

SESSION D:

AUTOMATIC CONTROL, IT AND ARTIFICIAL INTELLIGENCE

Drug clustering upon toxicity effects on liver spheroids – a deep learning approach

Miroslava Jordović Pavlović^{1*}, Tijana Geroski², Milica Nikolić^{1,3}

¹ University of Kragujevac, Faculty of Mechanical and Civil Engineering in Kraljevo, Kraljevo, Serbia

² University of Kragujevac, Faculty of Engineering, Kragujevac, Serbia

³ University of Kragujevac, Institute for Information Technologies Kragujevac, Kragujevac, Serbia

ARTICLE INFO

* Correspondence: jordovicpavlovic.m@mfkv.kg.ac.rs

DOI: <https://doi.org/10.46793/ET26.D11JP>

PAGES: D87 – D96

ABSTRACT

Motivation for this research stems from drug-induced liver injury (DILI) being a leading cause of drug failure and market withdrawal, which calls for reliable, non-invasive hepatotoxicity assessment. The basis for this research was the work performed by Dubinsky et al., who established a deep learning (DL) model to predict ATP values (non-invasive estimation of cell viability), assess damage to liver spheroids, and predict hepatotoxicity based on spheroid images. They created a public database containing approximately 20,000 brightfield microscopy images of human liver spheroids exposed over a period of up to 7 days to 108 drug compounds for assessing cell viability. Our work focused on clustering applied compounds by the intensity of damage caused to liver spheroids (level of predicted ATP values), forming groups of compounds that have similar effects on liver cells and potentially similar mechanisms of toxicity. Then we explored whether selected drugs belong to the same class of pharmacotherapeutic classification – analgesics, antipyretics, antibiotics, etc. The model can detect if some drugs behave differently than expected, leading to detecting errors and revealing potential new phenomenological causes. Further, we analysed the temporal dependence of drugs on cell viability (toxicity), providing insights into the dynamics of effects and toxicity progression over time. Future work will investigate drug effects in combination, since co-exposure of drugs can lead to different outcomes of hepatotoxicity, in comparison to individual exposure.

KEYWORDS

Liver spheroid, Hepatotoxicity, Deep learning, Clustering of drugs by toxicity effects, Toxicity dynamics

1. INTRODUCTION

The liver represents the main organ in charge of detoxification processes in the organism, responsible for processing ingested food, drinks and medications and, therefore, it is exposed to a potential harmful influence by these processed substances but also can be affected by environmental pollutants [1], [2]. Changing of human habits can significantly reduce the risk of harmful influences originating from food, drinks and environmental pollution, but adverse effects of medications are more difficult to avoid [3]. It is clinically confirmed that some drugs can cause hepatotoxicity [4], but there is also evidence of toxic effects of some drugs in a laboratory set up, affecting liver cells function and metabolisms, although they have not yet been confirmed as highly toxic drugs in clinical testing [5]. The early detection of detrimental effects is particularly important as well as analysis of applied drugs, specifically its molecular structure and chemical composition, that could be further linked to causing damage to the liver.

Since the goal is to eliminate or reduce liver toxicity caused by applied medications, researchers established a safer approach in studying the field - *in vitro* testing systems, such as liver spheroids and other liver tissue models [6]. Development and further improvement of the *in vitro* testing system, that can at a certain level recapitulate behaviour of living organs, significantly improved understanding of biological mechanisms in different organs, multi-organ systems, and testing and studying of mechanisms alerted by applied drugs. In that sense, drug testing became more effective and in accordance with the 3R principles (Replace, Reduce, Refine), which are part of the ethical guidance in animal research [7].

Hepatotoxicity stands for chemical-driven liver damage and can appear when the liver is exposed to substances like medications, alcohol, supplements or industrial (environmental) chemicals. In this paper our focus is on the analysis of toxic drug effects to the liver spheroids, therefore, drugs hepatotoxicity. Overall symptoms of injured and malfunctioned liver, such as nausea, vomiting, yellowing of the skin, swelling of the abdominal area, appear with the increased level of liver injury - middle to high [3], moving investigation of drugs hepatotoxicity to the early phase - analysis of drug toxic effects to the liver cells after exposure and assessing liver damage from low to middle and high levels in laboratory conditions [8]. Because it is difficult to detect and diagnose low level liver injury in patients, a list of well-known drugs that can have hepatotoxic effects, together with the list of drugs that cause liver damage in the laboratory conditions can be used to create a computational *in silico* model to draw a conclusion about specific molecular structure section and chemical composition responsible for these adverse effects, which could be further useful in analysing novel drugs hepatotoxicity levels computationally prior to experimental testing.

Literature survey revealed that further optimisation of testing drug safety is supported by development of *in silico* physics-based and AI-based computational models. Machine learning approaches have proven effective across a wide range of scientific and engineering domains [9], [10], [11] demonstrating their capability to model complex non-linear relationships in high-dimensional data. Geci et al. 2025 [12] integrated *in vitro* and *in silico* approaches to predict DILI upon drug exposures. They used an *in silico* model to predict pharmacokinetics of applied drugs and linked it to experimentally measured cytotoxicity, mitochondrial toxicity and bile salt export pump (BSEP) inhibition and obtained high DILI predictivity. A retrospective analysis of 241 drugs showed that the ratio of the *in vivo* maximum plasma concentration (C_{max}) to the lowest *in vitro* toxicity strongly correlates with clinical DILI risks. Then they used the *in silico* model to predict C_{max} and to analyse DILI levels, prospectively. Dynamic simulation through BSEP inhibition can identify drugs potentially causing DILI. Such an integrative, mechanistic approach enhanced DILI prediction and interpretability, offering an animal-free alternative for early drug development. Bergen et al. 2025 [13] analysed toxicogenomic resource for DILI prediction by developing ToxPredictor, a toxicogenomic framework combining RNA-seq data from primary human hepatocytes with pharmacokinetic data to predict dose-resolved DILI risks and safety margins, providing mechanistic insights into hepatotoxic pathways, enabling early de-risking and actionable safety decisions. The model flagged several compounds that later failed in late-stage clinical development because of liver toxicity, including drugs missed by traditional animal testing. This suggests human-cell transcriptomics may provide more clinically relevant toxicity signals than conventional preclinical approaches. Xiao et al. 2025 [14] also used toxicogenomic data but linked them to protein-protein interaction networks, explicitly modelling biological relationships among genes. The results obtained from this research have raised a significant observation that incorporating known molecular pathways can significantly improve prediction of hepatotoxicity and strengthen biological interpretability. Han et al. 2025 [15] created a graph neural network (GNN) model and combined it with foundation-model representations of molecules, which was pre-trained with approximately 10 million chemicals in the PubChem database [16]. They showed that the model can predict hepatotoxicity and that resulting platform enables both toxicity prediction and mechanistic exploration. Malka-Markovitz et al. 2025 [17] developed physics-based liver model to analyse DILI processes that integrates patient-specific anatomical data to predict acetaminophen-induced liver injury as a demonstration of its capability - a multi-scale model bridging vascular, lobular and cellular scales to simulate dynamic blood flow, drug transport, and injury mechanisms that accurately reflect clinically observed spatial heterogeneity. Even though AI models are these days in the focus of researching hepatotoxicity, physics-based models are still dominant when it comes to mechanistic explanation, since they can properly address that gap by connecting molecular events, cellular damage, tissue architecture, and organ-level outcomes within one framework, improving confidence in model predictions and supporting regulatory acceptance.

