioning that the 75% and 25% proportion between training and validation is the most used one in big data analysis [74].

IFPTML-LDA model

The IFPTML N2D3S model utilizes as input variables the PTOs specified in the previous section to codify information of the putative N2D3Ss with their corresponding subsystems NDD and NPs. Combining objective function $f(v_{\mathrm{i}j}, v_{\mathrm{n}j})_{\mathrm{obs}}$ and reference function $f(v_{\mathrm{i}j}, v_{\mathrm{n}j})_{\mathrm{ref}}$ and adding the IF PTOs $\Delta\Delta D(D_{\mathrm{c}1}, D_{\mathrm{c}2}, D_{\mathrm{d}k})$, we obtained the output function $f(v_{\mathrm{i}j}, v_{\mathrm{n}j})_{\mathrm{calc}}$. This function carries out dataset crosscut classification of NDD and

NP information. The generic equation for the IFPTML linear model is the following (Equation 4):

$$\begin{aligned} f\left(v_{\mathrm{i}j}, v_{\mathrm{n}j}\right)_{\mathrm{calc}} &= a_0 + a_1 \cdot f\left(v_{\mathrm{i}j}, v_{\mathrm{n}j}\right)_{\mathrm{ref}} \\ &+ \sum_{k=1, j=1}^{k=k_{\max}, j=j_{\max}} a_{k,j} \cdot \Delta D\left(D_{k\mathrm{i}}\right)_{c_{\mathrm{d}j}} \\ &+ \sum_{k=1, j=1}^{k=k_{\max}, j=j_{\max}} a_{k,j} \cdot \Delta D\left(D_{k\mathrm{n}}\right)_{c_{\mathrm{n}j}} \\ &+ \sum_{k=1, j=1}^{k=k_{\max}, j=j_{\max}} a_{k,j} \cdot \Delta\Delta D\left(D_{k\mathrm{i}}, D_{k\mathrm{n}}\right)_{c_{\mathrm{d}j}, c_{\mathrm{n}j}} \end{aligned} \qquad (4)$$

Generalities for IFPTML model training and validation series

In many big data systems, the linear discriminant analysis (LDA) model is the most commonly used tool to seek the preliminary model because of the simplicity of this technique. In this regard, within this model we applied a forward stepwise (FSW) [75] process that can select automatically the most essential input variables for N2D3Ss. We obtained all results by using the software STATISTICA 6.0 [74]. Afterwards, we applied the expert-guided selection (EGS) heuristic [76] in order to retrain the LDA method using the most crucial parameters selected by the FSW process along with other missing aspects. All IFPTML models were obtained by calculating different statistical parameters, specifically sensitivity (Sn), specificity (Sp), accuracy (Ac), chi-square (χ^2), and the p-level [77,78].

IFPTML-LDA vs cross linear model

In the Introduction section, we indicated the use of ML approaches as a promising strategy in order to tackle practical problems of nanotechnology, such as reducing the number of experiments [79-84]. In this paper the IFPTML method was used to combine preclinical assays of NDDs and NPs. Speck-Planche et al. described multiple IFPTML approaches regarding toxicity and drug delivery of NPs with a large number of species under a wide variety of experimental conditions. However, this study did not take into account the NDDs [54,69,85]. In contrast, Nocedo-Mena et al. reviewed an IFPTML method to explore the activity of NDDs against numerous species and under different assay conditions; but this research they did not consider NPs as part of the system [86]. Accordingly, these models could not take into consideration both components (NDD and NPs) of the N2D3Ss. In our group, Dieguéz-Santana et al. for the first time applied successfully the IFPTML technique to study the combination of multiple antibacterial drugs and preclinical assays on the cytotoxicity of NPs [10]. In this

paper, we used this new approach to develop complex N2D3Ss containing NDDs and NPs, taking into account, among other things, NDD assays, NP types including coating agents, and NP morphologies. To complete the IF scaling process, we calculated the objective function $f(v_{ij}, v_{nj})_{obs} = f(v_{ij})_{obs} \cdot f(v_{nj})_{obs}$. The main purpose of this function is to increase the effect of certainty and maintain the homogeneity of scales. Once the PTOs were obtained, we applied ML methods so as to fit $f(v_{ij}, v_{nj})_{obs}$ and to achieve the IFPTML models. As indicated in the previous section, we classified the preclinical NDD assays, c_{dj}, onto two different partitions (subsets) of variables $\mathbf{c}_I$ and $\mathbf{c}_{II}$. The partition $\mathbf{c}_I$ defines the biological characteristics; it contains, among other things, c_{d0} = biological activity parameters of NDDs (e.g., IC_{50}, K_i, potency, and time) and c_{d1} = type of proteins involved in the NDs. The partition $\mathbf{c}_{II}$ defines the data quality; it contains, among other things, c_{d4} = type of target and c_{d5} = type of assay. For the preclinical NP cytotoxicity assays, $\mathbf{c}_{nj}$ forms only one partition $\mathbf{c}_{III}$, which describes its nature and involves c_{n0} = biological activity parameters of the NPs (e.g., CC_{50}, IC_{50}, LC_{50}, and EC_{50}), c_{n1} = cell lines, c_{n2} = NP morphology, and c_{n3} = NP synthesis conditions. In addition, we acquired two types of IFPTML-LDA model for designing the N2D3Ss. On the one hand, we obtained the IFPTML-LDA by calculating the PTOs $\Delta D_k(\mathbf{c}_j)$ as the difference between the average value $\langle D_k(\mathbf{c}_j)\rangle$ and the partition $\mathbf{c}_n$ within of their own set. As result, the best IFPTML-LDA model found is as follows (Equation 5):

$$
\begin{aligned}
f\left(v_{dij}, v_{nij}\right)_{calc} = &-4.46387 + 16.30655 \cdot f\left(v_{dij}, v_{nij}\right)_{ref} \\
&+0.00003 \cdot \Delta DPSA\left(\mathbf{c}_I\right)_{dj} \\
&+0.0045 \cdot \Delta Dt\left(\mathbf{c}_{III}\right)_{nj} \\
&+0.00062 \cdot \Delta DLnp\left(\mathbf{c}_{III}\right)_{nj} \\
&+0.00675 \cdot \Delta DVnpu\left(\mathbf{c}_{III}\right)_{nj} \\
&+0.00431 \cdot \Delta DVxcoat\left(\mathbf{c}_{III}\right)_{nj} \\
&-0.00537 \cdot \Delta DVvdwMGcoat\left(\mathbf{c}_{III}\right)_{nj}
\end{aligned}
\tag{5}
$$

$$
N_{train} = 375000; \quad \chi^2 = 24273.63; \quad p\text{-level} < 0.05
$$

On the other hand, we tested the possibility to improve the results of statistical parameters for the IFPTML-LDA algorithm. To this end, we calculated the PTOs $\Delta D_k(\mathbf{c}_j)$ by performing all possible combinations among the average values $\langle D_k(\mathbf{c}_j)\rangle$ of both vectors $\mathbf{D}_{nk}$ and $\mathbf{D}_{dk}$ with each partition. As a result, we obtained three different combinations of crossing PTOs for each sample, one for NDDs ($\Delta D_{dk}(\mathbf{c}_{III})$) and two for NPs ($\Delta D_{nk}(\mathbf{c}_I)$ and $\Delta D_{nk}(\mathbf{c}_{II})$). For simplicity, they are named "IFPTML-LDA with cross" (see more details in Figure 1). The

best IFPTML-LDA found with the cross model is the following (Equation 6):

$$
\begin{aligned}
f\left(v_{dij}, v_{nij}\right)_{calc} = &-4.44505 + 14.28457 \cdot f\left(v_{dij}, v_{nij}\right)_{ref} \\
&+0.00216 \cdot \Delta DPSA\left(\mathbf{c}_I\right)_{cn_j} \\
&+0.00241 \cdot \Delta Dt\left(\mathbf{c}_{III}\right)_{cn_j} \\
&+0.01201 \cdot \Delta DLnp\left(\mathbf{c}_{III}\right)_{cn_j} \\
&+0.16549 \cdot \Delta DVnpu\left(\mathbf{c}_{III}\right)_{cn_j} \\
&-0.02389 \cdot \Delta DVxcoat\left(\mathbf{c}_{III}\right)_{cn_j} \\
&+0.04902 \cdot \Delta DVvdwMGcoat\left(\mathbf{c}_{III}\right)_{cn_j} \\
&+2.040821 \cdot \Delta DEnpu\left(\mathbf{c}_{II}\right)_{c_{dn}} \\
&+0.03229 \cdot \Delta DAMRcoat\left(\mathbf{c}_{II}\right)_{c_{dn}}
\end{aligned}
\tag{6}
$$

$$
N_{train} = 375000; \quad \chi^2 = 43587.01; \quad p\text{-level} < 0.05
$$

The output function $f(v_{dij}, v_{nij})_{calc}$ provides a real numeric value that will probably be applied to counting N2D3Ss. This function was acquired by calculating the objective function $f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))_{obs}$ with the ML method making use of the PTOs. The characteristic of the IFPTML models was defined by the statistical parameters sensibility (Sn), specificity (Sp), accuracy (Ac), chi-square test (χ^2), and p-level [74]. The results summary collected in Table 1 contains the statistical parameters for the best models found (Equation 2) for each sample (standard IFPTML-LDA and IFPTML-LDA with cross) are collected in Table 1. The statistical parameters obtained for both methods were in the accuracy range described for the classification model of ML algorithms [77,78]. The standard IFPTML-LDA contains all indispensable variables for defining the NDD structures and the most significant parameters for NPs, such as morphology, size, and assay conditions, among other things. In the IFPTML-LDA with cross system, we included not only all essential variables but also two crossing PTOs. These new PTOs were chosen by the FSW method, which can select the most influential variable in the system under study.

