Razvoj i primjena modela za procjenu ekotoksikoloških rizika bioaktivnih kemijskih spojeva

Lovrić, Mario

Sveučilište u Zagrebu
PRIRODOSLOVNO-MATEMATIČKI FAKULTET

Mario Lovrić

RAZVOJ I PRIMJENA MODELA ZA PROCJENU EKOTOKSIKOLOŠKOG RIZIKA BIOAKTIVNIH KEMIJSKIH SPOJEVA

DOKTORSKI RAD

Mentori:
dr. sc. Bono Lučić
prof. dr. sc. Göran Klobučar

Zagreb, 2021.

University of Zagreb
FACULTY OF SCIENCE

Mario Lovrić

DEVELOPMENT AND APPLICATION OF MODELS FOR ECOTOXICOLOGICAL RISK ASSESSMENT OF BIOACTIVE CHEMICAL COMPOUNDS

DOCTORAL DISSERTATION

Supervisors:
dr. sc. Bono Lučić
prof. dr. sc. Göran Klobučar

Zagreb, 2021

Prije svega veliko hvala mojim mentorima dr. sc. Boni Lučiću i prof. dr. sc. Göranu Klobučaru, koji su omogućili nastanak ovog rada i uložili sve svoje znanje u njega, dali (bes)konačni niz korisnih savjeta, kao i znatno doprinosili kvaliteti publiciranih radova koji su nastali. Zahvaljujem i institucijama koje podržavaju njihov neumorni rad, Institutu Ruđer Bošković (IRB) i Biološkom odsjeku Prirodoslovno-matematičkog fakulteta (PMF) Sveučilišta u Zagrebu.

Ovaj rad ne bi bio moguć bez bogatstva podataka koji su nam dani od strane Hrvatskih voda i tamošnjih suradnika na našim radovima, dr. sc. Draženke Stipaničev i Siniše Repeca, kao ni bez bioloških testova i analiza napravljenih na PMFu i IRBu, gdje bih posebno istaknuo dr. sc. Sanju Babić i dr. sc. Josipa Barišića.

Jedno veliko hvala i dr. sc. Olgi Malev koja je bila pokretač ove suradnje kemičara i biologa i neumorna kolegica u znanstvenom radu.

Sadržaj ove disertacije je niz nevjerojatnih ishoda potaknutih znatiželjom i kreativnih razgovora. Uz moje mentore i sve suradnike, jako je velik broj ljudi čije ideje i misli su na kraju evolulirale u sastavne dijelove ovog rada.

Zato veliko hvala prijateljima dr. sc. Joelu Koenki, koji me je potaknuo da krenem s programerskim radom i Damiru Bučaru koji mi je otkrio svijet strojnog učenja.

Zahvljujem se i Know-Centru, koji je podržao izradu ovog rada, te svojim kolegama dr.techn. Andi Rexhi, dr.techn. Romanu Kernu, Kristini Pavlović, Tomislavu Đuričiću, Emanuelu Laciću, Valnei Sindičić Đuretec, Univ.-Prof. Dr. Stefanie Lindstaedt i mnogim drugima koji su pomagali u rješavanju misterija strojnog učenja i kemoinformatike. Zahvaljujem i dr. sc. Petru Žuveli s kojim sam razvijao algoritme i imao niz znanstvenih razmjena.

Jedna posebna zahvala ide mojoj obitelji i supruzi Hani, koji su neumorna podrška i pokretač moje znatiželje i ona "nevidljiva ruka" kad "programski kod ne radi".

"The first step is to establish that something is possible; then probability will occur."
Elon Musk

Sadržaj

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Kemijski odsjek

Doktorska disertacija

SAŽETAK

RAZVOJ I PRIMJENA MODELA ZA PROCJENU EKOTOKSIKOLOŠKOG RIZIKA BIOAKTIVNIH KEMIJSKIH SPOJEVA

Mario Lovrić

U disertaciji je razvijen novi računalni postupak procjene ekotoksikološkog rizika vodenih staništa uzrokovan bioaktivnim kemijskim onečišćivalima sastavljen od tri cjeline: utvrđivanja stanja, određivanja rizika od postojećim metodama, te razvoja i primjene novih QSAR modela za procjenu rizika. Utvrđeno je da su vodena staništa rijeke Save opterećena kemijskim onečišćivalima od rastućeg značaja za okoliš. S obzirom na nedostatke postojećih QSAR metoda za procjenu rizika razvijenih na malom broju spojeva, novi robusniji QSAR modeli razvijeni su na nekoliko tisuća kemikalija iz baza ToxCast i Tox21 s brojnim i osjetljivijim toksičnim učincima izmjerenim na staničnim linijama i embrijima zebrice. Novi modeli temeljeni su na molekularnim deskriptorima i strukturnim otiscima, a dobiveni su logističkom regresijom, neuronskim mrežama i slučajnim šumama koje su dale ponajbolje modele. Dobiveni modeli pokazuju dobru generalizaciju na vanjskim skupovima onečišćivala, a njihova kvaliteta provjerena je novouvedenim metrikama. Procjene ekotoksikološkog rizika od spojeva utvrđenih u rijeci Savi temeljena na novim modelima toksičnosti pokazuju dobro slaganje s postojećim metodama, uz značajno povećanje kemijskog strukturnog prostora kojeg pokrivaju.