In this paper, we created a multimodal explainable AI model for hepatotoxicity prediction standing between two major directions in drug hepatotoxicity research - image-based toxicity prediction and chemoinformatic toxicity approach, as we combined both in one analysis. Image-based analysis and open-source data provided by Dubinsky et al. 2026 [1] enable biological aspects for the analysis closer to reality because many DILI mechanisms emerge only after interaction with liver cells. Our research doesn't analyse gene expression changes, but morphological changes which are more clinically observable and sensitive to downstream cellular damage. In comparison against Physiologically Based Pharmacokinetic (PBPK)/mechanistic models, our approach learns kinetics from features, which is a lack of the model and might be improved in future research. By fusing features extracted from images with physicochemical features extracted for applied drugs to the liver spheroids in our machine learning (ML) framework, our goal is to observe from images what happened to the liver spheroids and then analyse physicochemical features to explain why this happened - identify the mechanisms behind the cellular processes and sequence of drug molecular structure and chemical composition responsible for it. Further we identify toxic effects by drug groups and explore whether a related group of drugs develops the same mechanism of toxicity upon assessment of cell viability in liver spheroids.

2. METHODOLOGY

To create a ML framework that captures both morphological changes in liver spheroids and physicochemical properties of applied drugs we started with research conducted by Dubinsky et al. 2026 [1], since they created a public database containing approximately 20,000 brightfield microscopy images of human liver spheroids exposed to 108 drugs over a period of up to 7 days. Out of these 20,000 brightfield images, 2359 images were labelled and used in their model for training and testing purposes. This labelled set includes images taken from four types of microscopes. Images taken from

one microscope, echo rebel, were used for training purposes in their model (1022 images). The rest of them, from the other three microscopes, were used for testing the model. They developed a DL model for predicting ATP values by doing non-invasive estimation of cell viability, assessing damage to liver spheroids upon obtained morphology and predicting hepatotoxicity upon spheroid images.

By analysing labelled dataset, we concluded that in total there are 21 different drugs applied to the spheroids (all in the test part, 1337 images) and they are not equally distributed by days (most of the images when drugs were applied were taken at day 7). In Dubinsky's training set, drugs are not explicitly defined, but rather grouped in four merck compounds groups and for that reason not applicable to our concept, where we tend to extract physicochemical features for specific drugs. The other issue with our approach is that the set contains more images without applied drugs, namely medium and dimethyl sulfoxide (dmsO), and that can cause disbalance in our analysis. Therefore, from the test labelled set, we created our own set by taking out only images when drugs were applied and randomly adding to it a certain amount of medium and dmsO images. In total, out of 1337 images we extracted 320 images with applied drugs, plus randomly selected from the remaining image number 68 medium and 28 dmsO images. Our dataset in total is composed out of 416 images, containing 21 drugs. Features extracted from images, 512 in total per image, were obtained by running our dataset on Dubinsky's model.

The next phase in our approach was to extract physicochemical features for selected 21 drugs. For it, we used the RDKit library [18]. In total 25 physicochemical features were selected, and they are provided in Table 1. All features are dimensionless, except molecular weight (MolWt) which is in grams and topological polar surface area (TPSA) which is in square angstrom.

Table 1: List of physicochemical features extracted from RDKit library

No.	Name of the feature in RDKit library	Meaning	Interpretation of increased value
1.	MolWt	Molecular weight	Reduced membrane permeability, biliary excretion, cholestasis and fibrosis risk
2.	MolLogP	Lipophilicity	Higher membrane permeability, accumulation, fatty liver, mitochondrial toxicity, cholestasis, and reactive metabolite formation
3.	TPSA	Topological polar surface area	Forming hydrogen bonds. More polar, lower membrane permeability, lower lipophilicity
4.	NumHDonors	Number of hydrogen bond donors	More polar, lower membrane permeability, lower lipophilicity, increasing immune-mediated interactions
5.	NumHAcceptors	Number of hydrogen bond acceptors	More polar, lower membrane permeability, lower lipophilicity, increasing interactions with metabolic enzymes or immune pathways
6.	NumRotBonds	Number of Rotatable Bonds	Extensive CYP-mediated oxidation, formation of reactive metabolites
7.	HeavyAtomCount	Number of heavy (non-hydrogen) atoms	Larger molecular size, increasing molecular complexity, biliary excretion, cholestasis, chronic accumulation, reactive intermediates
8.	FractionCSP3	Fraction of sp ³ -hybridized carbons	lower aromaticity, reducing formation of reactive metabolites, decreasing mitochondrial and oxidative stress
9.	RingCount	Number of ring structures (aromatic or aliphatic)	Rigid and planar structure, reactive metabolite formation, mitochondrial stress and cholestasis risk
10.	CalcNumAromaticRings	Number of Aromatic Rings	planar, rigid, and lipophilic structure, reactive metabolite formation, mitochondrial stress, cholestasis
11.	GetFormalCharge	Formal Charge of the Molecule	Polar, low membrane permeability and hepatic uptake, transport-mediated or immune-mediated pathways or lipophilicity-driven hepatotoxicity
12.	NumHeteroatoms	Number of heteroatoms	Polarity and hydrogen-bonding capacity, immune-mediated toxicity
13.	BertzCT	Bertz Complexity Index	More complex molecular topology, reactive metabolite formation, higher metabolic burden and potential hepatotoxicity
14.	BalabanJ	Balaban's Topological Index	More rigid, interconnected, and aromatic topology, reactive metabolite formation, mitochondrial stress, cholestasis
15.	aromatic_ratio	Aromatic atom ratio	More planar, rigid, hepatic accumulation, reactive metabolite formation, mitochondrial stress, cholestasis
16.	CalcNumAliphaticRings	Number of aliphatic rings	Reducing planarity, aromatic reactive metabolites, lower hepatotoxicity risk
17.	CalcNumSaturatedRings	Number of saturated rings	Reducing planarity, aromatic reactive metabolites, lower hepatotoxicity risk
18.	len(pattern_nitro)	Number of nitro groups	Reactive metabolite formation, nitro reduction, idiosyncratic toxicity, DNA/protein adduct formation
19.	len(pattern_aniline)	Number of aniline substructures	Bioactivation to reactive nitroso, hydroxylamine, and iminoquinone intermediates, protein adducts, idiosyncratic hepatotoxicity, CYP-mediated oxidation
20.	len(pattern_quinone)	Number of quinone substructures	Increasing electrophilicity, ROS generation, oxidative stress, protein and DNA adduct, mitochondrial toxicity, cholestasis and idiosyncratic toxicity
21.	len(pattern_michael)	Number of Michael acceptor motifs	Increasing electrophilicity, increasing oxidative stress, protein adducts, mitochondrial toxicity, covalent binding and idiosyncratic hepatotoxicity
22.	len(pattern_halide)	Number of halide substituents	Increasing lipophilicity, slowing clearance, undergoing bioactivation, cholestasis and mitochondrial stress