The IFPTML-LDA model in this paper had Sn and Sp values of 70%–73% in both training and validation series. The IFPTML-LDA with cross model showed significantly higher Sn and Sp values of 70%–80% in both series. By only adding two PTOs to the standard model, the IFPTML-LDA Sp value was improved by almost 7% in the training/validation series. However, the Sp and Sn values of the "with cross" model are slightly unbalanced in comparison with the standard model; yet, the Sp and Sn values remain approximately constant within the same training and validation series.

Table 1: IFPTML-LDA N2D3S model results summary.

Data				Stat.	Param.	Without cross Subset predicted		Param.	With cross Subset predicted	
Sample	Set	Subset	Param.		(%)	0	1	(%)	0	1
1	t	0	Sp		73	255190	94292	72.2	252534	97042
		1	Sn		71	7398	18120	74.4	6517	18907
	v	0	Sp		73.3	85369	31125	72.3	84183	32315
		1	Sn		70.3	2522	5984	73.9	2218	6284
2	t	0	Sp		70	244548	105076	79.5	277907	71717
		1	Sn		62.1	9528	15848	70.1	7584	17792
	v	0	Sp		70	81640	35009	79.7	92929	23720
		1	Sn		63.1	3081	5270	70.7	2451	5900
3	t	0	Sp		70.6	246551	102809	79.6	277921	71439
		1	Sn		62.3	11616	15974	70.1	7668	17972
	v	0	Sp		70.7	82370	34174	79.6	92726	23818
		1	Sn		62.7	3828	5300	70.4	2500	5956
Avg.	t	0	Sp		71.2	248763	100726	77.1	269454	80066
		1	Sn		65.1	9514	16647	71.5	7256	18224
	v	0	Sp		71.3	83126	33436	77.2	89946	26618
		1	Sn		65.4	3144	5518	71.7	2390	6047

Linear vs non-linear IFPTML models

In order to obtain the artificial neural network (ANN) model, we used the same PTO variables as in the LDA model. As an alternative to the non-linear models, we created the ANN by using the same software STATISTICA. The ANN can also be used as a new strategy to confirm and validate the linear hypothesis. Both are comparable because the linear neural network (LNN) techniques are analogous to LDA models and they are linear equations. Accordingly, the IFPTML-LNN model is a useful tool to assess the degree of strength of the linear relationship between PTOs and the N2D3S objective function. The IFPTML-LNN models in this work showed lower Sn and Sp values of 64%–65% in the training and validation series, compared with the IFPTML-LDA models, see details in Table 2.
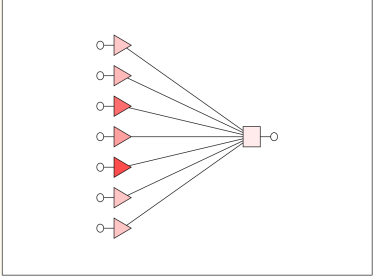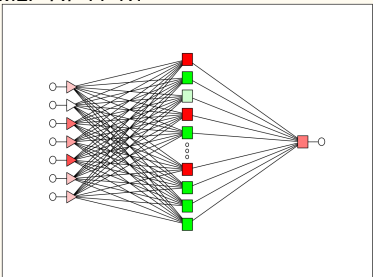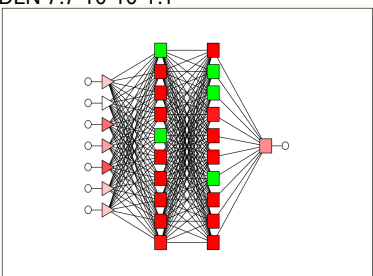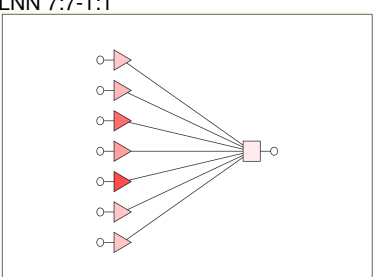
Analogous to the IFPTML-LDA model, the values of the statistical parameters Sp and Sn are considerably balanced and stay steady when comparing training and validation series. Also, we obtained two types of non-linear models, the multilayer perceptron (MLP) and the depth learning network (DLN). The MLP is made up by seven PTOs as input layer, a hidden layer with eleven neurons, and an output layer. The most notable difference is that the DLN involves two hidden layers, each one with ten neurons. Both MLP and DLN showed high Sp and Sn values of 85%–86% in the training and validation series. If we compare the linear IFPTML-ANN model with non-linear models based on the results of statistical parameters, we can confirm that N2D3S is a non-linear system. Another result obtained in the development of the ANN is the area under receiver operating characteristic (AUROC) (Figure 4) [74]. The AUROC curve values are 0.93–0.94 for both MLP and DLN models in the training and validation series. The AUROC values of the non-linear models are remarkably different from the random (RND) curve with AUROC = 0.5 [74].

Robustness analysis of IFPTML models

The design of the N2D3Ss involve the combination of a large amount of data on preclinical assays of NDDs and NPs. Because of the nature of this big data system, we divided the information fusion dataset into three samples. In the previous section, we discussed the best model obtained for IFPTML-LDA, IFPTML-LDA with cross and IFPTML-ANN. In this section, a robustness analysis for the three samples is given (see Table 3). In general, the number of cases (*n*) used in training and validation series for all models presented the lowest standard deviation (SDV), which indicated that most of the data in a sample tend to be clustered near its mean [87]. In contrast, the high value of SDV for the DLN model indicates that the data was distributed over a wide range of values. In addition, all

Table 2: The best result of IFPTML-ANN N2D3Ss models found.

Sample	IFPTML-ANN Models[a]	Subset	Stat.	Val. (%)	$f(v_{ij}(c_{d0}), v_{nj}(c_{n0}))$ Pred.	Observed 1	0	AUROC
01	LDA 7:7-1:1 FSTW + EGS	t	Sp	0	73.0	94272	255178	—
			Sn	1	71.0	18057	7367	
		v	Sp	0	73.3	31125	85319	—
			Sn	1	70.3	5980	2522	
	MLP 7:7-11-1:1 BP96b	t	Sp	0	86.1	300836	48740	0.943
			Sn	1	85.8	3610	2181	
		v	Sp	0	86.1	100278	16220	0.934
			Sn	1	86.2	1173	7329	
	DLN 7:7-10-10-1:1 BP100,CG20b	t	Sp	0	85.8	299942	49634	0.945
			Sn	1	85.8	3621	21803	
		v	Sp	0	85.9	100103	16395	0.933
			Sn	1	86.3	1168	7334	
	LNN 7:7-1:1 PI	t	Sp	0	65.0	227184	122392	0.744
			Sn	1	64.7	8971	16453	
		v	Sp	0	65.1	75788	40710	0.733
			Sn	1	64.1	3055	5447	

models presented similar SDV values in the same training and validation series. Interestingly, the LDA model showed significantly lower values of SDV for Sp (>1), compared with the SDV for Sn (>4) in the training and validation series. However, the SDV values for the LDA cross model were contrary to those of LDA, with lower SDV values for Sn and higher values for
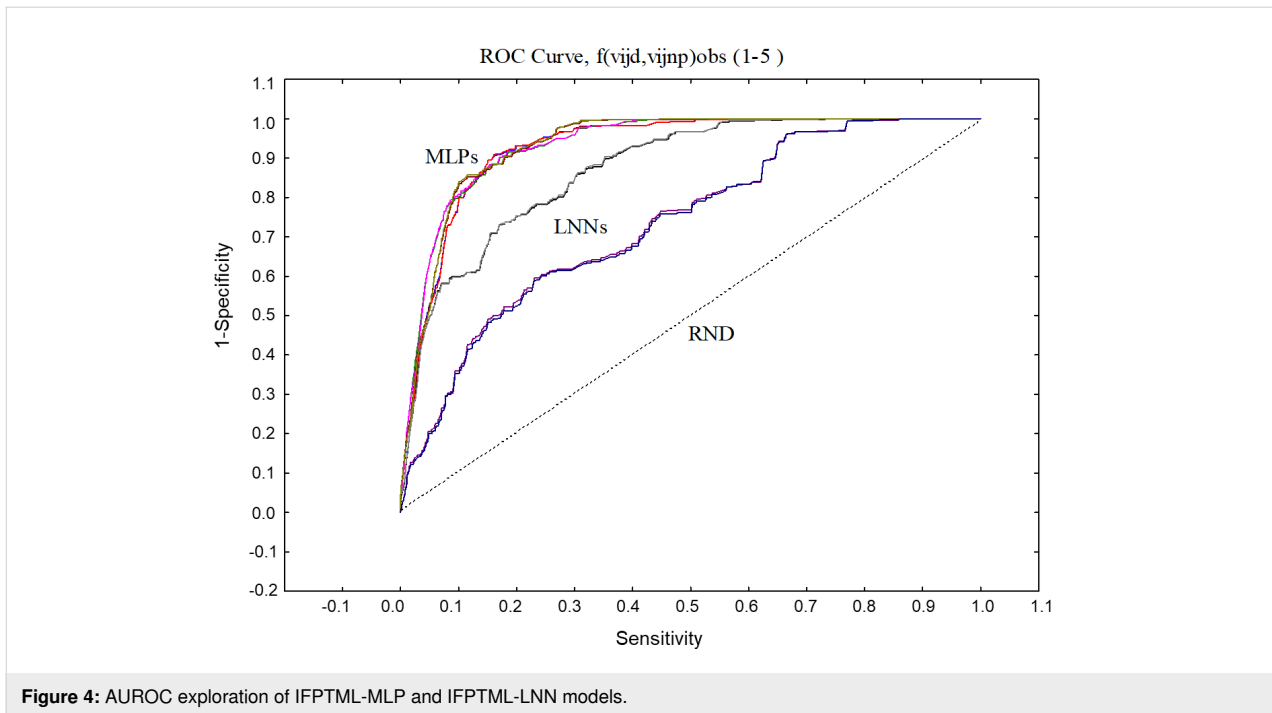
Figure 4: AUROC exploration of IFPTML-MLP and IFPTML-LNN models.

