A MULTI-OMICS VISIBLE DEEP NETWORK FOR DRUG ACTIVITY PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Drug discovery is a challenging task, characterized by a significant amount of time between initial development and market release, with a high rate of attrition at each stage. Computational virtual screening, powered by machine learning algorithms, has emerged as a promising approach for predicting therapeutic efficacy of drugs. However, the complex relationships between features learned by these algorithms can be challenging to decipher. We have devised a neural network model for the prediction of drug sensitivity, which employs a biologically-informed visible neural network (VNN), leveraging multi-omics data and molecular descriptors. The trained model can be scrutinized to investigate the biological pathways that play a fundamental role in prediction, as well as the chemical properties of drugs that influence sensitivity We have extended the model to predict drug synergy, resulting in favorable outcomes while retaining interpretability. Given the often unbalanced nature of publicly available drug screening datasets, our model demonstrates superior performance compared to state-of-the-art visible machine learning algorithms.

1 INTRODUCTION

Understanding the molecular features influencing drug sensitivity is the key element for developing personalized therapies (Liu et al., 2021). Machine learning models can exploit screening datasets to develop predictive algorithms useful to associate omics features with drug response such as Genomics of Drug Sensitivity in Cancer (GDSC) (Yang et al., 2012). Most of these studies use models as "black boxes" optimized for prediction accuracy without the possibility of interpreting the biological mechanisms underlying predicted outcomes. Recently, Ideker and colleagues proposed a "visible neural network" to address this issue (Kuenzi et al., 2020). The model, called *DrugCell*, encodes cells genotypes into a network constituted by modules organized according to the Biological Process Gene Ontology (GO) hierarchy (Ashburner et al., 2000). Interpreting the activity of each module allows the association between specific biological pathways and drug response to be discovered. Besides its great novelty, there are several possibilities to improve this biologically informed approach. First, DrugCell relies on somatic single nucleotide variations profiles of the screened models. Second, it is important to consider the unbalanced nature of the data since sensitivity values tend to be skewed toward values representing a lack of sensitivity. Finally, it should be important for an explanation tool to allow the subsetting of specific cell lines on which queries can be performed. The use of multiple therapies in patient care has become increasingly popular in recent years, as it has the potential to offer both greater efficacy and fewer side effects (Jaaks et al., 2022).

In order to advance the research in this area, we propose a novel approach to drug activity prediction using a Multi-Omics Visible Drug Activity prediction model, or MOViDA. MOViDA extends the existing DrugCell's visible network approach by incorporating pathway activity from gene expression and copy number variation data. This allows for a more comprehensive understanding of the interactions between drugs and biological systems, leading to more accurate predictions of drug activity. To ensure that the data used to train MOViDA is representative and balanced, the training algorithm employs a random sampler based on a multinomial distribution. This helps to account for skewness in the input dataset. In addition, MOViDA also enhances the interpretability of drug descriptions by using fingerprints and molecular descriptors. These descriptors relate the 3D molecular structure of drugs to their physical-chemical and pharmacokinetic properties, making it easier to understand the impact that a drug may have on a biological system. To further enhance the biolog-

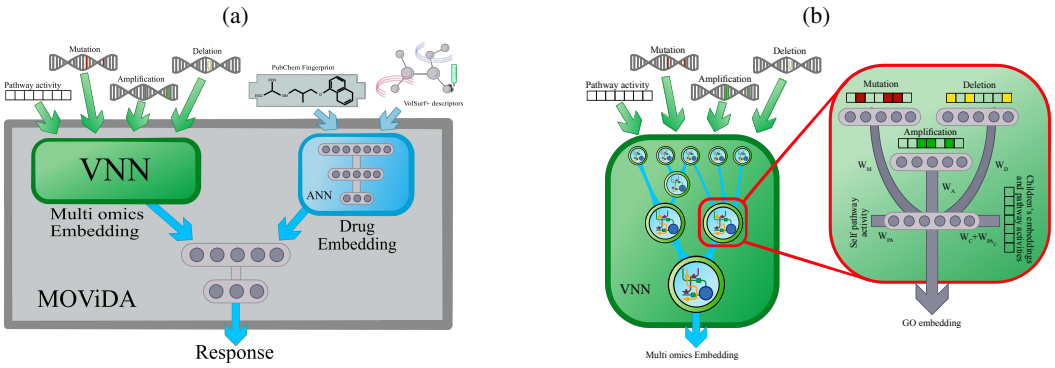


Figure 1: **MOViDA architecture**. a) The network is composed of three distinct sub-networks. The Multi-omics Embedding net utilizes cell line multi-omics profiles. The Drug Embedding net receives PubChem fingerprints and VolSurf+ molecular descriptors. The final layers combine the embeddings and predict the AUC. b) The Multi-omics Embedding net comprises a set of modules which are connected according to the Biological Process Gene Ontology Hierarchy. Each GO sub-module takes in input the cell line multi-omics profile, considering those genes associated to former.

ical interpretability of the model, we developed an *ad hoc* network explanation method to score the pathways affecting sensitivity prediction in specific sets of cell lines. Finally, we have also extended MOViDA to make drug synergy predictions.

2 METHODS

2.1 NETWORK ARCHITECTURE

MOViDA is a feedforward deep neural network that predicts the drug sensitivity of a cell line, which structure is separated into two branches (Figure 1a): a Visible Neural Network (*VNN*) and a feedforward Artificial Neural Network (*ANN*). The *ANN* on the right branch is a shallow neural network exploiting a combination of PubChem fingerprints (pub, 2021) and molecular descriptors relating 3D molecular shape with physical-chemical and pharmacokinetic properties (Crivori et al., 2000). Drug features are encoded into a three-layers neural architecture with 100, 50, and 6 nodes. The *VNN* (Figure 1b) on the left branch represents a Biological Process Gene Ontology hierarchy composed of five layers and 2086 GOs. Each GO is connected to more generic GO ancestors and is represented by a sub-submodule composed of a set of $k+1$ nonlinear units. k units are connected to the input layer and the output of previous layers. Each unit also receives a normalized gene set enrichment score (NES) of that GO term computed from gene expression. This value is concatenated with the activation of the k units and fed to the next layer in the hierarchy. The input layer is composed of nodes of three different kinds: mutations, amplification, and deletions. Each GO sub-module is connected to genes annotated with that term. The activation of the units at the root of the hierarchy represents a multi-omic cell line embedding. The training phase aims at learning the weights of each subsystem. In particular, every unit of each module s has the following output:

$$IM_s = W_M M_s \qquad IA_s = W_A A_s \qquad ID_s = W_D D_s \qquad INES_s = W_{NES} NES_s$$

$$IC_c = W_c E_c + W_{NES_c} NES_c \quad E_s = f(IM_s + IA_s + ID_s + INES_s + \sum_{c \in desc(s)} IC_c + b_s)$$

where: M_s, A_s, and D_s are binary vectors that describe mutation, amplification, and deletion status of genes associated with subsystem s and W_M, W_A and W_D are the corresponding weights; W_{NES} is the weight of the normalized enrichment score NES_s of the term s resulting from gene expression; W_c and W_{NES_c} are the weights associated to the embedding E_c and NES_c of child c of the considered subsystem. E_s is the embedding of a subsystem s, which is a nonlinear transformation f of the inputs consisting of hyperbolic tangent and batch normalization, and b_s is bias term.