23.	len(pattern_epoxide)	Number of epoxide sub-structures	Covalent binding, reactive toward nucleophiles, protein adducts, reactive metabolites, mutagenicity, carcinogenicity and idiosyncratic hepatotoxicity
24.	len(pattern_thiophene)	Number of thiophene rings	Undergo S-oxidation or ring epoxidation, covalent binding, immune-mediated liver injury and idiosyncratic hepatotoxicity
25.	len(pattern_phenol)	Number of phenol substructures	Undergo phase I oxidation, protein/DNA adducts, idiosyncratic liver injury

Out of 416 images we randomly selected two images for each drug (42 in total) and 8 medium images and 4 dmso images to serve as a test set in our framework. By excluding these images, the remaining set of 362 images was used as a training set. With a defined training dataset, we performed training using XG-Boost algorithm and applied 5-fold cross-validation (ratio 80% training to 20% testing). Hyperparameter optimisation was performed using the Optuna framework with 100 trials of Bayesian search (TPE sampler), minimising mean absolute error (MAE) across cross-validation folds. Output from the model is predicted ATP values. As a reference ATP values, we used actual experimental measurement for these samples. We calculated accuracy from our model and MAE and RMSE. We presented obtained values in the Results section.

Our training and testing set were combined and analysed upon morphological features to which we fused extracted physicochemical features. In total, the generated csv file contained 416 images (rows) and 541 features (additional 4 features - drug name, drug concentration, day at which image is taken, check whether image is control image, then 512 image features and 25 physico-chemical features). We clustered the dataset into four clusters with the plan to map it to four classes, representing toxicity levels - no toxicity, low toxicity, medium toxicity and high toxicity. Mapping of clusters was performed upon calculated average ATP level per cluster, such as cluster with highest average ATP value is assigned to no toxicity class, imposing that cluster with lowest ATP value is assigned to high toxicity class.

3. RESULTS

The optimised model achieved a cross-validation MAE of 58.78 (normalised MAE of 20.5% relative to the standard deviation of ATP values in the training set), RMSE of 84.7, and R^2 of 0.909. On the external test set, comprising 21 drugs not seen during training, the model demonstrated notably higher predictive accuracy, with MAE of 24.50 (NMAE of 8.6%), RMSE of 52.19, and R^2 of 0.968. The agreement between predicted and reference ATP values is illustrated in Figure 1, where most data points closely follow the identity line ($y = x$), indicating no systematic bias across the full range of ATP values. Minor local deviations are observed in the intermediate range (200–600 ATP units), which can be attributed to intra-compound variability between the two representative images selected per drug for the test set. The most influential features were predominantly morphological (z-space) features derived from spheroid images, with physico-chemical descriptors contributing a smaller but non-negligible share, consistent with the hypothesis that spheroid morphology encodes most of the drug-induced viability information. The higher R^2 on the external test set (0.957) compared to cross-validation (0.909) can be attributed to the composition of the test set, which contains two representative images per drug, selected to cover diverse exposure conditions, whereas cross-validation folds include a broader range of samples with higher intra-compound variability in morphological features across different concentrations and time points.

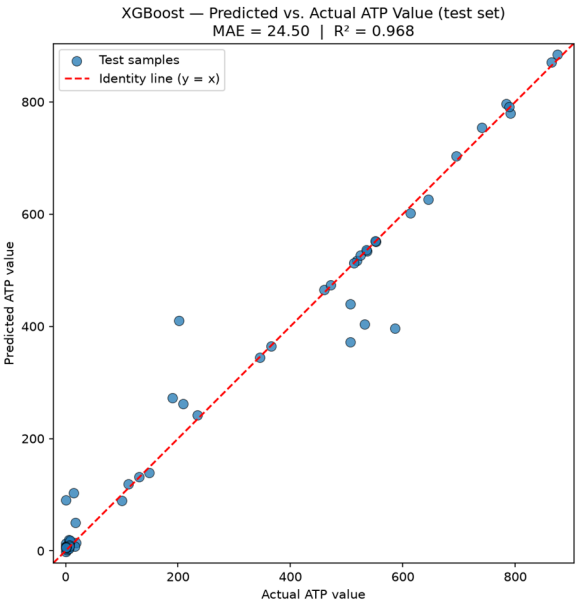


Figure 1: Comparison of predicted ATP values obtained from the XG-Boost model against corresponding experimentally measured ATP values, including MAE and R^2 .