Sp. It is worth mentioning that both MLP 1 and LNN models yielded statistical parameters close to its mean, in other words these models are robust. Furthermore, using the IFPTML-ANN model, we also obtained AUROC values as results. After doing the robustness analysis, we can confirm that all AUROC values for all ANN models are robust. In addition, the AUROC graphic (Figure 4) gives evidence to this because of the similarity of the curve shapes.

The results reveal the strength of the linear hypothesis. Nevertheless, the statistical parameters of the obtained linear model are not satisfactorily at all. As a result, in the IFPTML-LDA with cross model, we enlarged the number of input variables from seven to nine. Thus, we did not obtain substantial change. Therefore, we tested more complex non-linear models so as to improve the Sp and Sn values. The IFPTML-MLP 7:7-11-1:1 model, containing seven input variables in the input layer and eleven neurons in the hidden layer, yielded the best statistical parameters of Sn and Sp values (Table 3). The IFPTML-DLN model, which involves two hidden layers, yielded similar result as IFPTML-MLP 7:7-11-1:1.

Taking into account all the aforementioned results, we can consider both IFPTML-MLP and IFPTML-DLN as the best models with remarkably higher values of Sp and Sn of 85%–86% and AUROC values of 0.93–0.94. However, the DLN model is more complex and yields only a non-significant improvement of statistical parameters in comparison with the MLP model. Thus, we can confirm that N2D3Ss require the

MLP model. This selection is supported by the principle of parsimony, prioritizing the simplest explanations among all possible ones [88]. In Table 4, an input variable sensitivity analysis concerning NDDs, NPs, and the corresponding subsystems are shown for the IFPTML-ANN model. The IFPMTL-LNN model involves almost all significant parameters according to the EGS criteria. The majority of parameters provide a substantial influence on the sensitivity ≥ 1 [74]. In many cases, the value of sensitivity analysis is slightly higher with a sensitivity of 1.00–1.08. Nevertheless, the EGS perspective fails in the selection of ΔDPSA(c_I) and ΔDt(c_{III}) variables. In this regard, the IFPTML-ANN model suggests that those variables do not affect any model. In contrast, the IFPTML-LNN yielded the lowest value of sensitivity of 1.00–1.13, which would underline the need for a complex model in N2D3Ss. The DLN model involves the essential variables in accordance with the EGS criteria; however, they have remarkably higher sensitivity values of 0.96–2.03. The MLP yielded the highest values of sensitivity between 1.13 and 2.57.

IFPTML-LDA for N2D3S simulation

In this section, we employed the IFPTML-LDA technique to calculate the probability values for some selected cases of N2D3Ss. The linear model was chosen for its simplicity and the slight improvement of the non-linear model. The value of probability p(N2D3S$_{in}$)$_{cdj.cnj}$ was obtained for N2D3Ss, created by the combination of the i-th AD$_i$ and the n-th NP$_n$, which are likely to have a desired level of biological activity under both assay conditions c_{dj} and c_{nj}. This simulation experiment

Table 3: Result summary of N2D3Ss alongside average of three samples and standard deviations.

	Model	t			v			AUROC (t/v)
		Sp	Sn	n	Sp	Sn	n	
AVG	LDA	71.2	65.1	375000	71.3	65.4	125000	—
	LDA cross	77.1	71.5	375000	77.2	71.7	125000	—
	MPL 1	85.1	85.0	375000	85.1	85.1	125000	0.937/0.925
	DNL	79.2	79.0	375000	79.2	79.3	125000	0.893/0.879
	LNN	65.0	64.9	375000	65.1	64.9	125000	0.748/0.737
	Model	t			v			AUROC (t/v)
		Sp	Sn	n	Sp	Sn	n	
SDV	LDA	1.587	5.082	0	1.739	4.277	0	—
	LDA cross	4.244	2.483	0	4.244	1.940	0	—
	MLP 1	1.266	1.217	0	1.940	1.102	0	0.010/0.010
	DLN	8.489	8.568	0	8.584	8.727	0	0.069/0.071
	LNN	0.100	0.153	0	0.153	0	0	0.005/0.003

Table 4: IFPTML-ANN model input variable sensitivity analysis for different subsystems with their corresponding variables.

Sub-systems	Variables	LNN		MLP		DLN					
		t	v	t	v	t	v	t	v	t	v
NDD$_S$&NP	$\mathbf{f}(c_{d0},c_{n0})_{ref}$	1.02	1.02	1.32	1.33	1.46	1.45	1.25	1.24	1.38	1.40
NDDs	$\Delta DPSA(\mathbf{c}_I)$	0	0	0	0	0	0	0	0	0	0
	$\Delta Dt(\mathbf{c}_{III})$	0	0	0	0	0	0	0	0	0	0
	$\Delta DL_{np}(\mathbf{c}_{III})$	1.00	1.00	1.14	1.13	1.08	1.08	1.08	1.08	1.60	1.59
NP	$\Delta DV_{npu}(\mathbf{c}_{III})$	1.00	1.00	2.22	2.22	0.92	0.92	1.06	1.05	1.24	1.25
	$\Delta DV_{xcoat}(\mathbf{c}_{III})$	1.00	1.00	1.96	1.98	1.45	1.47	1.45	1.48	1.99	2.03
	$\Delta DV_{vdw}MG_{coat}(\mathbf{c}_{III})$	1.13	1.13	2.57	2.54	1.44	1.43	1.24	1.24	1.91	1.90

involved in total $N_{N2D3S} = 88$ systems vs a total of $N_{NDDs} = 123$ drugs. Many of these drugs are NDDs with known anti-neurodegenerative activity, generally for Alzheimer and Parkinson diseases. Some of these NDDs are approved by the Food and Drug Administration, while others have been shown to be active in several assays. In addition, the simulation also contained cytotoxicity assays against multiple cell lines, the type of NPs, their coating, and the time of each assay. In this context, we calculated a total of $N_{tot} = N_{NDDs}·N_{NP} = 22·218 = 4796$ values of probability, which were able to predict successfully putative N2D3Ss.

Figure 5 depicts the results in a three-color scale according to the value of probability: the green section indicates high probability (0.61–0.98), yellow low-to-middle probability (0.17–0.60), and red very low probability (<0.17). Assays that