(113 stranica, 43 slika, 34 tablica, 227 literaturnih navoda, jezik izvornika: Hrvatski)

Rad je pohranjen u Središnjoj kemijskoj knjižnici, Horvatovac 102a, Zagreb te Nacionalnoj i sveučilišnoj knjižnici, Hrvatske bratske zajednice 4, Zagreb.

Ključne riječi: ekotoksikologija/ Sava/ QSAR/ strojno učenje/ procjena rizika/ prioritizacija

Mentori: dr. sc. Bono Lučić (v. zn. sur.), prof. dr. sc. Goran Klobučar (red. prof. traj. zv.)

Rad prihvaćen: 2. lipnja 2021.
Ocjenitelji
1. prof. dr. sc. Branimir Bertoša
2. izv. prof. dr. sc. Šime Ukić
3. prof. dr. sc. Bojan Hamer

University of Zagreb
Faculty of Science
Department of Chemistry

Doctoral Thesis

ABSTRACT

DEVELOPMENT AND APPLICATION OF MODELS FOR ECOTOXICOLOGICAL RISK ASSESSMENT OF BIOACTIVE CHEMICAL COMPOUNDS

Mario Lovrić

A novel computational approach for assessing the ecotoxicological risk for water habitats caused by bioactive chemical pollutants is developed in this work. It consists of three parts: the determination of current ectoxicological status, the risk assessment based on existing methods, and the development and application of new QSAR models for risk assessment. Water habitats of the Sava River were found to be burdened with chemical pollutants of growing importance to the environment. Given the shortcomings of existing QSAR risk assessment methods developed on a small number of compounds, new, more robust QSAR models were developed herein on several thousand chemical compounds from the ToxCast and Tox21 databases which were assessed on more sensitive toxic endpoints measured on zebrafish embryos and cell lines. The newly developed models are based on molecular descriptors and structural fingerprints and are obtained by logistic regression, neural networks and random forests which gave the best models. The trained models show good generalization on external sets of chemical compounds. Their quality is validated by newly introduced metrics. Ecotoxicological risk assessment of compounds identified in the Sava River, is based on the new toxicity models which agree well with the existing methods, supported by a significant increase of structural space covered.

(113 pages, 43 figures, 34 tables, 227 references, original in Croatian)

Thesis deposited in Central Chemical Library, Horvatovac 102A, Zagreb, Croatia and National and University Library, Hrvatske bratske zajednice 4, Zagreb, Croatia.

Keywords: ekotoxicology/ Sava/ QSAR/ machine learning/ risk assessment/ prioritization

Supervisor: dr. sc.Bono Lučić (higher res. assoc.), prof. dr. sc. Goran Klobučar (full prof.)

Thesis accepted: 2 June 2021

Reviewers:
prof. dr. sc. Branimir Bertoša
assoc. prof. dr. sc. Šime Ukić
prof. dr. sc. Bojan Hamer

§ 1. UVOD

Na tržištu je EU prisutno više od 100 000 registriranih kemijskih spojeva od kojih je njih 30000 – 70000 u svakodnevnoj upotrebi. [1] EU godišnje proizvodi prosječno 329 milijuna tona kemikalija (2007. - 2016.), od toga je prosječno 124 milijuna tona procijenjeno kao opasno za okoliš (http://ec.europa.eu/eurostat, preuzeto 29. travnja 2021. god.). Velik broj ovih kemijskih spojeva dospijeva u okoliš gdje mogu imati različito djelovanje na sve organizme koji taj okoliš nastanjuju, uključujući i ljude. Kako bi se zaštitilo zdravlje ljudi i okoliš, Europska je komisija donijela uredbu REACH [2] (Uredba o registraciji, evaluaciji, autorizaciji i ograničavanju kemikalija). Ta je uredba stupila na snagu u lipnju 2007., a pridruživanjem EU postala je i regulatorna/pravna stečevina Republike Hrvatske. REACH zahtjeva da se sve tvari proizvedene ili uvezene u Europu, u količini većoj od 1 tone godišnje, prije uporabe registriraju. [2] Podnositelji registracije moraju dostaviti fizikalno-kemijske, toksikološke, ekotoksikološke i okolišne podatke o tvarima, ovisno o proizvedenoj količini. U tu i druge svrhe se procjenjuje mogućnost da tvar prouzroči štetne učinke na zdravlje ljudi i okoliš te njezina granična razina djelovanja. Nadalje, provodi se i procjena njezinih perzistentnih, bioakumulativnih i toksičnih (PBT) te vrlo perzistentnih i vrlo bioakumulativnih (vPvB) svojstava. Ukoliko bi se nastavile koristiti konvencionalne metode ispitivanja opasnosti kemijskih spojeva, zahtjevi za dobivanje informacija naznačenih (definiranih) uredbama REACH i CLP (Uredba o razvrstavanju, označivanju i pakiranju) [3] uključivali bi intenzivno provođenje ispitivanja na pokusnim životinjama. Izračunato je da bi za jednu kemijsku tvar, bez prethodno postojećih podataka i bez pokušaja minimiziranja ispitivanja na životinjama, registracija mogla zahtijevati preko 5000 žrtvovanih životinja kako bi se dobili potrebni podaci. [4] Ipak, unatoč svim naporima testiranja, za 75 % kemikalija na tržištu (20000 do 70000) i za 2000 – 3000 kemikalija visokog proizvodnog volumena na tržištu (preko 1000 tona po godini) [5] nema dovoljno javno dostupnih podataka o toksičnosti i ekotoksičnosti potrebnih za "minimalnu" procjenu rizika u skladu s OECD smjernicama. [4,6,7]