Classification of clusters was performed upon calculated average ATP value by cluster. ATP values are taken from a labelled data set, since they were measured experimentally. ATP value is assigned to each row of dataset used in our model. Then the average value of ATP is calculated for liver spheroids belonging to individual clusters. In the trend of

high to low average value, clusters were mapped to four classes - no toxicity, low toxicity, medium toxicity, high toxicity (Figure 2). A high level of ATP means better viability of cells and indirectly, a less toxic environment.

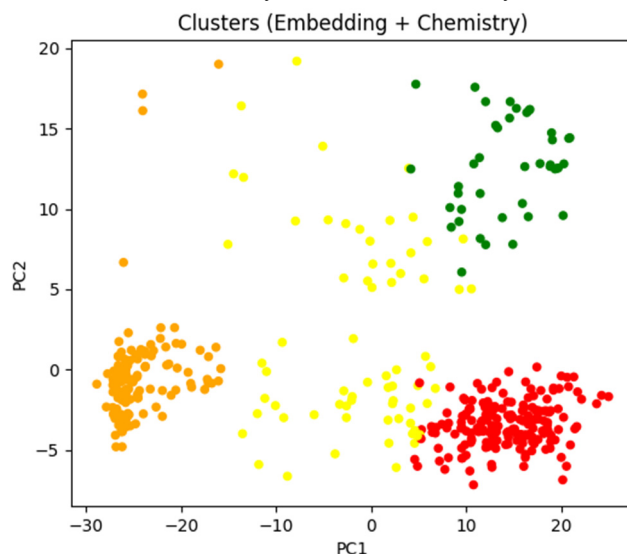


Figure 2: Clustering of used data in the model and mapping to four toxicity classes. Green cluster - no toxicity, yellow cluster - low toxicity, orange cluster - medium toxicity, red cluster - high toxicity

From Figure 2 can be concluded that the high-toxicity and the medium toxicity clusters are nicely grouped, while there are more dispersive groups of the no toxicity and the low toxicity clusters. The low toxicity class can correlate with all other three clusters, and it signals that it was the most challenging to distinguish from other classes.

We analysed the content of each cluster and elaborated on specific drug groups by exploring whether selected drugs belong to the same class of pharmacotherapeutic classification – analgesics, antipyretics, antibiotics, etc. The model can detect if some drugs behave differently than expected, which can be useful in detecting errors and potential new phenomenological causes. First, we analysed compounds that belong to the high toxicity class. All spheroids with applied rosiglitazone and spectinomycin belong to this class (100%). Majority of spheroids with applied methotrexate (93.75%), neostigmine (93.75%), carbamazepine (87.5%), minoxidil (87.5%), nadolol (81.25%), imipramine hcl (75%) and nitrofurantoin (68.75%) are presented in the high toxicity class. Around half of the spheroids with applied amitriptyline hydrochloride (56.25%), desipramine hydrochloride (50%) and pyridostigmine (43.75%) showed high-toxicity effects; other half of them is mainly in the medium toxicity class. A smaller presence of phenformin (31.25%) is noted in the high toxicity class (mostly represented in the low toxicity class). Procyclidine (25%), metergoline (25%) and tamoxifen (6.25%) are noted in the high toxicity class, but they are mainly positioned in the medium toxicity class. For most compounds that are present in the high toxicity class there is clinical evidence of high or moderate hepatotoxicity. Some of them are not clinically known to cause hepatotoxicity, but *in vitro* models showed they can cause morphological changes to exposed liver spheroids, which are dependent on concentration and time of compound exposure. Figure 3 summarises all spheroids of the high toxicity class collected by their pharmacotherapeutic type and clinically known evidence of hepatotoxic effects.

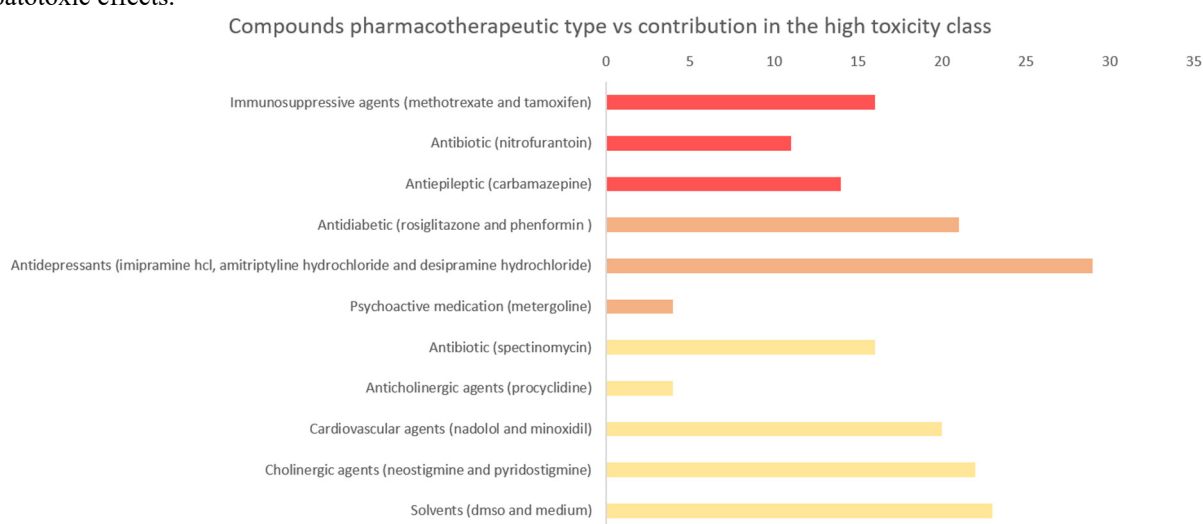


Figure 3: The high toxicity class spheroids grouped by pharmacotherapeutic types (number of occurrence) and clinically known hepatotoxic effects. Red colour - high hepatotoxic effects, orange colour - moderate hepatotoxic effects and yellow colour - low hepatotoxic effects

We did the same analysis for the medium toxicity class and observed less compounds with known high hepatotoxic effects, most of them with known moderate effects and several with low clinical evidence of hepatotoxicity, as it is presented in Figure 4.