have not been reported before, are represented in the original dataset to a very low extent, or whose combination of NDDs and NPs are meaningless were illustrated in white color to avoid an overestimation of results. The results of the IFPTML-LDA model pointed out some N2D3Ss as promising combinations for future additional assays. The resulting N2D3Ss shown in Figure 5 involve twenty different NDDs. The first ten are 1 = clozapine, 2 = galantamine, 3 = levodopa, 4 = apomorphine, 5 = fiduxosin, 6 = beagacestat, 7 = memoquin, 8 = mesodihydroguaired acid, 9 = tarenflubil, and 10 = huperzine A. The other ten NDDs are 11 = guanidinonaltrindole, 12 = semagacestat, 13 = huprine X, 14= carproctamide, 15= tacrine, 16 = tramiprosate, 17 = preladenant, 18 = piracetam, 19 = istradefylline, and 20 = rivastigmine. These systems include the following coating agents: PEG = polyethylene glycol, PVP = polyvinylpyrrolidone, PPF = propylammonium fragment, and
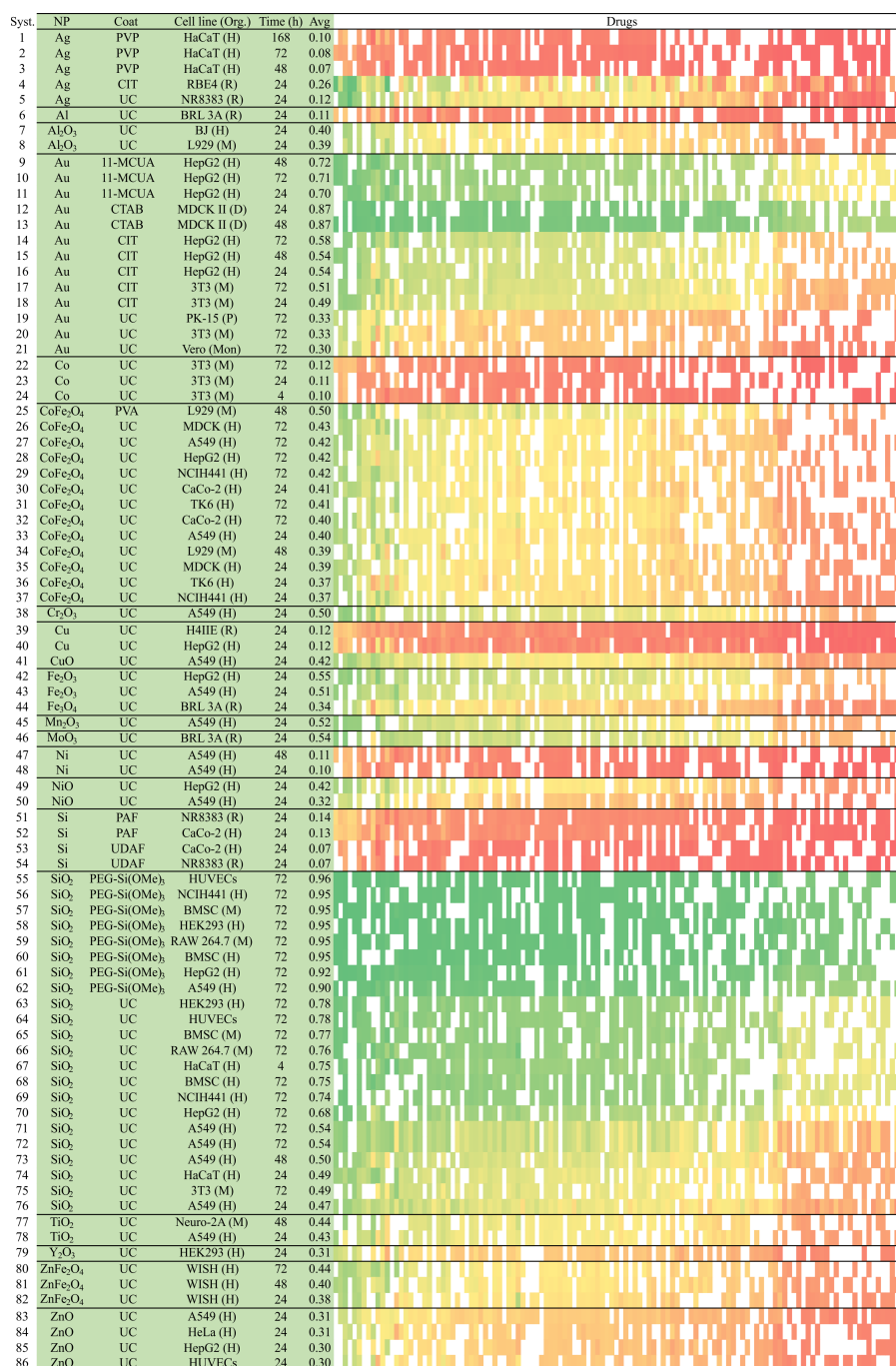
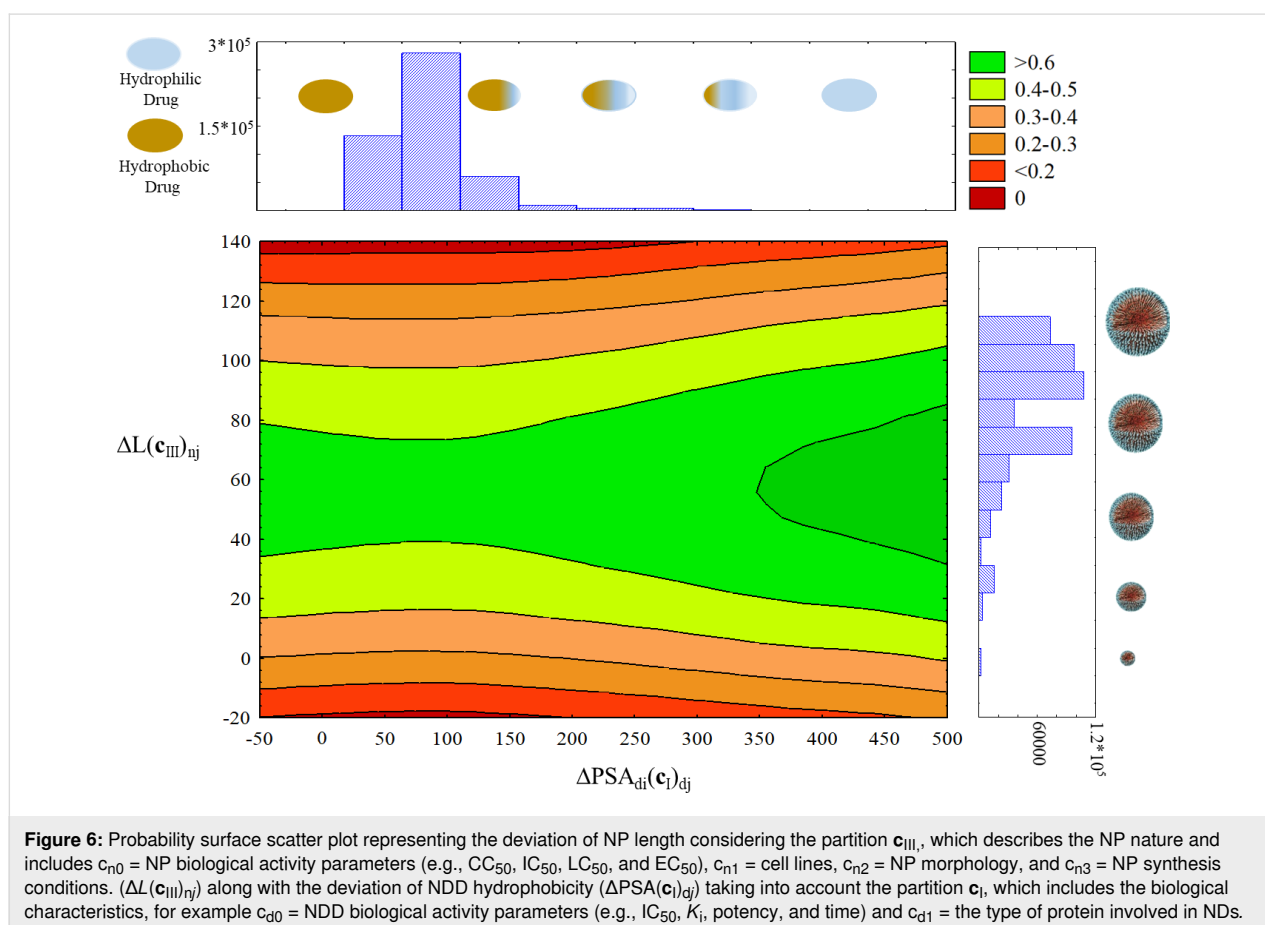| Syst.	NP	Coat	Cell line (Org.)	Time (h)	Avg	Drugs
1	Ag	PVP	HaCaT (H)	168	0.10	
2	Ag	PVP	HaCaT (H)	72	0.08	
3	Ag	PVP	HaCaT (H)	48	0.07	
4	Ag	CIT	RBE4 (R)	24	0.26	
5	Ag	UC	NR8383 (R)	24	0.12	
6	Al	UC	BRL 3A (R)	24	0.11	
7	Al_2O_3	UC	BJ (H)	24	0.40	
8	Al_2O_3	UC	L929 (M)	24	0.39	
9	Au	11-MCUA	HepG2 (H)	48	0.72	
10	Au	11-MCUA	HepG2 (H)	72	0.71	
11	Au	11-MCUA	HepG2 (H)	24	0.70	
12	Au	CTAB	MDCK II (D)	24	0.87	
13	Au	CTAB	MDCK II (D)	48	0.87	
14	Au	CIT	HepG2 (H)	72	0.58	
15	Au	CIT	HepG2 (H)	48	0.54	
16	Au	CIT	HepG2 (H)	24	0.54	
17	Au	CIT	3T3 (M)	72	0.51	
18	Au	CIT	3T3 (M)	24	0.49	
19	Au	UC	PK-15 (P)	72	0.33	
20	Au	UC	3T3 (M)	72	0.33	
21	Au	UC	Vero (Mon)	72	0.30	
22	Co	UC	3T3 (M)	72	0.12	
23	Co	UC	3T3 (M)	24	0.11	
24	Co	UC	3T3 (M)	4	0.10	
25	$CoFe_2O_4$	PVA	L929 (M)	48	0.50	
26	$CoFe_2O_4$	UC	MDCK (H)	72	0.43	
27	$CoFe_2O_4$	UC	A549 (H)	72	0.42	
28	$CoFe_2O_4$	UC	HepG2 (H)	72	0.42	
29	$CoFe_2O_4$	UC	NCIH441 (H)	72	0.42	
30	$CoFe_2O_4$	UC	CaCo-2 (H)	24	0.41	
31	$CoFe_2O_4$	UC	TK6 (H)	72	0.41	
32	$CoFe_2O_4$	UC	CaCo-2 (H)	72	0.40	
33	$CoFe_2O_4$	UC	A549 (H)	24	0.40	
34	$CoFe_2O_4$	UC	L929 (M)	48	0.39	
35	$CoFe_2O_4$	UC	MDCK (H)	24	0.39	
36	$CoFe_2O_4$	UC	TK6 (H)	24	0.37	
37	$CoFe_2O_4$	UC	NCIH441 (H)	24	0.37	
38	Cr_2O_3	UC	A549 (H)	24	0.50	
39	Cu	UC	H4IIE (R)	24	0.12	
40	Cu	UC	HepG2 (H)	24	0.12	
41	CuO	UC	A549 (H)	24	0.42	
42	Fe_2O_3	UC	HepG2 (H)	24	0.55	
43	Fe_2O_3	UC	A549 (H)	24	0.51	
44	Fe_2O_4	UC	BRL 3A (R)	24	0.34	
45	Mn_2O_3	UC	A549 (H)	24	0.52	
46	MoO_3	UC	BRL 3A (R)	24	0.54	
47	Ni	UC	A549 (H)	48	0.11	
48	Ni	UC	A549 (H)	24	0.10	
49	NiO	UC	HepG2 (H)	24	0.42	
50	NiO	UC	A549 (H)	24	0.32	
51	Si	PAF	NR8383 (R)	24	0.14	
52	Si	PAF	CaCo-2 (H)	24	0.13	
53	Si	UDAF	CaCo-2 (H)	24	0.07	
54	Si	UDAF	NR8383 (R)	24	0.07	
55	SiO_2	PEG-Si(OMe)$_3$	HUVECs	72	0.96	
56	SiO_2	PEG-Si(OMe)$_3$	NCIH441 (H)	72	0.95	
57	SiO_2	PEG-Si(OMe)$_3$	BMSC (M)	72	0.95	
58	SiO_2	PEG-Si(OMe)$_3$	HEK293 (H)	72	0.95	
59	SiO_2	PEG-Si(OMe)$_3$	RAW 264.7 (M)	72	0.95	
60	SiO_2	PEG-Si(OMe)$_3$	BMSC (H)	72	0.95	
61	SiO_2	PEG-Si(OMe)$_3$	HepG2 (H)	72	0.92	
62	SiO_2	PEG-Si(OMe)$_3$	A549 (H)	72	0.90	
63	SiO_2	UC	HEK293 (H)	72	0.78	
64	SiO_2	UC	HUVECs	72	0.78	
65	SiO_2	UC	BMSC (M)	72	0.77	
66	SiO_2	UC	RAW 264.7 (M)	72	0.76	
67	SiO_2	UC	HaCaT (H)	4	0.75	
68	SiO_2	UC	BMSC (H)	72	0.75	
69	SiO_2	UC	NCIH441 (H)	72	0.74	
70	SiO_2	UC	HepG2 (H)	72	0.68	
71	SiO_2	UC	A549 (H)	72	0.54	
72	SiO_2	UC	A549 (H)	72	0.54	
73	SiO_2	UC	A549 (H)	48	0.50	
74	SiO_2	UC	HaCaT (H)	24	0.49	
75	SiO_2	UC	3T3 (M)	72	0.49	
76	SiO_2	UC	A549 (H)	24	0.47	
77	TiO_2	UC	Neuro-2A (M)	48	0.44	
78	TiO_2	UC	A549 (H)	24	0.43	
79	Y_2O_3	UC	HEK293 (H)	24	0.31	
80	$ZnFe_2O_4$	UC	WISH (H)	72	0.44	
81	$ZnFe_2O_4$	UC	WISH (H)	48	0.40	
82	$ZnFe_2O_4$	UC	WISH (H)	24	0.38	
83	ZnO	UC	A549 (H)	24	0.31	
84	ZnO	UC	HeLa (H)	24	0.31	
85	ZnO	UC	HepG2 (H)	24	0.30	
86	ZnO	UC	HUVECs	24	0.30	