A third neural network combines the multi-omics and drug embeddings predicting the cell's response to the drug, measured as the area under the dose-response curve (AUC). During the training phase, the input data was split into three sets: training (80%), testing (10%), and validation (10%) sets.

2.2 DATASETS

We used the Genomics of Drug Sensitivity in Cancer database (GDSC) (Yang et al., 2012) and the Cancer Therapeutics Response Portal v2 (CTRP) (Basu et al., 2013) to collect 383,998 triplets representing cell line, drug, and cell survival after treatment measure as AUC value. Overall, our dataset contains 889 cell lines and 684 drugs. Each drug is represented by 1009 variables, 881 molecular fragments from PubChem fingerprints (pub, 2021) , and 128 molecular descriptors from the software VolSurf+ (Crivori et al., 2000), as detailed in Table S1. To represent the molecular properties of a cell line, we use the mutation and copy number profiles stored in three binary vectors, where the value corresponds to the presence or absence of a mutation/deletion/amplification in a particular gene in a given cell line, which were downloaded from the GDSC data portal (Yang et al., 2012). We selected 4870 (top 2.5%) frequently mutated genes in cancer using the pan-cancer compendium encompassing 33 cancer types and more than 10,000 tumor-normal exome pairs (Ellrott et al., 2018). Analogously, 2612 and 3625 genes contained in focal recurrently amplified copy number segments and deleted copy number segments respectively, selected as described in (Iorio et al., 2016). These genes were further filtered for those associated with at least one GO term present in the MOViDA hierarchy, obtaining 2931 and 2097 genes for amplifications and deletions, respectively. Gene expression was also used to compute a normalized enrichment score (NES) using single-sample gene set test using the Mann–Whitney–Wilcoxon Gene Set test (mww-GST) available in the yaGST package (Frattini et al., 2018). NES is an estimate of the probability that the expression of a gene in the geneset is greater than the expression of a gene outside this set: $NES = 1 - \frac{U}{mn}$ where m is the number of genes in a gene set, n is the number of those outside the gene set, $U = mn + m(m+1) - T$, and T is the sum of the ranks of the genes in the gene set.

For drug combination, we used Therapeutic Target Database (TTD) (Zhou et al., 2021) to identify potential synergies among drug targets and then used the dataset of pharmaceutical synergies specific to breast, colon, and pancreatic cancer cells created by Jaaks et al. (Jaaks et al., 2022) for validation. To further advance our model's capabilities, we extended its application to predict drug combination therapies utilizing the dataset presented by O'Neil et al. (O'Neil et al., 2016). We selected the cell lines and drugs with available features, resulting in a dataset of 32 compounds and 32 cell lines, totaling 13376 instances of triplets, with 1296 instances considered synergistic. To assess the synergistic interaction between drugs, we employed Loewe Additivity score (LOEWE, 1953), utilizing a threshold of 30 to differentiate synergistic from non-synergistic outcomes. To overcome limitations posed by the limited size of training set, we took steps to reduce the number of input features, specifically by excluding copy number information from the input.

2.3 DATA UNBALANCE STRATEGIES

Drug sensitivity data exhibits a significant skewness, characterized by many screens with low sensitivity outcomes (AUC close to 1) and very few with high sensitivity (AUC close to 0). To mitigate the potentially deleterious effects of this data unbalancing during the training, we used a weighted random sampler based on a multinomial distribution estimated from the data.

The AUC sensitivity scores are divided into twelve equally spaced bins between 0 and 1.2, and we used the inverse frequencies with additive smoothing to fix the weights of the multinomial sampler:

$$f_i = \frac{s_i}{\sum_{j=0}^{c} s_j} \qquad v_i = \frac{1}{f_i} + \epsilon \qquad w_i = \frac{v_i}{\sum_{j=0}^{c} v_j} \tag{1}$$

Where s_i is the number of samples in bin i, c is the number of bin, f_i is the relative frequency of the bin i, ϵ is the smoothing penalty term. We also used a *weighted loss* function to give more importance to errors associated with lower scores of ground truth and predictions. Hence we adopted the following *double-weighted MSE loss*.

$$L(p,t) = max(w_{c_p}, w_{c_t}) * (p-t)^2 \tag{2}$$

Here, p is model prediction, t is the ground truth, w_{c_p} and w_{c_t} are weights associated with corresponding c_p and c_t bins as computed in Equation (1). This loss function guarantees higher weights for errors when either the ground truth or the prediction are in a class with few samples.

If an evaluation measure conceived for balanced datasets was used, a trivial system assigning the majority of the items the values with the highest frequency may even outperform genuinely engineered systems. Evaluating measures that can handle imbalance are recommended to avoid these cases, so we adopted a measure developed in the field of ordinal regression (Baccianella et al., 2009), which is the *macroaverage MSE*, based on a sum of the classification errors across classes. This is motivated by the fact that sensitivity classes can be considered ordinal variables, and the ordering between the values is significant, as they represent discretized degrees of sensitivity.

$$MSE^M = \frac{1}{c}\sum_{j=0}^{c}\frac{1}{|c_j|}\sum_{x\in c_j}(p_x - t_x)^2 \tag{3}$$

where c_j represents the set of samples in class j, c is the number of classes, t_x id the ground truth of sample x and p_x is its prediction. The macroaverage MSE does not depend on the frequency of each class, as every class contributes to $1/c$ of the total measure. Therefore trivial assignments are penalized, whereas to have better MSE^M the errors in all classes should be minimized.

2.4 Model Explanation

The Biology informed nature of MOViDA allows performing accurate post-hoc analyses, enabling us to identify the biological processes that contribute to the prediction of a cell line's drug sensitivity the most. Therefore, we developed an interpretation score, relative improvement score (***RIS***), specifically tailored for our model that measures the relative contribution of a sub-module concerning its children in the GO hierarchy. This score is inspired by *ablative brain surgery* (Meyes et al., 2019), which involves removing components of the brain while keeping its functions intact.

First, we calculate the prediction for a specific drug-cell line pair. Then we recalculate the prediction after silencing the output of each sub-module one by one, setting weights and biases to 0. Similarly, we silence all children subsystems for each GO. In the case of leaf nodes, we silence the corresponding inputs. The ***RIS*** score expresses the importance of a term during the prediction phase and its ability to combine the information from its children, comparing the deviations from the actual prediction of the ablated models. The ***RIS*** is computed as follows:

$$s_f = |p_f - p| \qquad s_{c_f} = |p_{c_f} - p| \qquad RIS = \frac{s_{c_f} - s_f}{s_{c_f} + s_f} \tag{4}$$

where p is the effective prediction for a specific drug-cell line pair, p_f and p_{c_f} is the prediction obtained by silencing a GO subsystem and its children, respectively. RIS is the interpretation score. For a given GO subsystem, positive RIS values correspond to a larger deviation in predictions when silencing children compared to the father. The advantages of RIS over the score adopted in DrugCell (RLIPP) are that: i) it can be calculated for each individual drug-cell line pair; ii) there are multiple ways to aggregate these values, by drug or by specific cell line types.