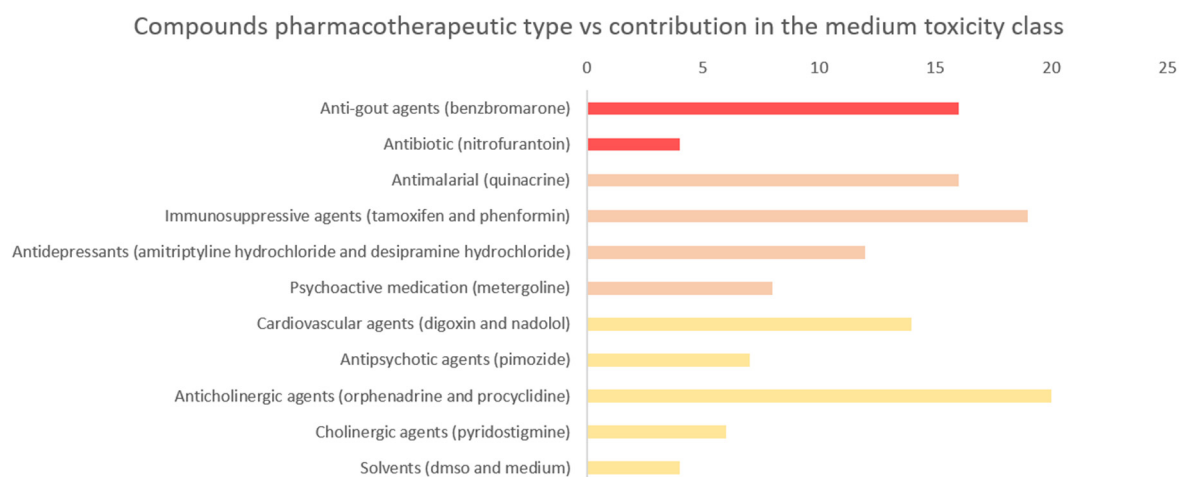


Figure 4: The medium toxicity class spheroids grouped by pharmacotherapeutic types (number of occurrence) and clinically known hepatotoxic effects. Red colour - high hepatotoxic effects, orange colour - moderate hepatotoxic effects and yellow colour - low hepatotoxic effects

Spheroids belonging to the low toxicity class point out most of them with low clinically known hepatotoxic effects, as it is depicted in Figure 5.

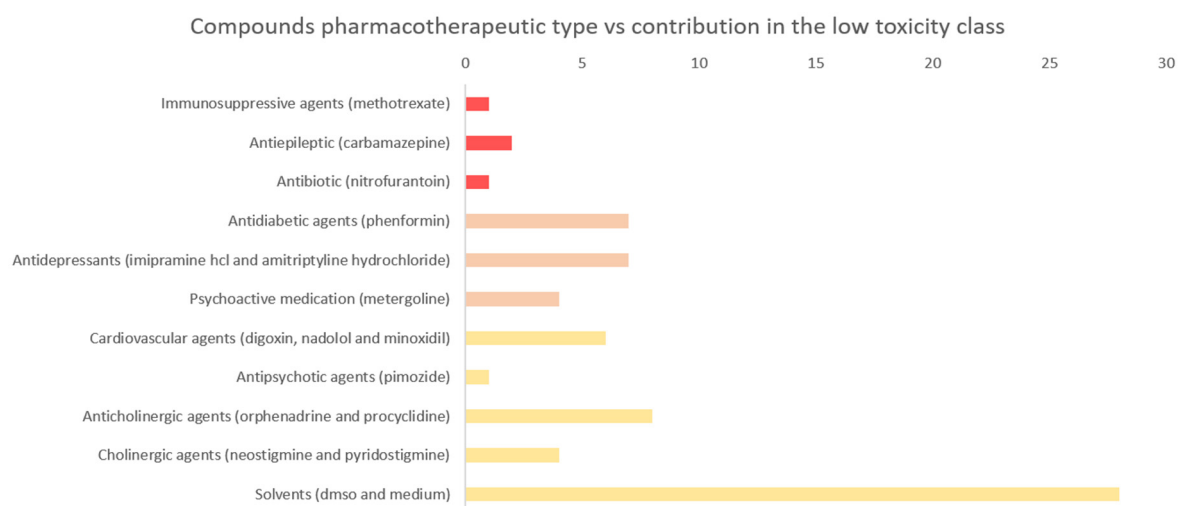


Figure 5: The low toxicity class spheroids grouped by pharmacotherapeutic types (number of occurrence) and clinically known hepatotoxic effects. Red colour - high hepatotoxic effects, orange colour - moderate hepatotoxic effects and yellow colour - low hepatotoxic effects

Finally, in the no toxicity class we observed only spheroids exposed to dmsa and medium, which are known as neutral when it comes to hepatotoxicity. Dmsa can occasionally show low toxicity effects, depending on concentration and time of exposure, but in general can be considered as non-toxic. Medium does not show any toxic effect. However, presence of medium and dmsa spheroids in all other classes in smaller amounts can signal that changes in spheroids do not originate from exposure to compounds, but can be due to other laboratory conditions, poor nutrition to the cells, poor image quality, etc. To gain a more profound conclusion on it, examination of spheroid images relating to the dmsa and medium images belonging to the other toxicity classes needs to be performed.

By analysing obtained results presented in Figures 3-5, it could be concluded that the most critical compounds pharmacotherapeutic type in causing hepatotoxicity are immunosuppressive agents, antiepileptic drugs, anti-gout agents and some antibiotics. Moderate hepatotoxic effects are causing antidiabetics and antidepressants, psychoactive medications and antimalarial drugs. Cardiovascular agents, cholinergic and anticholinergic agents, antipsychotic agents and some antibiotics are not very harmful to the liver. Dmsa and medium are recognised as non-toxic to the liver.

The five most toxic drugs in our dataset, with clinical evidence of it (methotrexate, tamoxifen, nitrofurantoin, carbamazepine and benzbromarone), were analysed in terms of the five most influential physicochemical features. We created a heat map of values normalised to average values of all drugs in our dataset (Figure 6). Orange and red colours signalise values above the total average, while green colours stand for values below the total average.