Figure 5: IFPTML-LDA N2D3Ss experiment simulation.

UAF = undecylazide fragment. The symbol UC = uncoated represents non-coated N2D3Ss. Interestingly, a high value of prediction involves PEG-Si(OMe)$_3$ as NP coating with p(N2D3S$_{in}$)$_{cdj.cnj}$ = 0.80–0.99 for the majority of NDDs. Another important factor that may affect the value of probability is the type of NP. It appears that metal oxide compounds such as SiO_2 and TiO_2 along with PEG-Si(OMe)$_3$NP coating for almost all NDDs are likely to be promising for further assays. Double metal oxide compounds such as $CoFe_2O_4$ and $ZnFe_2O_4$ obtained intermediate probability values p(N2D3S$_{in}$)$_{cdj.cnj}$ = 0.17–0.70 against TK6 (H) and WISH (H). In general, the least advantageous combinations are metal NPs with all NDDs, which give low values of probability (p(N2D3S$_{in}$)$_{cdj.cnj}$ = 0.02–0.35). It is worth mentioning that all predictions carried out by this method should be used with caution and require experimental corroboration. The potential

utility of the IFPTML method is to speed up experimental studies and to provide inexpensive preliminary results for a large database of N2D3Ss. This approach offers an efficient and powerful tool to direct experimental research as an alternative to tedious trial-and-error tests.

In addition, the determination of the probability value distribution in a generic sense for the unique pairs of NP cytotoxicity assays and NDDs was carried out. For this, we depict the surface scatterplot of probability values against histograms of NP length along with NDD hydrophobicity (Figure 6). Generally, a third of the probability values remains in the dark green zone, which represents promising N2D3Ss for further assay. It is worth mentioning that most of the cases (white dots) are hydrophobic drugs (on the left of the graph). This feature is one of the most important physicochemical properties for drugs in order to cross the BBB [89]. High lipophilicity can contribute to excessive distribution volumes, increased metabolic liability, and lower unbound drug concentration in the plasma and/or brain; it may also negatively affect pharmaceutical properties, in particular solubility [90]. Most NDDs of this database are in the PSA_{di} range of 60–120 $Å^2$. Stephen et al. suggested that CNS drugs should have a PSA value below 90 $Å^2$ for a decent BBB perme-

ability, among other physicochemical characteristics such as number of hydrogen bond donors, molecular size, and shape, with smaller contributions from hydrogen bond acceptors [89]. Although this type of graphic is clearly a simplification of the whole database, it offers simple guidelines for researchers concerned with designing NDD compounds or libraries with improved probability of BBB penetration. The size of the vast majority of NPs for NDD delivery in this database is in the range of 70–115 nm. Recently, Chithrani et al. [91] have demonstrated that size, coating, and surface charge of nanoparticles have a crucial impact on the intracellular uptake process. Similarly, Shilo et al. have investigated the influence of NP size on the probability to cross the BBB by using the endothelial brain cell method. The results indicated that the intracellular uptake of NPs strongly depends on the NP size. This characteristic has a direct impact on biomedical applications. When NPs serve as carriers for drug delivery through encapsulation, a larger NP size (70 nm) is needed. However, when NPs serve as carriers by binding drug molecules to their surface, a larger free surface area is required; therefore, the optimal size would be 20 nm [92]. This principle suggests that a high number of the NPs in our database are proper drug delivery carriers by drug encapsulation.

Figure 6: Probability surface scatter plot representing the deviation of NP length considering the partition c_{III}, which describes the NP nature and includes c_{n0} = NP biological activity parameters (e.g., CC_{50}, IC_{50}, LC_{50}, and EC_{50}), c_{n1} = cell lines, c_{n2} = NP morphology, and c_{n3} = NP synthesis conditions. ($\Delta L(c_{III})_{nj}$) along with the deviation of NDD hydrophobicity ($\Delta PSA(c_I)_{dj}$) taking into account the partition c_I, which includes the biological characteristics, for example c_{d0} = NDD biological activity parameters (e.g., IC_{50}, K_i, potency, and time) and c_{d1} = the type of protein involved in NDs.

Thus, the design of new N2D3Ss based on multiple preclinical assays of NP cytotoxicity and NDDs has been carried out successfully. This database involves a high structural and biological diversity, which may help to distinguish active from non-active N2D3Ss. Experimentally, the IFPTML-LDA method predicted with high probability $p(N2D3S_{in})_{cdj.cnj} > 0.81$ all examples reported in Table 5. The results support our initial premise that the IFPTML additive approach is able to carry out an

Table 5: IFPTML analysis of experimentally tested N2D3S compounds.

Drug[a]	NP	c_{d0} = activity	$\Delta DPSA(c_l)$	Obs.[b]	Pred.[c]	p[d]	L (nm)[e]
Metal/n.a.							
2234684	Ag	Time (h)	0.57	1	1	0.88	12.50
2376472	Ag	Time (h)	4.30	1	1	0.88	12.50
2234683	Ag	Time (h)	0.57	1	1	0.88	12.50
Metal oxide/n.a.							
3769671	TiO_2	Cp (nm)	0	1	1	0.94	56
Levodopa	TiO_2	Time (h)	−3.5	1	1	0.93	56
Sch-58261	TiO_2	Time (h)	−1	1	1	0.93	56
2180030	TiO_2	EC_{20} (nm)	0	1	1	0.93	56
Levodopa	TiO_2	Time (h)	−3.5	1	1	0.93	56
Sch-58261	TiO_2	Time (h)	−1	1	1	0.93	56
2234689	TiO_2	Time (h)	0.3	1	1	0.93	56
Morin	TiO_2	Time (h)	0	1	1	0.93	56
Metal/elliptical							
Datiscetin	Ag	Time (h)	0.3	1	1	0.81	36.8
2234993	Ag	Time (h)	0.4	1	1	0.81	36.8
1240582	Ag	Time (h)	−1.7	1	1	0.81	36.8
1241456	Ag	Time (h)	−2.1	1	1	0.81	36.8
Metal oxide/elliptical							
2180030	Yb_2O_3	EC_{20} (nm)	0	1	1	0.90	62.1
Levodopa	Yb_2O_3	Time (h)	−3.5	1	1	0.90	62.1
3769671	CeO_2	Cp (nm)	0	1	1	0.90	44.8
Metal oxide/needle							
3747225	La_2O_3	Time (h)	2.8	1	1	0.89	65.8
3769671	La_2O_3	Cp (nm)	0	1	1	0.88	65.8
Meta/rod							
3218426	Au	Activity (%)	−2.0	1	1	0.93	37.8
Congo red	Au	Inhibition (%)	3.6	1	1	0.93	37.8
3218189	Au	Activity (%)	−2.0	1	1	0.93	37.8
3580774	Au	Activity (nm)	0	1	1	0.93	37.8
Metal oxide/pyramidal							
PGA[f]	TiO_2	Time (h)	−18	1	1	0.91	6.5
Apomorphine	TiO_2	Time (h)	−17	1	1	0.91	50
1801682	TiO_2	Time (h)	−20	1	1	0.91	50