To further investigate the model, we have inspected all the elements that compose the inputs describing the drugs. The importance score of each feature was performed using DeepLift, whose implementation is based on the algorithm of Shrikumar et al. (Shrikumar et al., 2017) and gradient formulation proposed by Ancona et al. (Ancona et al., 2017).

2.5 Drug Combination Strategies through Relevant Subsystems

By combining multiple drugs, it is possible to harness cancer's resistance to particular anticancer drugs, and create a more effective treatment plan. A recent effort toward understanding the combined effects of drugs has been carried out by Jaaks et al. (Jaaks et al., 2022) by creating a huge dataset of pharmaceutical synergies specific to breast, colon, and pancreatic cancer cells.
Taking into account a drug and a cell line evaluated in the combination dataset, we used the RIS score calculated from a drug/cell-line pair. We selected the top 5 enriched GO terms along with associated genes. From the collection of drug targets *Therapeutic Target Database* (Zhou et al., 2021), we determine the drugs targeting the genes associated with the previous selection of GO terms, marking them as potentially synergistic for that drug-cell line pair.
We then compared our predictions on the synergistic dataset, marking the right (TP) and wrong (FP) combinations and comparing TP over FP and the ratio of synergistic drugs to non-synergistic drugs

to understand if the former was significantly higher than the latter. We took all drug combinations studied for a specific cell line and drug and counted how many of these combinations were synergistic (S) and how many are not (NS). We applied the binomial test on TP over (TP + FP), which is the Precision metric, with probability equal to S/(NS+S), thus accounting for the number of synergistic combinations. The p-values for the binomial test and the enrichment scores $\frac{TP/(TP+FP)}{S/(NS+S)}$ of the above-described tests are used in the volcano plots reported in the Results section.

2.6 Extension for Drug Synergy Prediction

We have extended our model to see its actual efficacy in predicting synergistic effects of drug combinations as well. For this purpose, we exploited the concept of Siamese neural networks, which involves the use of identical subnetworks, in this case the drug embedding branch of our model. The final ANN concatenates the cell line and drugs embedding. In this way, the order of past drugs is important, so we doubled the initial dataset by considering the two possible combinations of drug pairs. In addition, the problem was set up as a binary class prediction, distinguishing the cases where there is synergy or not. Given the presence of imbalance in the dataset, we used focal loss (Lin et al., 2017) and the previously described weighted random sampler to counteract this phenomenon.

3 Results

3.1 Dataset Unbalance

MOViDA is trained to predict the drug response of a cellular model (represented by its molecular profile), measured as area under the curve (AUC). AUC combines information about the potency and efficacy of the drug into a single measure (Fallahi-Sichani et al., 2013). A value close to 0 means high sensitivity, a value close to 1 represents no effect of the drug, if higher than 1, the drug has the effect of promoting cell viability. Besides the high interest in accurate predictions for drugs with high sensitivity, the majority of drug screens present AUC values that are close to 1, therefore, presenting a marked skewness. Figure S1a shows the distribution of the AUC values present in the dataset after binning AUC values into 12 bins (ten bins for the interval between 0 and 1 and other two for values greater than 1): class 0 (AUC scores in the range [0.0, 0.1]) is 80 times less populated than class 9 (scores in the range [0.9, 1.0]). To mitigate this effect, our approach considers a weighted random sampler and a double-weighted loss (Section 2.3). Both use the weights calculated as a function of the inverse frequencies of each class plus a smoothing term ϵ. Figure S1b shows the number of samples for all classes: besides the raw case (no weights), different scenarios are depicted by varying the $epsilon$ parameter that affects the weights. The ideal scenario lies between the raw case and the perfectly balanced dataset (with $epsilon$ set to 0), which, on the contrary, could produce too many sample repetitions. After parameter tuning, we chose the value of ϵ to 80 as a good compromise, producing, on average, a four-fold repetition for the samples in a less represented class.

3.2 Performance and comparison with DrugCell

For the comparison, we used pre-trained DrugCell, downloaded from `https://github.com/idekerlab/DrugCell`. The accuracy was evaluated by measuring the Spearman and Pearson correlation between the predicted AUC values and the actual ones. To assess the issue of unbalance, 100 samples from each class were sampled, repeating this process for 1000 runs. The results show that Pearson and Spearman correlation was 0.88 and 0.89 respectively. DrugCell model showed good results as well, with Pearson correlation of 0.85 and Spearman correlation of 0.86 (Table S2). We show in Figure 2b that MOViDA has a significantly lower macroaverage MSE (MSE^M) (60%, 0.02 vs. 0.05), which accounts for the class unbalance. Indeed MOViDA can make much more accurate predictions of the AUC in classes with fewer training examples (classes between zero to 6). Those classes are the most meaningful ones as they represent cases of high sensitivity to drugs. Whereas for classes 7-9, MOViDA behaves similarly to DrugCell. To better explain this behavior, by recasting the regression as a classification problem in terms of prediction of AUC interval classes, the confusion matrix in the Figures S2a and S2b show that DrugCell tends to over-estimate the majority class. In contrast, MOViDA reaches better accuracy along the cells on the diagonal.

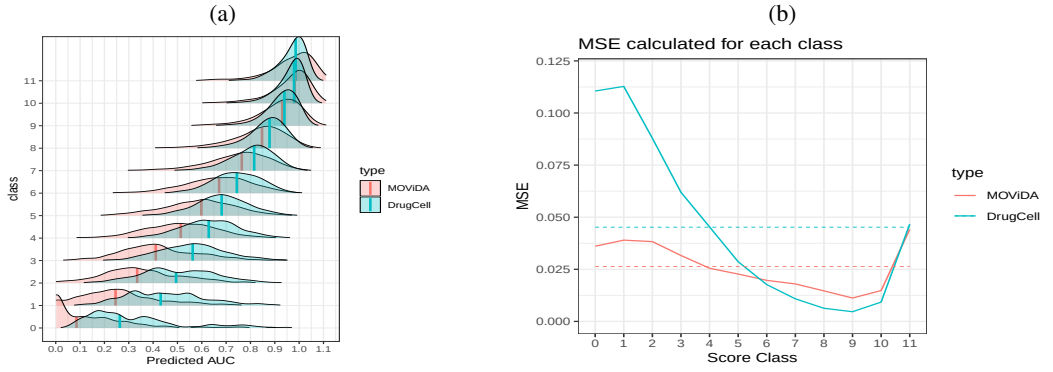


Figure 2: **Evaluation and comparison** a) Distribution of predicted AUC compared to ground truth classes. b) MSE calculated for each class, dashed lines correspond to macroaverage MSE.