	MolLogP	MolWt	NumRotBonds	BalabanJ	FractionCSP3
Methotrexate	0.09996795	1.347759652	2.02247191	1.011640138	0.645475574
Tamoxifen	2.233300389	1.101833782	1.797752809	1.205752899	0.595823607
Nitrofurantoin	0.027375724	0.706311387	0.674157303	1.25066186	0.322737787
Carbamazepine	1.261592548	0.700721017	0	1.392603259	0
Benzbromarone	2.032433342	1.257723552	0.674157303	1.333619881	0.303753212

Figure 6: Heat map of five top physicochemical features for five selected drugs normalised to average feature values of all drugs in used dataset

From Figure 6 we can observe that tamoxifen and benzbromarone show high lipophilicity (high MolLogP) which correlates with higher hepatic accumulation. Tamoxifen, benzbromarone and methotrexate are bigger molecules (high MolWt) that link to higher cholestasis risk (tamoxifen and benzbromarone) and fibrosis risk (methotrexate). Tamoxifen and methotrexate are flexible (high NumRotBonds) opposite to carbamazepine which is very rigid. Flexibility increases oxidation and formation of reactive metabolites that contribute to hepatotoxicity (tamoxifen). Increased BalabanJ feature indicated more complex connectivity and all five drugs are above total average value, meaning they have more complex molecular topology linked to diverse metabolic pathways and increased risk of toxicity. Higher values of FractionCSP3 means more saturated compounds, while lower values explain flat, aromatic structure. All five drugs have FractionCSP3 values below the total average, which could relate to mitochondrial stress, reactive metabolites, and cholestasis and therefore, increased hepatotoxicity risk. Overall, tamoxifen [19], benzbromarone [20] and carbamazepine [21] stand for lipophilic, aromatic and bulky molecules that cause metabolic changes in liver spheroids. Nitrofurantoin is not driven by lipophilicity, and even being small molecule and polar, causes chronic autoimmune-like liver injury, driven by nitro reduction and oxidative stress [22]. Methotrexate with high molecular weight and flexibility upon chronic exposure leads to liver fibrosis and steatosis. Methotrexate lacks the structural indicators for acute hepatotoxicity, but its size, polarity, and mechanism create a chronic hepatotoxicity pathway, meaning that bidirectional mechanism of methotrexate depends on dosage [23]. Similarly, due to low lipophilicity, nitrofurantoin can show bidirectional behaviour in hepatotoxicity, but it depends on time of exposure.

Finally, we analysed temporal dependence of drugs to cell viability (toxicity), providing the insights in the dynamic of effects and toxicity progress over time. We analysed only compounds known to be highly toxic to the liver - methotrexate, tamoxifen, nitrofurantoin, carbamazepine and benzbromarone. Since our dataset (used in training and testing of the model) contains only images taken at day 7, we used additional spheroid images exposed to these five selected drugs at other days from Dubinsky's unlabelled set. To analyse toxic effects, we used our trained model to predict the level of ATP for each image. In total 15 new images were taken from an unlabelled dataset - for each drug at days 1, 4 and 7.

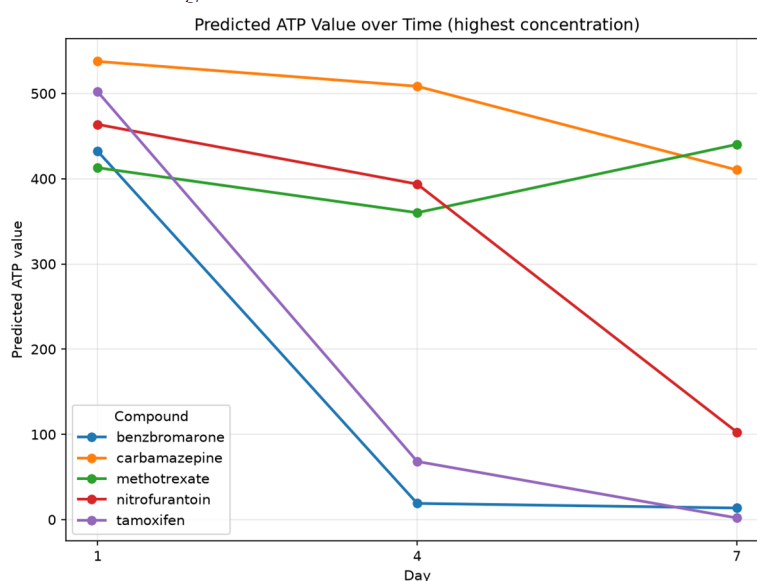


Figure 7: Temporal dependence of cell viability (predicted ATP level) for selected five drugs

The predicted ATP trajectories (Figure 7) revealed distinct toxicity dynamics: benzbromarone and tamoxifen exhibited a rapid decline in viability already by day 4, nitrofurantoin showed a delayed but pronounced decrease between days 4 and 7, whereas carbamazepine displayed a slow, gradual reduction, retaining comparatively high viability throughout. Because the training data contained drug samples only at day 7 (while control samples spanned days 1 to 21), predictions for days 1 and 4 represent an extrapolation along the temporal dimension for drugs. To assess the reliability of these predictions, the model outputs were validated against the reference Dubinsky predictions for the same images, Figure 8, accessed through the image identifiers and rescaled to the ATP value range by linear calibration. The validation yielded an overall R^2 of 0.891 (MAE 51.33), confirming that the model reproduces the reference ATP values for days 1 and 4 with reasonable accuracy despite the absence of drug samples at these days in training of the model. This result supports the interpretation that the predictions are driven primarily by spheroid morphology, which implicitly encodes the temporal state of the spheroid, rather than by the day feature itself, thereby justifying the temporal toxicity analysis.

A particular case worth highlighting is methotrexate, which exhibited anomalous behaviour across all stages of the analysis. Unlike the other four compounds, whose predicted ATP trajectories followed a monotonically decreasing pattern consistent with progressive cell damage, methotrexate displayed a non-monotonic profile, with ATP values slightly decreasing by day 4 and then rising again by day 7, contrary to the expected toxicity progression, as it can be noticed in Figure 7. This compound was also the most consistent outlier throughout the study: it produced the largest single prediction error in the test-set evaluation, and in the temporal validation against the reference Dubinsky predictions it again showed the largest deviations, particularly at day 4 (absolute error of 171), with the model systematically underestimating its ATP values.