Table 5: IFPTML analysis of experimentally tested N2D3S compounds. (continued)

Metal oxide/irregular

3350757	TiO_2	Time (h)	−5.3	1	1	0.93	21
3747225	TiO_2	Time (h)	2.8	1	1	0.93	21
1243007	TiO_2	Time (h)	−0.7	1	1	0.92	21
3769671	TiO_2	Cp (nm)	0	1	1	0.92	21
Levodopa	TiO_2	Time (h)	−3.5	1	1	0.92	21

Metal Oxide/pseudo-spherical

2376474	CeO_2	Time (h)	3.9	1	1	0.89	8
3747225	CeO_2	Time (h)	2.8	1	1	0.89	8
3769671	CeO_2	Cp (nm)	0	1	1	0.89	8
Levodopa	CeO_2	Time (h)	−3.5	1	1	0.89	8
Sch-58261	CeO_2	Time (h)	−1.0	1	1	0.89	8

Metal/spherical

2151181	Au	ED_{50} (mg/kg)	−0.4	1	1	0.94	42.9
1222303	Au	ED_{50} (mg/kg)	−0.4	1	1	0.94	42.9
2181911	Au	Activity (%)	1.6	1	1	0.90	42.9
3397881	Au	Inhibition (%)	−1.1	1	1	0.90	42.9
3785241	Au	Inhibition (%)	−1.5	1	1	0.90	42.9
3947919	Au	Activity (%)	1.0	1	1	0.90	42.9
3817925	Au	Inhibition (%)	−0.7	1	1	0.90	42.9
3612821	Au	Inhibition (%)	0.3	1	1	0.90	42.9
2159510	Au	Activity (%)	−0.8	1	1	0.90	42.9
2415095	Au	Inhibition (%)	0.5	1	1	0.90	42.9
436483	Au	Inhibition (%)	1.5	1	1	0.90	42.9
2159511	Au	Activity (%)	−1.2	1	1	0.90	42.9
2349470	Au	Activity (%)	−1.8	1	1	0.90	42.9
3127906	Au	Activity (%)	0.6	1	1	0.90	42.9
Propidium	Au	Inhibition (%)	0.4	1	1	0.90	42.9

Metal oxide/spherical

3218188	SiO_2	Activity (%)	91	1	1	0.97	12.5
3087679	SiO_2	Inhibition (%)	69	1	1	0.97	60
3233831	SiO_2	Inhibition (%)	58	1	1	0.97	44
510384	SiO_2	K_i (nm)	−30	1	1	0.97	47.5
81999	SiO_2	K_i (nm)	−40	1	1	0.97	36.8
3218425	SiO_2	Activity (%)	91	1	1	0.97	70
55401	SiO_2	K_i (nm)	−31	1	1	0.97	37
3233829	SiO_2	Inhibition (%)	58	1	1	0.97	36.8
3087678	SiO_2	Inhibition (%)	69	1	1	0.97	3.4
3769671	SiO_2	Cp (nm)	0	1	1	0.99	5.5
2234689	SiO_2	Time (h)	37	1	1	0.99	36.8
2234690	SiO_2	Time (h)	37	1	1	0.99	16.4

[a]ChEMBL ID or drug name; the name of the drug is depicted if it is available, otherwise the ChEMLID code of the drug is indicated, which can be easily consulted by accessing the CheMBL website. [b]Class. Obs: $f(v_{ij}, v_{nj})_{obs}$. [c]Class. Pred: $f(v_{ij}, v_{nj})_{pred}$. [d]p: probability calculated as $p(\text{N2D3S}_{in}/\mathbf{c}_{dj}, \mathbf{c}_{nj})_{pred} = 1/(1 + \exp[−f(v_{ij}, v_{nj})_{calc}]$. [e]$L$ (nm): NP length. [f]PGA: phloroglucin aldehyde.

appropriate recognition of N2D3Ss involving additive and synergic cases.

Conclusion

N2D3Ss are a promising and plausible tool to help conventional NDDs cross the BBB. AI/ML algorithms can be instrumental in expediting the process of designing N2D3Ss. However, scientific literature lacks a sufficient number of real N2D3S experimental cases that characterize complex applications. In this context, the IFPTML model, encompassing both NDDs and NP models, could offer a practical solution. This approach has successfully addressed the challenges posed by the vast number of combinations of NP and NDD compounds and the wide range of conditions to be tested in N2D3S discovery. The results of the IFPTML-LDA and IFPTML-ANN techniques showed satisfactory performance, achieving Sp values of 73.0%–86.1% and Sn values of 70.0%–86.2% in the training and validation series, comprising 375,000 and 125,000 cases, respectively. Moreover, both models are easily accessible and provide logical solutions for predicting putative N2D3Ss. The most successful outcome was observed using non-linear models, specifically, the IFPTML-MLP model, which displayed Sn and Sp values of 85.8–86.2% and an AUROC value of 0.94 in the training and validation series. Furthermore, the analysis of three N2D3Ss samples yielded low SDV values, confirming the robustness of both IFPTML-LDA and IFPTML-ANN. In summary, the IFPTML models offer an initial solution for a rapid and less arduous pre-screening of putative N2D3Ss. This approach is widely utilized to minimize resource costs and save experimental time that would otherwise be spent on testing all possible combinations.

Supporting Information

Supporting Information File 1

Detailed dataset information.
[https://www.beilstein-journals.org/bjnano/content/supplementary/2190-4286-15-47-S1.xlsb]

Funding

Author Contributions

Shan He: formal analysis; investigation; methodology; validation; visualization; writing – original draft. Julen Segura Abarrategi: data curation; resources. Harbil Bediaga: data curation; funding acquisition. Sonia Arrasate: conceptualization; funding acquisition; supervision; writing – review & editing. Humberto González-Díaz: conceptualization; funding acquisition; supervision; writing – review & editing.

ORCID® iDs

Shan He - https://orcid.org/0000-0001-6965-6276
Julen Segura Abarrategi - https://orcid.org/0009-0006-3098-5543
Harbil Bediaga - https://orcid.org/0000-0002-9055-0721
Sonia Arrasate - https://orcid.org/0000-0003-2601-5959
Humberto González-Díaz - https://orcid.org/0000-0002-9392-2797

Data Availability Statement

The data generated and analyzed during this study is openly available in the Figshare repository at https://doi.org/10.6084/m9.figshare.25144544.

Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2024.10.v1

References

1. Chowdhury, A.; Kunjiappan, S.; Panneerselvam, T.; Somasundaram, B.; Bhattacharjee, C. *Int. Nano Lett.* **2017,** *7,* 91–122. doi:10.1007/s40089-017-0208-0
2. Murray, C.; Lopez, A. D. *Bull. W. H. O.* **1994,** *72,* 447–480. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2486698/
3. Zhang, Q.; Qian, W.-J.; Knyushko, T. V.; Clauss, T. R. W.; Purvine, S. O.; Moore, R. J.; Sacksteder, C. A.; Chin, M. H.; Smith, D. J.; Camp, D. G.; Bigelow, D. J.; Smith, R. D. *J. Proteome Res.* **2007,** *6,* 2257–2268. doi:10.1021/pr0606934
4. Aslan, M.; Ozben, T. *Curr. Alzheimer Res.* **2004,** *1,* 111–119. doi:10.2174/1567205043332162
5. Hanumanthappa, R.; Venugopal, D. M.; P C, N.; Shaikh, A.; B.M, S.; Heggannavar, G. B.; Patil, A. A.; Nanjaiah, H.; Suresh, D.; Kariduraganavar, M. Y.; Raghu, S. V.; Devaraju, K. S. *ACS Omega* **2023,** *8,* 47482–47495. doi:10.1021/acsomega.3c04312
6. Contestabile, A.; Ciani, E.; Contestabile, A. *Neurochem. Res.* **2008,** *33,* 318–327. doi:10.1007/s11064-007-9497-4
7. Agnihotri, T. G.; Jadhav, G. S.; Sahu, B.; Jain, A. *Drug Delivery Transl. Res.* **2022,** *12,* 3104–3120. doi:10.1007/s13346-022-01173-y
8. Calzoni, E.; Cesaretti, A.; Polchi, A.; Di Michele, A.; Tancini, B.; Emiliani, C. *J. Funct. Biomater.* **2019,** *10,* 4. doi:10.3390/jfb10010004
9. Polchi, A.; Magini, A.; Mazuryk, J.; Tancini, B.; Gapiński, J.; Patkowski, A.; Giovagnoli, S.; Emiliani, C. *Nanomaterials* **2016,** *6,* 87. doi:10.3390/nano6050087
10. Diéguez-Santana, K.; González-Díaz, H. *Nanoscale* **2021,** *13,* 17854–17870. doi:10.1039/d1nr04178a
11. Cacciatore, I.; Ciulla, M.; Fornasari, E.; Marinelli, L.; Di Stefano, A. *Expert Opin. Drug Delivery* **2016,** *13,* 1121–1131. doi:10.1080/17425247.2016.1178237
12. Asefy, Z.; Hoseinnejhad, S.; Ceferov, Z. *Neurol. Sci.* **2021,** *42,* 2653–2660. doi:10.1007/s10072-021-05234-x
13. Shayganfard, M. *Curr. Pharm. Biotechnol.* **2022,** *23,* 538–551. doi:10.2174/1389201022666210622111028