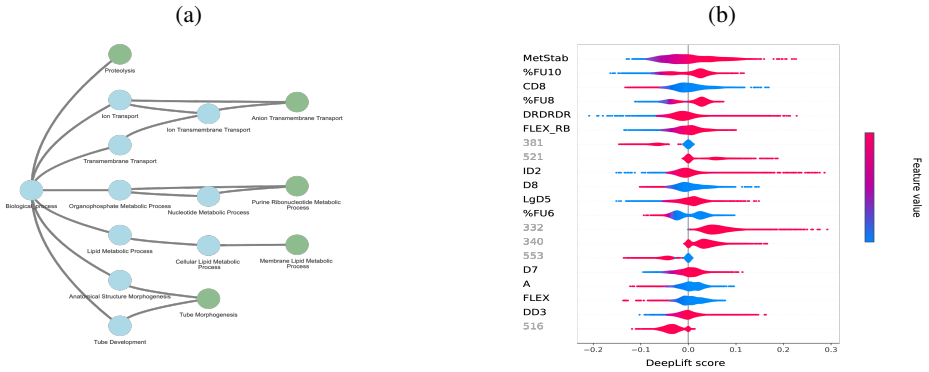


Figure 3: **Explainability**. a) Top 5 RIS score associated to GOs (green nodes), considering the ALLSIL cell line and GSK1070916 drug. The whole sub-tree (blue nodes) is displayed. b) Deep Lift drug feature interpretation of Liver tissue. The color represents the feature value for a specific prediction. If feature values are low (blue) on the left and high (red) on the right of the violin, AUC is directly dependent on the feature. Feature names are colored differently if they are VolSurf+ descriptors or PubChem fingerprint bits in black and gray, respectively.

The performance of MOViDA was further evaluated (Figures S3) through comparative analyses, varying the type of drug representation used as input (Morgan Fingerprint, PubChem Fingerprint, and VolSurf+ descriptors). Results indicated that MOViDA exhibited the lowest MSE^M compared to other models. We also evaluated MOViDA with a larger artificial neural network (ANN) for drug embedding. We found that MOViDA with fewer parameters generalized better. Finally, MOViDA was compared with a standard shallow network consisting of 6 linear layers with ReLU activation functions to assess whether the trade-off between explainability and performance exists. The results indicated that MOViDA and the shallow network performed similarly, with comparable MSE^M.

3.3 The RIS score identifies pathway dependencies in specific cellular models

We have implemented a novel interpretability score called relative improvement score (RIS), which relates deviations from prediction by first removing a GO module and then the modules associated with its children. This score also has the advantage of being calculated for each specific cell line-drug prediction, so we can show which GOs are most predictive for a specific tissue (represented as a group of cells) or drug. Among the leukemia cell lines, we selected the ALLSIL cell line highly sensitive to GSK1070916, an ATP-competitive inhibitor of Aurora kinase, which is important during cell division. The RIS scores associated with this prediction revealed that *anion transmembrane*

transport (GO:0098656) is among the most important modules for prediction (Figure 3a). The overexpression of ATP-binding cassette (ABC) transporters, particularly ABCG2, contributes to reduced cytotoxicity of GSK1070916 (Wu et al., 2021). The family of these genes is responsible for transporting substances across the cell membrane using the energy produced by ATP electrolysis. Interestingly, ALLSIL is ABCC9 mutant which, together with ABCG2, is downregulated in this cell line. Similarly, *proteolysis* (GO:0006508) has a high RIS score for this cell line-drug pair. This can be attributed to AURKB (aurora kinase B) phosphorylating caspase-2 by mediating its proteolysis (Lim et al., 2021). As a result, cell division is not stopped. In our case, GSK1070916 inhibits AURKB promoting apoptosis of the cancer cell. AURKB is over-expressed in the ALLSIL cell line. An high RIS score is also reported for the *positive regulation of the reactive oxygen species metabolic process (ROS)* pathway (GO:2000379) associated with the DB cell line (lymphoma) when administered with Dinaciclib (an inhibitor of CDK1, CDK2, CDK5 and CDK9). Interestingly, it has been recently reported that the inhibition of CKD leads to increased mitochondrial ROS levels, confirming this pathway's importance in the cellular response to this exposure (Riess et al., 2021).

3.4 Drug features interpretability

As a complementary interpretation step, we can also measure the impact of individual drug features on the model's predictions utilizing DeepLift score (Shrikumar et al., 2017). Figure 3b shows the 20 most important features of our model. The importance lies in the variability of the score for the various cell lines: the more it varies, the more significant it is for predictions. The most relevant feature is the VolSurf+ descriptor METSTAB for all the cell lines. Such descriptor refers to metabolic stability (measured on human liver microsomes), mostly due to isoform 3A4 of the cytochrome P450 system. We have noticed a direct relationship between such a feature with the AUC. This means low values for metabolic stability (thus, fast CYP3A4-mediated metabolism) for high-sensitivity drugs. This agrees with the absorption, distribution, metabolism, and excretion (ADME) profile of many anticancer drugs, most of which are metabolized in the liver by CYP3A4. Several features refer to drug lipophilicity; among these, the VolSurf+ descriptors D8 and CD8 refer to highly lipophilic regions of the molecules, and characteristics of active molecules (low AUC values).

Two features refer to molecular flexibility, namely the VolSurf+ descriptors FLEX and FLEX_RB. Given their lift values, we can argue that for most of the predictions, flexibility is inversely related to AUC, whereas the number of rotatable bonds is directly related to AUC. Although it is uncommon to have an opposite behavior for these two features, an attempt to generalization may be that anticancer drugs are generally flexible but with a low number of rotatable bonds (compared to the overall number of bonds). The VolSurf+ descriptors %FU8 and %FU10 can measure the percent of the unionized fraction at a given pH (8 or 10). According to violin colors, the system identified a direct relationship with AUC; in other words, many anticancer drugs have strong or weak acid groups, that is reflected onto the significant presence of ionized species at basic pH.

3.5 Drug Combination predictions

MOViDA predictions can be used to uncover potential drug synergies. Given the interpretation scores for specific drug-cell line pairs, we select genes involved in GO terms with highest scores and prioritize as potential combinations the drugs targeting these genes (Methods 2.5), reported in Supplementary Material. Volcano plot in Figure 4a shows cell line-drug pairs for which the candidate molecules are enriched for experimentally validated synergistic drugs. For example, the synergy predictions associated with the breast cancer cell line JIMT1 and the drug MK-2206, a highly selective inhibitor of Akt1/2/3among, has among the top 5 scoring GO categories the GO:0007169 (transmembrane receptor protein tyrosine kinase signaling pathway) and the GO:0007584 (response to nutrient). Our model selects Lapatinib, PD173074, Axitinib, Linsitinib, Sapitinib, and OSI-027 as potential candidates for combination therapy with MK-2206. They are all tyrosine kinase inhibitors involved in tumor cell growth. The association between these drugs and MK-2206 is well documented in the literature, as many tyrosine kinases are part of the PI3 kinase-AKT cascade, affecting mTOR activity (Lara et al., 2015). Another relevant combination consists of Navitoclax and Vorinostat associated with the MDAMB231 cell line. The latter is an HDAC inhibitor, which decreases the expression of BCL2 family proteins, as described by (Duan et al., 2005). Since Navitoclax is an inhibitor of this anti-apoptotic protein family, it has been shown that its efficacy, combined with Vorinostat, can induce apoptosis in cancer cells (Nakajima et al., 2016).