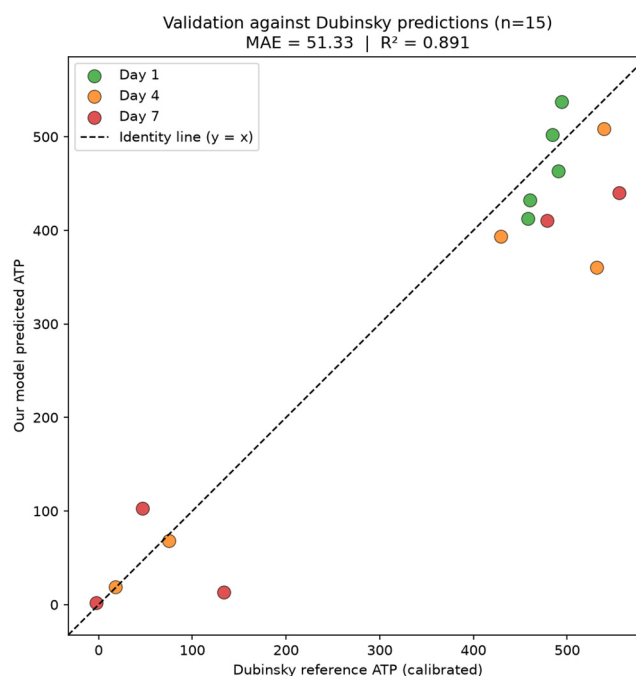


Figure 8: Validation of our model outputs against referenced Dubinsky's model for predicted ATP levels for 15 new samples of five selected drugs at days 1, 4 and 7, including MAE and R^2

Methotrexate is an antimetabolite (chemotherapy / immunosuppressant) whose cytotoxic mechanism acts primarily through inhibition of folate metabolism and DNA synthesis, an intracellular biochemical mode of action that may not be fully reflected in the morphological appearance of the spheroid [24]. This offers a plausible explanation for the model's reduced accuracy: since the predictions are driven predominantly by morphological latent features, a compound whose toxicity is not accompanied by a proportional change in spheroid morphology is inherently more difficult to characterise. Rather than representing a model failure, this discrepancy illustrates one of the stated objectives of the framework, the ability to detect compounds that behave differently than expected, thereby flagging potential errors or revealing new phenomenological causes that merit dedicated experimental investigation.

4. CONCLUSION

Our research showed that even though a lot of information can be extracted from liver spheroid images, it is important to take into account physicochemical features of applied drugs, since it would bring a bit better clustering of data, but more importantly can provide insight into reasoning why some changes in liver spheroids happens, instead of pointing out they happened. Interpretation of obtained results can give better insight into underlying mechanisms leading to hepatotoxic changes. We analysed pharmacotherapeutic types of drugs from used dataset and observed that immunosuppressive agents, antiepileptic drugs, anti-gout agents and some antibiotics showed highest toxicity risk. Antidiabetics and antidepressants indicated moderate toxicity risk, while cardiovascular agents, cholinergic and anticholinergic agents, antipsychotic agents and some antibiotics revealed not very harmful effects to the liver. We then analysed values of the five most

influential physicochemical features for five most toxic drugs in our dataset (methotrexate, tamoxifen, nitrofurantoin, carbamazepine and benzbromarone) to show alignment with observed experimental and clinical data and to look for possible deviations from expected results. Our analysis explained potential bidirectional behaviour of methotrexate and nitrofurantoin in relation to hepatotoxic effects. Temporal analysis of five most toxic drugs from new external test on the trained model explained changes in liver spheroids and more profound hepatotoxicity effects with time but also captured bidirectional behaviour of methotrexate. Future research will include testing of additional drugs and analysis of their effects, as well as extending of developed ML framework to analyse liver spheroids co-exposures. In that way, we would be able to analyse combined effects of applied drugs on liver spheroids in terms of potential hepatotoxicity risk. Additional analysis will be dedicated to the influence of time and concentration of applied drugs for co-exposure scenarios.

ACKNOWLEDGEMENTS

This research was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, contract numbers 451-03-34/2026-03/200108 (Faculty of Mechanical and Civil Engineering in Kraljevo, University of Kragujevac), 451-03-33/202603/200107 (Faculty of Engineering, University of Kragujevac) and 45103-33/2026-03/200378 (Institute for Information Technologies Kragujevac, University of Kragujevac) This research is in line with the strategic plan of the Republic of Serbia 2030, particularly in the section 2.2.2. Development of human resources, goal 3: Good health.