14. Verma, R.; Sartaj, A.; Qizilbash, F. F.; Ghoneim, M. M.; Alshehri, S.; Imam, S. S.; Kala, C.; Alam, M. S.; Gilani, S. J.; Taleuzzaman, M. *Curr. Drug Metab.* **2022,** *23,* 447–459. doi:10.2174/1389200223666220608142506

15. Syed, A. A.; Reza, M. I.; Singh, P.; Thombre, G. K.; Gayen, J. R. *Curr. Drug Metab.* **2021,** *22,* 561–571. doi:10.2174/1389200222666210203182716

16. Yu, Y.; O'Rourke, A.; Lin, Y.-H.; Singh, H.; Eguez, R. V.; Beyhan, S.; Nelson, K. E. *ACS Infect. Dis.* **2020,** *6,* 2120–2129. doi:10.1021/acsinfecdis.0c00196

17. Pribut, N.; Kaiser, T. M.; Wilson, R. J.; Jecs, E.; Dentmon, Z. W.; Pelly, S. C.; Sharma, S.; Bartsch, P. W., III; Burger, P. B.; Hwang, S. S.; Le, T.; Sourimant, J.; Yoon, J.-J.; Plemper, R. K.; Liotta, D. C. *ACS Infect. Dis.* **2020,** *6,* 922–929. doi:10.1021/acsinfecdis.9b00524

18. Wang, X.; Perryman, A. L.; Li, S.-G.; Paget, S. D.; Stratton, T. P.; Lemenze, A.; Olson, A. J.; Ekins, S.; Kumar, P.; Freundlich, J. S. *ACS Infect. Dis.* **2019,** *5,* 2148–2163. doi:10.1021/acsinfecdis.9b00295

19. Cooper, C. J.; Krishnamoorthy, G.; Wolloscheck, D.; Walker, J. K.; Rybenkov, V. V.; Parks, J. M.; Zgurskaya, H. I. *ACS Infect. Dis.* **2018,** *4,* 1223–1234. doi:10.1021/acsinfecdis.8b00036

20. Duncan, G. A.; Bevan, M. A. *Nanoscale* **2015,** *7,* 15332–15340. doi:10.1039/c5nr03691g

21. Zhou, H.; Cao, H.; Matyunina, L.; Shelby, M.; Cassels, L.; McDonald, J. F.; Skolnick, J. *Mol. Pharmaceutics* **2020,** *17,* 1558–1574. doi:10.1021/acs.molpharmaceut.9b01248

22. Sun, L.; Yang, H.; Cai, Y.; Li, W.; Liu, G.; Tang, Y. *J. Chem. Inf. Model.* **2019,** *59,* 973–982. doi:10.1021/acs.jcim.8b00551

23. Kolesov, A.; Kamyshenkov, D.; Litovchenko, M.; Smekalova, E.; Golovizin, A.; Zhavoronkov, A. *Comput. Math. Methods Med.* **2014,** 781807. doi:10.1155/2014/781807

24. Heider, D.; Senge, R.; Cheng, W.; Hüllermeier, E. *Bioinformatics* **2013,** *29,* 1946–1952. doi:10.1093/bioinformatics/btt331

25. Manganelli, S.; Leone, C.; Toropov, A. A.; Toropova, A. P.; Benfenati, E. *Chemosphere* **2016,** *144,* 995–1001. doi:10.1016/j.chemosphere.2015.09.086

26. Toropova, A. P.; Toropov, A. A.; Rallo, R.; Leszczynska, D.; Leszczynski, J. *Ecotoxicol. Environ. Saf.* **2015,** *112,* 39–45. doi:10.1016/j.ecoenv.2014.10.003

27. Toropova, A. P.; Toropov, A. A.; Veselinović, A. M.; Veselinović, J. B.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. *Ecotoxicol. Environ. Saf.* **2016,** *124,* 32–36. doi:10.1016/j.ecoenv.2015.09.038

28. Rybińska-Fryca, A.; Mikolajczyk, A.; Puzyn, T. *Nanoscale* **2020,** *12,* 20669–20676. doi:10.1039/d0nr05220e

29. Le, T. C.; Yin, H.; Chen, R.; Chen, Y.; Zhao, L.; Casey, P. S.; Chen, C.; Winkler, D. A. *Small* **2016,** *12,* 3568–3577. doi:10.1002/smll.201600597

30. Ahmadi, S.; Toropova, A. P.; Toropov, A. A. *Nanotoxicology* **2020,** *14,* 1118–1126. doi:10.1080/17435390.2020.1808252

31. Ojha, P. K.; Kar, S.; Roy, K.; Leszczynski, J. *Nanotoxicology* **2019,** *13,* 14–34. doi:10.1080/17435390.2018.1529836

32. Sizochenko, N.; Gajewicz, A.; Leszczynski, J.; Puzyn, T. *Nanoscale* **2018,** *10,* 20867–20868. doi:10.1039/c8nr07975g

33. Tasi, D. A.; Csontos, J.; Nagy, B.; Kónya, Z.; Tasi, G. *Nanoscale* **2018,** *10,* 20863–20866. doi:10.1039/c8nr02377h

34. Villaverde, J. J.; Sevilla-Morán, B.; López-Goti, C.; Alonso-Prados, J. L.; Sandín-España, P. *Sci. Total Environ.* **2018,** *634,* 1530–1539. doi:10.1016/j.scitotenv.2018.04.033

35. Sizochenko, N.; Leszczynska, D.; Leszczynski, J. *Nanomaterials* **2017,** *7,* 330. doi:10.3390/nano7100330

36. Manganelli, S.; Benfenati, E. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2017,** *1601,* 275–290. doi:10.1007/978-1-4939-6960-9_22

37. Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H.-M.; Toropov, A.; Leszczynska, D.; Leszczynski, J. *Nat. Nanotechnol.* **2011,** *6,* 175–178. doi:10.1038/nnano.2011.10

38. Sizochenko, N.; Mikolajczyk, A.; Jagiello, K.; Puzyn, T.; Leszczynski, J.; Rasulev, B. *Nanoscale* **2018,** *10,* 582–591. doi:10.1039/c7nr05618d

39. Toropov, A. A.; Toropova, A. P.; Benfenati, E.; Gini, G.; Puzyn, T.; Leszczynska, D.; Leszczynski, J. *Chemosphere* **2012,** *89,* 1098–1102. doi:10.1016/j.chemosphere.2012.05.077

40. Gonzalez-Diaz, H.; Arrasate, S.; Gomez-SanJuan, A.; Sotomayor, N.; Lete, E.; Besada-Porto, L.; Ruso, J. M. *Curr. Top. Med. Chem.* **2013,** *13,* 1713–1741. doi:10.2174/1568026611313140011

41. Alonso, N.; Caamaño, O.; Romero-Duran, F. J.; Luan, F.; D. S. Cordeiro, M. N.; Yañez, M.; González-Díaz, H.; García-Mera, X. *ACS Chem. Neurosci.* **2013,** *4,* 1393–1403. doi:10.1021/cn400111n

42. Diez-Alarcia, R.; Yáñez-Pérez, V.; Muneta-Arrate, I.; Arrasate, S.; Lete, E.; Meana, J. J.; González-Díaz, H. *ACS Chem. Neurosci.* **2019,** *10,* 4476–4491. doi:10.1021/acschemneuro.9b00302

43. González-Díaz, H.; Riera-Fernández, P.; Pazos, A.; Munteanu, C. R. *BioSystems* **2013,** *111,* 199–207. doi:10.1016/j.biosystems.2013.02.006

44. González-Díaz, H.; Herrera-Ibatá, D. M.; Duardo-Sánchez, A.; Munteanu, C. R.; Orbegozo-Medina, R. A.; Pazos, A. *J. Chem. Inf. Model.* **2014,** *54,* 744–755. doi:10.1021/ci400716y

45. González-Díaz, H.; Riera-Fernández, P. *J. Chem. Inf. Model.* **2012,** *52,* 3331–3340. doi:10.1021/ci300321f

46. Concu, R.; D. S. Cordeiro, M. N.; Munteanu, C. R.; González-Díaz, H. *J. Proteome Res.* **2019,** *18,* 2735–2746. doi:10.1021/acs.jproteome.8b00949

47. Martínez-Arzate, S. G.; Tenorio-Borroto, E.; Barbabosa Pliego, A.; Díaz-Albiter, H. M.; Vázquez-Chagoyán, J. C.; González-Díaz, H. *J. Proteome Res.* **2017,** *16,* 4093–4103. doi:10.1021/acs.jproteome.7b00477