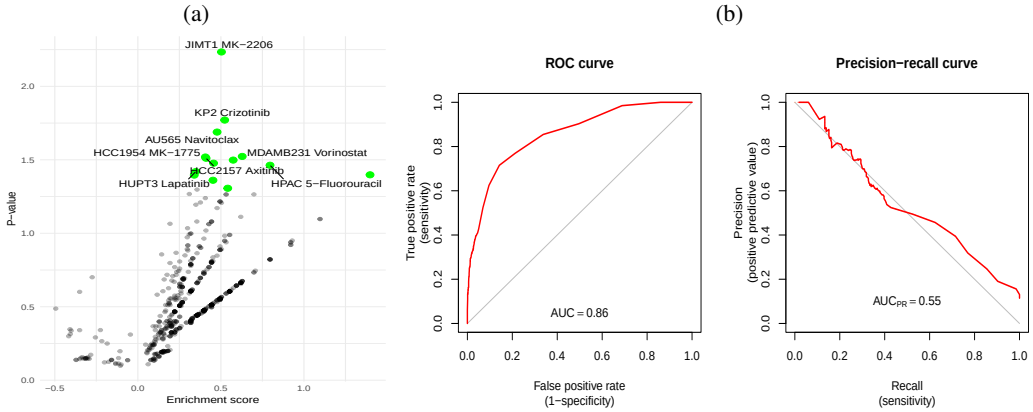


Figure 4: **Evaluation of Drug synergy prediction**. a) Enrichment scores against the p-values of binomial, testing the Precision of the model, using as the probability the percentage of synergistic combination. The green points correspond to drug-cell lines pairs that have a significant number of positive drug candidates. b) AUC scores were calculated for ROC and PR curves, which are graphical evaluations of binary classification model performance. PR shows precision vs recall, while ROC plots TPR (recall) vs FPR.

MOViDA was initially developed as a regressor for predicting drug sensitivity in cancer cells. However, in order to better support the discovery of new and effective cancer therapies, the model has been extended to classify drug synergy as well. To evaluate the performance of the extended model, we used two commonly used metrics in machine learning: area under the Receiver Operating Characteristics (ROC) and Precision-Recall (PR) Curves. The results showed that MOViDA was able to achieve a ROC of 0.86 and PR of 0.55 (Figure 4b), which indicates high accuracy in classifying drug synergy despite the unbalanced nature and small size of our dataset, which is often a challenge in drug discovery research.

4 Conclusion

In this manuscript, we presented MOViDA, a biologically inspired neural network architecture for the prediction of drug sensitivity of cellular models of cancer. The assessment of anti-cancer drugs and the identification of potential synergistic effects can be ideally assessed by using patient-derived cell lines (Liu et al., 2016). However, this process requires substantial time, and there is no guarantee of efficiency. The use of machine learning (ML) to exploit the variety of screening data already available, together with the knowledge of the molecular features of cellular models, can help to accelerate the process of drug prioritization for experimental validation (Dezső & Ceccarelli, 2020) and candidate combination therapies (Jaaks et al., 2022). The adoption of a biologically informed architecture has three main advantages: a) it allows to uncover the role of specific pathways in response to drug stimuli; b) it improves the trust in predictions, especially among non-ML experts and c) the efficient parameterization of our model can simplify the learning process rather than use arbitrarily overparameterized, architectures for prediction, simplifying interpretability. Most drug sensitivity prediction models only use gene expression data (Chen et al., 2021), however, the effect of single nucleotide mutations, DNA methylation and DNA copy number variation on drug sensitivity should also be considered. Here we have presented a visible neural with an improved accuracy level due to the use of multiple omics platforms and the better handling of unbalance of data. We also have developed an interpretability score that has the advantage of producing a value for every cell line-drug pair and, therefore, can be summarized in terms of cellular models derived from the same tissue/cancer subtype or at the level of individual drugs. We have shown that our score produces meaningful results that can be the subject of experimental follow-up. We have also introduced a set of features that can be directly related to behavior or chemical groups. We confirmed the importance of chemical features such as LogP, FLEX as well as %FU(4-10) also observed in the inhibition of glycoprotein (Broccatelli et al., 2011).

REFERENCES

Pubchem substructure fingerprint description. https://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.pdf, 2021.

Marco Ancona, Enea Ceolini, A. Cengiz Öztireli, and Markus H. Gross. A unified view of gradient-based attribution methods for deep neural networks. *CoRR*, abs/1711.06104, 2017. URL http://arxiv.org/abs/1711.06104.

Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Evaluation measures for ordinal regression. In *2009 Ninth international conference on intelligent systems design and applications*, pp. 283–287. IEEE, 2009.

Amrita Basu, Nicole E Bodycombe, Jaime H Cheah, Edmund V Price, Ke Liu, Giannina I Schaefer, Richard Y Ebright, Michelle L Stewart, Daisuke Ito, Stephanie Wang, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*, 154(5):1151–1161, 2013.

Fabio Broccatelli, Emanuele Carosati, Annalisa Neri, Maria Frosini, Laura Goracci, Tudor I. Oprea, and Gabriele Cruciani. A novel approach for predicting p-glycoprotein (abcb1) inhibition using molecular interaction fields. *Journal of Medicinal Chemistry*, 54(6):1740–1751, 2011. doi: 10.1021/jm101421d. URL https://doi.org/10.1021/jm101421d. PMID: 21341745.

Yaojia Chen, Liran Juan, Xiao Lv, and Lei Shi. Bioinformatics research on drug sensitivity prediction. *Frontiers in Pharmacology*, 12:799712, 2021.

P Crivori, G Cruciani, P A Carrupt, and B Testa. Predicting blood-brain barrier permeation from three-dimensional molecular structure. *J Med Chem*, 43(11):2204–2216, June 2000.

Zoltán Dezső and Michele Ceccarelli. Machine learning prediction of oncology drug targets based on protein and network properties. *BMC bioinformatics*, 21(1):1–12, 2020.

Hong Duan, Caroline A. Heckman, and Linda M. Boxer. Histone deacetylase inhibitors down-regulate bcl-2¡/i¿ expression and induce apoptosis in t(14;18) lymphomas. *Molecular and Cellular Biology*, 25(5):1608–1619, 2005. doi: 10.1128/MCB.25.5.1608-1619.2005. URL https://journals.asm.org/doi/abs/10.1128/MCB.25.5.1608-1619.2005.

Kyle Ellrott, Matthew H Bailey, Gordon Saksena, Kyle R Covington, Cyriac Kandoth, Chip Stewart, Julian Hess, Singer Ma, Kami E Chiotti, Michael McLellan, et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell systems*, 6(3):271–281, 2018.