REFERENCES

1. D. Dubinsky, S. Harel, A. Bein, A. Nyska, S. Ya'ari, F. Anas, I. Bentwich and L. Wolf, "Non-invasive quantification of viability in liver spheroids using deep learning", *Frontiers in Bioengineering and Biotechnology*, Vol. 14, pp. 1-18, <https://doi.org/10.3389/fbioe.2026.1797474>, (2026)
2. D. Dubinsky, S. Harel and QurisAI, "3D Liver Spheroids Brightfield 256x256 Images + ATP labels [Data set]. QurisAI", <https://doi.org/10.5281/zenodo.15120118>, (2025)
3. R. J. Andrade, N. Chalasani, E. S. Björnsson, A. Suzuki, G. A. Kullak-Ublick, P. B. Watkins, H. Devarbhavi, M. Merz, M. I. Lucena, N. Kaplowitz and G.P. Aithal, "Drug-induced liver injury", *Nature Reviews Disease Primers*, Vol. 5(1), <https://doi.org/10.1038/s41572-019-0105-0>, (2019)
4. LiverTox®: Clinical and Research Information on Drug-Induced Liver Injury [Internet]. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases, <https://www.ncbi.nlm.nih.gov/books/NBK547852/>, (2012)
5. V. Y. Soldatow, E. L. Lecluyse, L. G. Griffith and I. Rusyn, "In vitro models for liver toxicity testing", *Toxicology Research (Cambridge)*, Vol. 2(1), <https://doi.org/10.1039/C2TX20051A>, (2013)
6. C. C. Bell, V. M. Lauschke, S. U. Vorrink, H. Palmgren, R. Duffin, T. B. Andersson and M. Ingelman-Sundberg, "Transcriptional, Functional, and Mechanistic Comparisons of Stem Cell-Derived Hepatocytes, HepaRG Cells, and Three-Dimensional Human Hepatocyte Spheroids as Predictive In Vitro Systems for Drug-Induced Liver Injury", *Drug Metabolism and Disposition*, Vol. 45(4), <https://doi.org/10.1124/dmd.116.074369>, (2017)
7. Z. Benzait, Ö. Tomsuk, A. Ebrahimi, H. Ghorbanpoor, C. Özel, R. Didarian, B. Demir Cevizlidere, M. Kaya, T. Gur, N. Gasimzade, F. B. Ergen, A. E. Topal, S. Dabagh, R. Javanifar, N. Abdullayeva, O. Uysal, A. Eker Sariboyacı, F. Doğan Güzel, S. Hassan and H. Avci, "Liver-on-a-Chip (LoC) Models: Case Studies of Academic Platforms and Commercial Products", *Molecular Pharmacology*, Vol. 23(3), pp. 1367-1401, <https://doi.org/10.1021/acs.molpharmaceut.5c01122>, (2026)
8. P. A. Walker, S. Ryder, A. Lavado, C. Dilworth, C. and R. J. Riley, "The evolution of strategies to minimise the risk of human drug-induced liver injury (DILI) in drug discovery and development", *Archives of Toxicology*, Vol. 94(8), pp. 2559-2585, <https://doi.org/10.1007/s00204-020-02763-w>, (2020)
9. M. Jordović-Pavlović, M. Stanković, M. Popović, Ž. Čojbašić, S. Galović, D. Markushev, "The application of artificial neural networks in solid-state photoacoustics for the recognition of microphone response effects in the frequency domain", *Journal of Computational Electronics*, Vol. 19(3), pp.1268-1280, 2020, <https://doi.org/10.1007/s10825-020-01507-4>
10. V. Milićević, L. Kocova Lazarova, M. Jordović Pavlović, "The application of artificial intelligence in education—the current state and trends", *International Journal of Cognitive Research in Science, Engineering and Education*, Vol. 12(2), pp. 259-272, 2024, <https://orcid.org/0000-0002-5587-2717>
11. T. Šušteršič, V. Gribova, M. Nikolic, P. Lavalle, N. Filipovic, N. E. Vrana, "The Effect of Machine Learning Algorithms on the Prediction of Layer-by-Layer Coating Properties", *ACS Omega*, 8(5), pp. 4677-4686, 2023, <https://doi.org/10.1021/acsomega.2c06471>
12. R. Geci, A. Z. Sayin, S. Schaller and L. Kuepfer, "Integration of in vitro and in silico approaches enables prediction of drug-induced liver injury", *Archives of Toxicology*, Vol. 100(5), pp. 2029-2046, <https://doi.org/10.1007/s00204-026-04305-2>, (2026)

13. V. Bergen, K. Kodella, S. Srikrishnan, O. Barrandon, S. Anderson, M. Rogers-Grazado, C. Fowler, H. Beyene, N. Robichaud, T. Fulton, N. Lapchyk, M. Cortes, N. Plugis, M. Goddeeris and M. Zamanighomi, “A large-scale human toxicogenomics resource for drug-induced liver injury prediction”, *Nature Communications*, Vol. 16(1), <https://doi.org/10.1038/s41467-025-65690-3>, (2025)
14. T. Xiao, Y. Liu, K. Hu, K. Guo, M. Zhang, T. Wang, W. Lei, W. Wang, S. Zhou, Y. Hu and R. Su, “Drug-induced liver injury prediction based on graph convolutional networks and toxicogenomics”, *PLOS Computational Biology*, Vol. 21(9), <https://doi.org/10.1371/journal.pcbi.1013423>, (2025)
15. J. Han, W. Zhung, I. Jang, J. Lee, M. J. Kang, T. D. Lee, S.J. Kwack, K. B. Kim, D. Hwang, B. Lee, H. S. Kim, W. Y. Kim and S. Lee, “HepatoToxicity Portal (HTP): an integrated database of drug-induced hepatotoxicity knowledgebase and graph neural network-based prediction model”, *Journal of Cheminformatics*, Vol. 17(1), <https://doi.org/10.1186/s13321-025-00992-8>, (2025)
16. <https://pubchem.ncbi.nlm.nih.gov/>
17. A. Malka-Markovitz, S. Camara Dit Pinto, M. Cherkaoui, S. M. Levine, S. Anandasabapathy, G. K. Sood, S. Dhingra, G. Yujia, J. M. Vierling and N. R. Gallo “Multiscale modeling of drug-induced liver injury from organ to lobule”, *npj Digital Medicine*, Vol. 8, <https://doi.org/10.1038/s41746-025-01736-6>, (2025)
18. <https://www.rdkit.org/>
19. F. Zhao, P. Xie, J. Jiang, L. Zhang, W. An, Y. Zhan, “The effect and mechanism of tamoxifen-induced hepatocyte steatosis in vitro”, *International Journal of Molecular Sciences*, Vol. 15, No. 3, pp. 4019–4030, <https://doi.org/10.3390/ijms15034019>, (2014)
20. M. G. McDonald and A. E. Rettie, „Sequential metabolism and bioactivation of the hepatotoxin benzobromarone: formation of glutathione adducts from a catechol intermediate”, *Chemical Research in Toxicology*, Vol. 20, No. 12, pp. 1833–1842, <https://doi.org/10.1021/tx7001228>, (2007)
21. S. Higuchi, A. Yano, S. Takai, K. Tsuneyama, T. Fukami, M. Nakajima, T. Yokoi, “Metabolic Activation and Inflammation Reactions Involved in Carbamazepine-Induced Liver Injury”, *Toxicological Sciences*, Vol. 130, Issue 1, pp. 4–16, <https://doi.org/10.1093/toxsci/kfs222>, (2012)
22. National Institute of Diabetes and Digestive and Kidney Diseases, „Nitrofurantoin”, in *LiverTox: Clinical and Research Information on Drug-Induced Liver Injury*, Bethesda (MD), Bookshelf ID NBK548318, PMID 31643641, <https://www.ncbi.nlm.nih.gov/books/NBK548318/>, (2025)
23. D. Ezhilarasan, „Hepatotoxic potentials of methotrexate: understanding the possible toxicological molecular mechanisms”, *Toxicology*, Vol. 458, Article 152840, <https://doi.org/10.1016/j.tox.2021.152840>, (2021)
24. P. Hanoodi and M. Mittal, „Methotrexate”, in *StatPearls*, StatPearls Publishing, Treasure Island (FL), Bookshelf ID NBK556114, <https://www.ncbi.nlm.nih.gov/books/NBK556114/>, (2024)