48. Quevedo-Tumailli, V. F.; Ortega-Tenezaca, B.; González-Díaz, H. *J. Proteome Res.* **2018,** *17,* 1258–1268. doi:10.1021/acs.jproteome.7b00861

49. Santana, R.; Zuluaga, R.; Gañán, P.; Arrasate, S.; Onieva, E.; González-Díaz, H. *Nanoscale* **2020,** *12,* 13471–13483. doi:10.1039/d0nr01849j

50. Romero Durán, F. J.; Alonso, N.; Caamaño, O.; García-Mera, X.; Yañez, M.; Prado-Prado, F. J.; González-Díaz, H. *Int. J. Mol. Sci.* **2014,** *15,* 17035–17064. doi:10.3390/ijms150917035

51. Luan, F.; Cordeiro, M. N. D. S.; Alonso, N.; García-Mera, X.; Caamaño, O.; Romero-Duran, F. J.; Yañez, M.; González-Díaz, H. *Bioorg. Med. Chem.* **2013,** *21,* 1870–1879. doi:10.1016/j.bmc.2013.01.035

52. Gonzalez-Diaz, H. *Curr. Pharm. Des.* **2010,** *16,* 2598–2600. doi:10.2174/138161210792389261

53. Kleandrova, V. V.; Luan, F.; González-Díaz, H.; Ruso, J. M.; Speck-Planche, A.; Cordeiro, M. N. D. S. *Environ. Sci. Technol.* **2014,** *48,* 14686–14694. doi:10.1021/es503861x

54. Luan, F.; Kleandrova, V. V.; González-Díaz, H.; Ruso, J. M.; Melo, A.; Speck-Planche, A.; Cordeiro, M. N. D. S. *Nanoscale* **2014,** *6,* 10623–10630. doi:10.1039/c4nr01285b

55. Santana, R.; Zuluaga, R.; Gañán, P.; Arrasate, S.; Onieva, E.;
González-Díaz, H. *Nanoscale* **2019,** *11,* 21811–21823.
doi:10.1039/c9nr05070a

56. Santana, R.; Zuluaga, R.; Gañán, P.; Arrasate, S.; Onieva, E.;
Montemore, M. M.; González-Díaz, H. *Mol. Pharmaceutics* **2020,** *17,*
2612–2627. doi:10.1021/acs.molpharmaceut.0c00308

57. Urista, D. V.; Carrué, D. B.; Otero, I.; Arrasate, S.;
Quevedo-Tumailli, V. F.; Gestal, M.; González-Díaz, H.;
Munteanu, C. R. *Biology (Basel, Switz.)* **2020,** *9,* 198.
doi:10.3390/biology9080198

58. Romero-Durán, F. J.; Alonso, N.; Yañez, M.; Caamaño, O.;
García-Mera, X.; González-Díaz, H. *Neuropharmacology* **2016,** *103,*
270–278. doi:10.1016/j.neuropharm.2015.12.019

59. Ortega-Tenezaca, B.; González-Díaz, H. *Nanoscale* **2021,** *13,*
1318–1330. doi:10.1039/d0nr07588d

60. Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.;
Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.;
Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P.
Nucleic Acids Res. **2014,** *42,* D1083–D1090. doi:10.1093/nar/gkt1031

61. Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.;
Atkinson, F.; Bellis, L.; Overington, J. P. *Nucleic Acids Res.* **2015,** *43,*
W612–W620. doi:10.1093/nar/gkv352

62. Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.;
Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.;
Overington, J. P. *Nucleic Acids Res.* **2012,** *40,* D1100–D1107.
doi:10.1093/nar/gkr777

63. Gusenbauer, M.; Haddaway, N. R. *Res. Synth. Methods* **2020,** *11,*
181–217. doi:10.1002/jrsm.1378

64. Lu, Z. *Database* **2011,** baq036. doi:10.1093/database/baq036

65. Islamaj Dogan, R.; Murray, G. C.; Névéol, A.; Lu, Z. *Database* **2009,**
bap018. doi:10.1093/database/bap018

66. Li, Y.; Li, H.; Pickard, F. C., IV; Narayanan, B.; Sen, F. G.;
Chan, M. K. Y.; Sankaranarayanan, S. K. R. S.; Brooks, B. R.; Roux, B.
J. Chem. Theory Comput. **2017,** *13,* 4492–4503.
doi:10.1021/acs.jctc.7b00521

67. Xia, R.; Kais, S. *Nat. Commun.* **2018,** *9,* 4195.
doi:10.1038/s41467-018-06598-z

68. Na, G. S.; Chang, H.; Kim, H. W. *Phys. Chem. Chem. Phys.* **2020,** *22,*
18526–18535. doi:10.1039/d0cp02709j

69. Concu, R.; Kleandrova, V. V.; Speck-Planche, A.; Cordeiro, M. N. D. S.
Nanotoxicology **2017,** *11,* 891–906.
doi:10.1080/17435390.2017.1379567

70. Kleandrova, V. V.; Luan, F.; González-Díaz, H.; Ruso, J. M.; Melo, A.;
Speck-Planche, A.; Cordeiro, M. N. D. S. *Environ. Int.* **2014,** *73,*
288–294. doi:10.1016/j.envint.2014.08.009

71. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.;
Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.;
Davies, M.; Dedman, N.; Karlsson, A.; Magariños, M. P.;
Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R.
Nucleic Acids Res. **2017,** *45,* D945–D954. doi:10.1093/nar/gkw1074

72. Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.;
Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.;
Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.;
Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.;
Segura-Cabrera, A.; Hersey, A.; Leach, A. R. *Nucleic Acids Res.* **2019,**
47, D930–D940. doi:10.1093/nar/gky1075

73. Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. *J. Cheminf.* **2018,**
10, 4. doi:10.1186/s13321-018-0258-y

74. Hill, T.; Lewicki, P. *Statistics: Methods and Applications,* 1st ed.;
StatSoft, Inc.: USA, 2006.

75. Bendel, R. B.; Afifi, A. A. *J. Am. Stat. Assoc.* **1977,** *72,* 46–53.
doi:10.1080/01621459.1977.10479905

76. Gamberger, D.; Lavrac, N. *J. Artif. Intell. Res.* **2002,** *17,* 501–527.
doi:10.1613/jair.1089

77. Huberty, C. J.; Olejnik, S. *Applied MANOVA and discriminant analysis,*
2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006.
doi:10.1002/047178947x

78. Hanczar, B.; Hua, J.; Sima, C.; Weinstein, J.; Bittner, M.;
Dougherty, E. R. *Bioinformatics* **2010,** *26,* 822–830.
doi:10.1093/bioinformatics/btq037

79. Bian, L.; Sorescu, D. C.; Chen, L.; White, D. L.; Burkert, S. C.;
Khalifa, Y.; Zhang, Z.; Sejdic, E.; Star, A. *ACS Appl. Mater. Interfaces*
2019, *11,* 1219–1227. doi:10.1021/acsami.8b15785

80. Alafeef, M.; Srivastava, I.; Pan, D. *ACS Sens.* **2020,** *5,* 1689–1698.
doi:10.1021/acssensors.0c00329

81. Sun, B.; Fernandez, M.; Barnard, A. S. *J. Chem. Inf. Model.* **2017,** *57,*
2413–2423. doi:10.1021/acs.jcim.7b00272

82. Barnard, A. S.; Opletal, G. *Nanoscale* **2019,** *11,* 23165–23172.
doi:10.1039/c9nr03940f

83. He, J.; He, C.; Zheng, C.; Wang, Q.; Ye, J. *Nanoscale* **2019,** *11,*
17444–17459. doi:10.1039/c9nr03450a

84. Yan, T.; Sun, B.; Barnard, A. S. *Nanoscale* **2018,** *10,* 21818–21826.
doi:10.1039/c8nr07341d

85. Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S.
Nanomedicine (London, U. K.) **2015,** *10,* 193–204.
doi:10.2217/nnm.14.96

86. Nocedo-Mena, D.; Cornelio, C.; Camacho-Corona, M. d. R.;
Garza-González, E.; Waksman de Torres, N.; Arrasate, S.;
Sotomayor, N.; Lete, E.; González-Díaz, H. *J. Chem. Inf. Model.* **2019,**
59, 1109–1120. doi:10.1021/acs.jcim.9b00034

87. Leys, C.; Ley, C.; Klein, O.; Bernard, P.; Licata, L.
J. Exp. Soc. Psychol. **2013,** *49,* 764–766.
doi:10.1016/j.jesp.2013.03.013

88. Arnedo, M. *Bol. Soc. Entomol. Aragonesa* **1999,** *26,* 57–84.

89. Hitchcock, S. A.; Pennington, L. D. *J. Med. Chem.* **2006,** *49,*
7559–7583. doi:10.1021/jm060642i

90. Doan, K. M. M.; Humphreys, J. E.; Webster, L. O.; Wring, S. A.;
Shampine, L. J.; Serabjit-Singh, C. J.; Adkison, K. K.; Polli, J. W.
J. Pharmacol. Exp. Ther. **2002,** *303,* 1029–1037.
doi:10.1124/jpet.102.039255

91. Chithrani, B. D.; Ghazani, A. A.; Chan, W. C. W. *Nano Lett.* **2006,** *6,*
662–668. doi:10.1021/nl052396o

92. Shilo, M.; Sharon, A.; Baranes, K.; Motiei, M.; Lellouche, J.-P. M.;
Popovtzer, R. *J. Nanobiotechnol.* **2015,** *13,* 19.
doi:10.1186/s12951-015-0075-7