Mohammad Fallahi-Sichani, Saman Honarnejad, Laura M Heiser, Joe W Gray, and Peter K Sorger. Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nature chemical biology*, 9(11):708–714, 2013.

Véronique Frattini, Stefano M Pagnotta, Jerry J Fan, Marco V Russo, Sang Bae Lee, Luciano Garofano, Jing Zhang, Peiguo Shi, Genevieve Lewis, Heloise Sanson, et al. A metabolic function of fgfr3-tacc3 gene fusions in cancer. *Nature*, 553(7687):222–227, 2018.

Francesco Iorio, Theo A Knijnenburg, Daniel J Vis, Graham R Bignell, Michael P Menden, Michael Schubert, Nanne Aben, Emanuel Gonçalves, Syd Barthorpe, Howard Lightfoot, et al. A landscape of pharmacogenomic interactions in cancer. *Cell*, 166(3):740–754, 2016.

Patricia Jaaks, Elizabeth A. Coker, Daniel J. Vis, Olivia Edwards, Emma F. Carpenter, Simonetta M. Leto, Lisa Dwane, Francesco Sassi, Howard Lightfoot, Syd Barthorpe, Dieudonne van der Meer, Wanjuan Yang, Alexandra Beck, Tatiana Mironenko, Caitlin Hall, James Hall, Iman Mali, Laura Richardson, Charlotte Tolley, James Morris, Frances Thomas, Ermira Lleshi, Nanne Aben, Cyril H. Benes, Andrea Bertotti, Livio Trusolino, Lodewyk Wessels, and Mathew J.

Garnett. Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature*, 603(7899):166–173, Mar 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04437-2. URL https://doi.org/10.1038/s41586-022-04437-2.

Brent M Kuenzi, Jisoo Park, Samson H Fong, Kyle S Sanchez, John Lee, Jason F Kreisberg, Jianzhu Ma, and Trey Ideker. Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer cell*, 38(5):672–684, 2020.

P. N. Lara, J. Longmate, P. C. Mack, K. Kelly, M. A. Socinski, R. Salgia, B. Gitlitz, T. Li, M. Koczywas, K. L. Reckamp, and D. R. Gandara. Phase II Study of the AKT Inhibitor MK-2206 plus Erlotinib in Patients with Advanced Non-Small Cell Lung Cancer Who Previously Progressed on Erlotinib. *Clin Cancer Res*, 21(19):4321–4326, Oct 2015.

Yoon Lim, Dylan De Bellis, Jarrod J. Sandow, Luisa Capalbo, Pier Paolo DÁvino, James M. Murphy, Andrew I. Webb, Loretta Dorstyn, and Sharad Kumar. Phosphorylation by aurora b kinase regulates caspase-2 activity and function. *Cell Death & Differentiation*, 28(1):349–366, Jan 2021. ISSN 1476-5403. doi: 10.1038/s41418-020-00604-y. URL https://doi.org/10.1038/s41418-020-00604-y.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

Ruishan Liu, Shemra Rizzo, Samuel Whipple, Navdeep Pal, Arturo Lopez Pineda, Michael Lu, Brandon Arnieri, Ying Lu, William Capra, Ryan Copping, et al. Evaluating eligibility criteria of oncology trials using real-world data and ai. *Nature*, 592(7855):629–633, 2021.

Xiaoming Liu, Jiasheng Yang, Yi Zhang, Yun Fang, Fayou Wang, Jun Wang, Xiaoqi Zheng, and Jialiang Yang. A systematic study on drug-response associated genes using baseline gene expressions of the cancer cell line encyclopedia. *Scientific reports*, 6(1):1–12, 2016.

S. LOEWE. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung*, 3(6):285–290, Jun 1953.

Richard Meyes, Melanie Lu, Constantin Waubert de Puiseau, and Tobias Meisen. Ablation studies in artificial neural networks. *arXiv preprint arXiv:1901.08644*, 01 2019.

Wataru Nakajima, Kanika Sharma, Mark A. Hicks, Ngoc Le, Rikiara Brown, Geoffrey W. Krystal, and Hisashi Harada. Combination with vorinostat overcomes abt-263 (navitoclax) resistance of small cell lung cancer. *Cancer Biology & Therapy*, 17(1):27–35, 2016. doi: 10.1080/15384047.2015.1108485. URL https://doi.org/10.1080/15384047.2015.1108485. PMID: 26575826.

Jennifer O'Neil, Yair Benita, Igor Feldman, Melissa Chenard, Brian Roberts, Yaping Liu, Jing Li, Astrid Kral, Serguei Lejnine, Andrey Loboda, William Arthur, Razvan Cristescu, Brian B. Haines, Christopher Winter, Theresa Zhang, Andrew Bloecher, and Stuart D. Shumway. An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies. *Molecular Cancer Therapeutics*, 15(6):1155–1162, 06 2016. ISSN 1535-7163. doi: 10.1158/1535-7163.MCT-15-0843. URL https://doi.org/10.1158/1535-7163.MCT-15-0843.

Christin Riess, Dirk Koczan, Björn Schneider, Charlotte Linke, Katharina del Moral, Carl Friedrich Classen, and Claudia Maletzki. Cyclin-dependent kinase inhibitors exert distinct effects on patient-derived 2d and 3d glioblastoma cell culture models. *Cell Death Discovery*, 7(1):54, Mar 2021. ISSN 2058-7716. doi: 10.1038/s41420-021-00423-1. URL https://doi.org/10.1038/s41420-021-00423-1.

Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning important features through propagating activation differences. In *International conference on machine learning*, pp. 3145–3153. PMLR, 2017.

Zhuo-Xun Wu, Qiuyan Mai, Yuqi Yang, Jing-Quan Wang, Hansu Ma, Leli Zeng, Zhe-Sheng Chen, and Yihang Pan. Overexpression of human atp-binding cassette transporter abcg2 contributes to reducing the cytotoxicity of gsk1070916 in cancer cells. *Biomedicine & Pharmacotherapy*,

136:111223, Apr 2021. ISSN 0753-3322. URL https://www.sciencedirect.com/science/article/pii/S0753332221000081.

Wanjuan Yang, Jorge Soares, Patricia Greninger, Elena J. Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, Dave Beare, James A. Smith, I. Richard Thompson, Sridhar Ramaswamy, P. Andrew Futreal, Daniel A. Haber, Michael R. Stratton, Cyril Benes, Ultan McDermott, and Mathew J. Garnett. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*, 41(D1):D955–D961, 11 2012. ISSN 0305-1048. doi: 10.1093/nar/gks1111. URL https://doi.org/10.1093/nar/gks1111.

Ying Zhou, Yintao Zhang, Xichen Lian, Fengcheng Li, Chaoxin Wang, Feng Zhu, Yunqing Qiu, and Yuzong Chen. Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Research*, 50(D1):D1398–D1407, 10 2021. ISSN 0305-1048. doi: 10.1093/nar/gkab953. URL https://doi.org/10.1093/nar/gkab953.

A KEY POINTS

- The adoption of a biologically informed architecture allows to uncover the role of specific pathways in response to drug stimuli and it improves the trust in predictions.
- Machine learning methods aimed at predicting sensitivity to drugs should take in account the unbalanced nature of large-scale drug screening datasets.
- Molecular drug features and multi-omics embedding can significantly improve the accuracy of sensitivity predictions.

B AVAILABILITY AND IMPLEMENTATION

MOViDA is implemented in Python using PyTorch library and freely available for download at https://github.com/Luigi-Ferraro/MOViDA.
RIS score and drug features are archived on Zenodo https://doi.org/10.5281/zenodo.7406327

C COMPETING INTERESTS

No competing interest is declared.

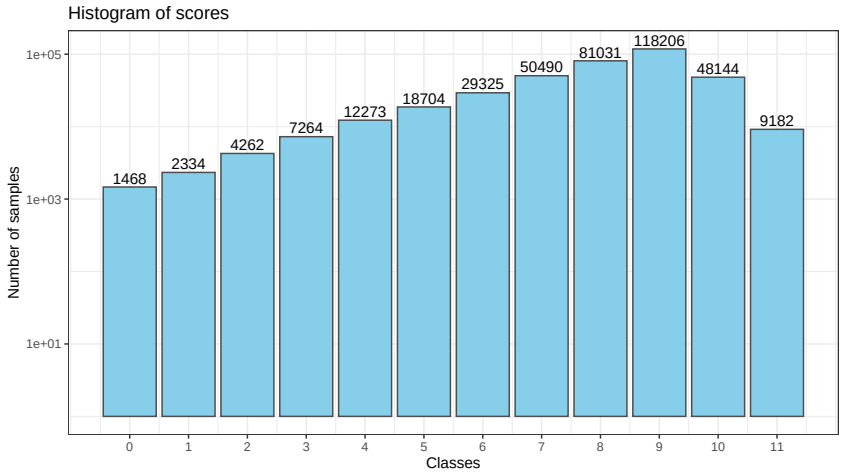
D AUTHOR CONTRIBUTIONS STATEMENT

L.F. and M.C. conceived the study, L.F. developed the code and generated the results, G.S., L.C. and E.C. analysed the results, L.F., E.C. and M.C. wrote the manuscript, all authors reviewed the manuscript.

E ACKNOWLEDGMENTS

We thank Associazione Italiana per la Ricerca sul Cancro for the support. We thank Molecular Discovery Ltd. for supplying the VolSurf+ software (https://www.moldiscovery.com/software/vsplus).

(a)



(b)

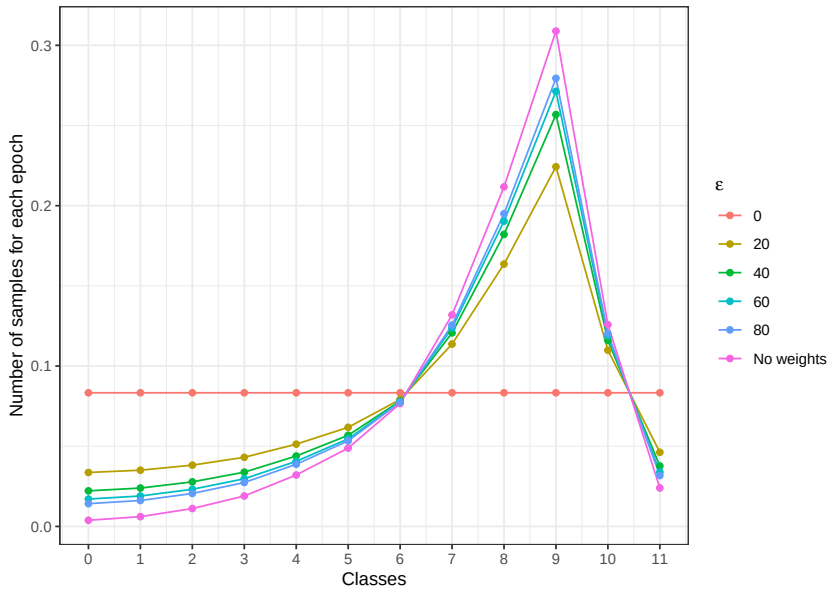
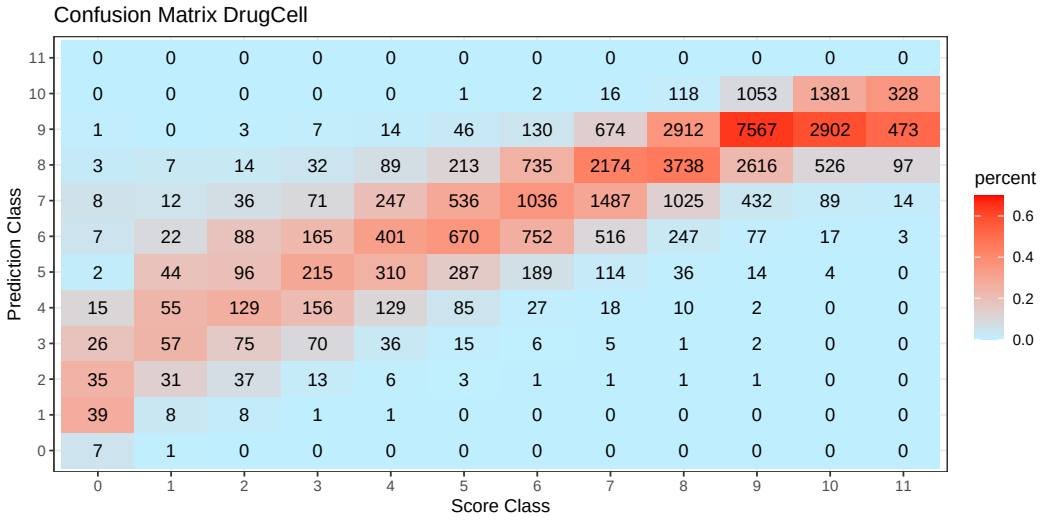


Figure S1: **Data Unbalancing review**. a) BarPlot distribution of AUCs divided in 12 classes, showing high data skewness, with very few examples for classes corresponding to high drug sensitivity (low AUC values). b) Probability to pick up a sample of a specific class by varying ϵ parameter. It is computed as $w_i * s_i$ normalized.

(a)



(b)

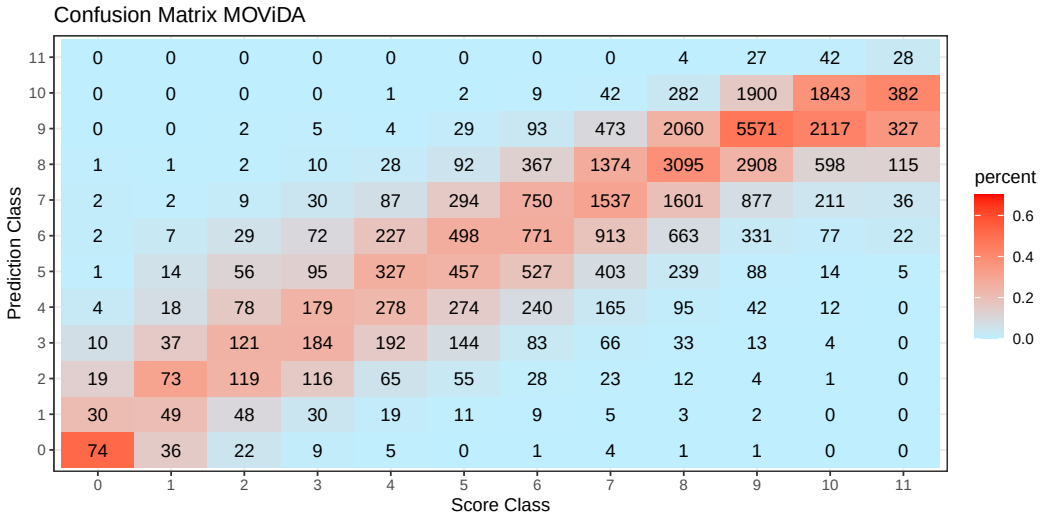


Figure S2: **Evaluation and comparison** Confusion matrices of DrugCell and MOViDA respectively, the percentage of each box is calculated based on the numerosity of the reference score class

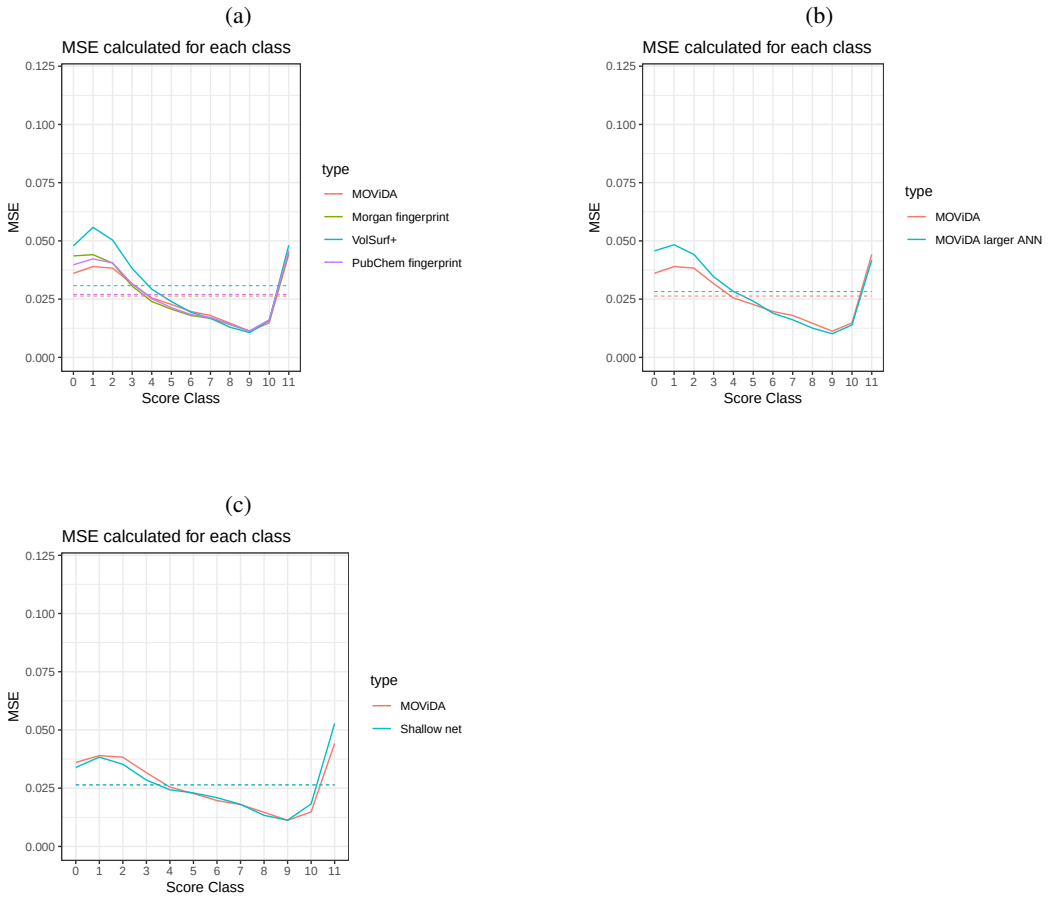


Figure S3: **Evaluation and comparison** MSE computed for each class with dashed line correspondind to the macroaverage MSE. Comparison of MOViDA: a) varying drug input (Morgan fingerprint, VolSurf+, PubChem fingerprint; b) with a larger ANN for drug embedding with 4 linear layers of 512, 128, 32, 8 nodes respectively; c) with a shallow network, composed of six layers and ReLU as activation functions.

Table S1: Illustration of VolSurf+ descriptors

VolSurf+ descriptors	Description
Size and shape descriptors	Molecular volume, surface, rugosity, globularity (how much the molecule is spheroidal), flexibility parameters
Hydrophilic regions descriptors	Molecular hydrophilic volumes, Capacity factors (ratio of hydrophilic surface over the total molecular surface)
Hydrophobic regions descriptors	Molecular hydrophobic volumes, Capacity factors (ratio of hydrophobic surface over the total molecular surface), the difference between the maximum conformational hydrophobic volumes and the hydrophobic volumes
INTEraction enerGY (INTEGY) moments	Unbalance between the center of mass of a molecule and the barycentre of its hydrophilic or hydrophobic regions
Descriptors of H-bond donor/acceptor regions	The molecular envelope generating attractive H-donor or H-bond acceptor interactions
Mixed descriptors	Hydrophilic-Lipophilic balance (ratio between hydrophilic and hydrophobic regions), Amphiphilic moment, Critical packing parameter (ratio between the hydrophilic and lipophilic part of a molecule), average molecular polarizability,, dispersion of chemical in water fluid, Molecular Weight, Log P 1octanol/water, Log P cyclohexane/water, Log D, Polar and Hydrophobic Surface Areas
Charge State descriptors	Number of Charged Centers, Available Uncharged Species, % unionised species
3D pharmacophoric descriptors (TOPP)	Dry, H-bond donor, H-bond acceptor and mixed Dry, H-bond donor and acceptor 3D triplets pharmacophoric areas
ADME model descriptors	Intrinsic solubility, Solubility at various pH, Solubility profiling coefficients (distinguish compounds that present similar solubility but different pH-depended profile or vice-versa), CACO2 permeability, Skin permeability, % of protein binding, Volume of Distribution, High Throughput Screening Flag

Table S2: Comparison between MOViDA and DrugCell using: macroaverage MSE (MMSE), Pearson correlation and Spearman correlation

Model	MMSE	Pearson	Spearman
MOViDA	0.02	0.88	0.89
DrugCell	0.05	0.85	0.86