Kopnene su vode najizloženije negativnom utjecaju onečišćenja jer su one krajnji recipijenti otpadnih voda iz kućanstava, poljoprivrednih aktivnosti i industrije. Kopnene vode su stoga onečišćene tisućama spojeva koji predstavljaju nepoznati rizik za te ekosustave i posljedično za ljudsko zdravlje. [1] Nova su istraživanja pokazala da je polovica europskih kopnenih voda nad kojima se provodi monitoring ugrožena onečišćenjem. [8] Postoji općenito suglasje da se

dobro ekološko i kemijsko stanje kopnenih voda, koje je određeno kao konačni cilj Okvirne direktive o vodama, ODV [9], neće postići u predviđenom vremenskom roku. [10] Nedavne studije upozoravaju da je to u velikoj mjeri posljedica načina na koji se trenutačno provodi kemijska i ekološka procjena kvalitete/onečišćenja kopnenih voda, koju zahtijeva ODV. Trenutačni postupci ne pružaju sveobuhvatne odgovore o negativnim učincima smjesa onečišćujućih tvari antropogenog porijekla prisutnih u vodenim ekosustavima. Također ne pružaju zadovoljavajuću informaciju o rizicima koje one predstavljaju za vodene organizme i ljude. Unatoč velikim naporima, biološke metode koje se trenutno koriste za utvrđivanje kvalitete vodenog okoliša daju nepotpune, djelomične, pa čak i kontradiktorne rezultate. [10]

Informacije o toksičnosti spojeva koji dospijevaju u vodeni okoliš, a koje su temelj za donošenje svih procjena njihovog toksikološkog i ekotoksikološkog rizika, rezultat su toksikoloških testiranja pojedinačnih spojeva, koja se provode na nekolicini test-organizama u zadnjih nekoliko desetljeća. Testiranja toksičnosti mogu biti regulatorna i akademska, no zakonska se regulativa u pravilu oslanja samo na regulatorna, koja se provode putem standardiziranih testova toksičnosti i ekotoksičnosti. [11]

Budući da niti kemijske analize niti dosadašnje tradicionalno određivanje ekološkog statusa analizom sastava zajednica vodenih organizama ne može kvalitetno identificirati toksične učinke mješavina kemijskih onečišćivala (KO) prisutnih u vodenom okolišu, kao ni njihov ukupni toksični potencijal, potrebno je rješavanju tog problema pristupiti kombinacijom kemijskog nadzora i djelotvornom analizom bioloških učinaka. [12] Posljednjih je godina posebna pozornost posvećena kemijskim onečišćivalima od rastućeg značaja za okoliš (KORZ) (engl. *contaminants of emerging environmental concern*) čiji se unos u kopnene vode neprestano povećava, dok se istovremeno smanjuje unos tradicionalnih kemijskih onečišćivala (TKO) čija se koncentracija nadzire. [13] Popis tradicionalnih kemijskih onečišćivala sastoji se od 45 unosa, od kojih 42 organskih, a preostale su tri grupe spojeva: olovni, nikleni i živini spojevi. [14] Farmaceutici, droge te metaboliti farmaceutika i droga (FADM) označavaju podkategoriju KO. Sve navedene kategorije KO, TKO, KORZ i FADM prikazane su na slici 1. Definicija KORZ potrebna je kako bi se razlikovao rizik od negativnih učinaka takvih spojeva u okolišu od rizika uzrokovanih TKO (od kojih se redovno prati njih 45). KORZ uključuju farmaceutike i proizvode za osobnu njegu, kao i veterinarske proizvode, pesticide, endokrine modulatore, opijate i industrijske spojeve s još nepoznatim rizicima po okoliš i ljude. [10] Ti su spojevi u pravilu sintetizirani tako da imaju određen način djelovanja na žive organizme (engl. *mode of action;* MoA), a povrh toga mnogi se i sporo metaboliziraju te se stoga vrlo dugo zadržavaju u okolišu. [15]

Slika 1. Kategorizacija kemijskih onečišćivala korištena u ovom radu.

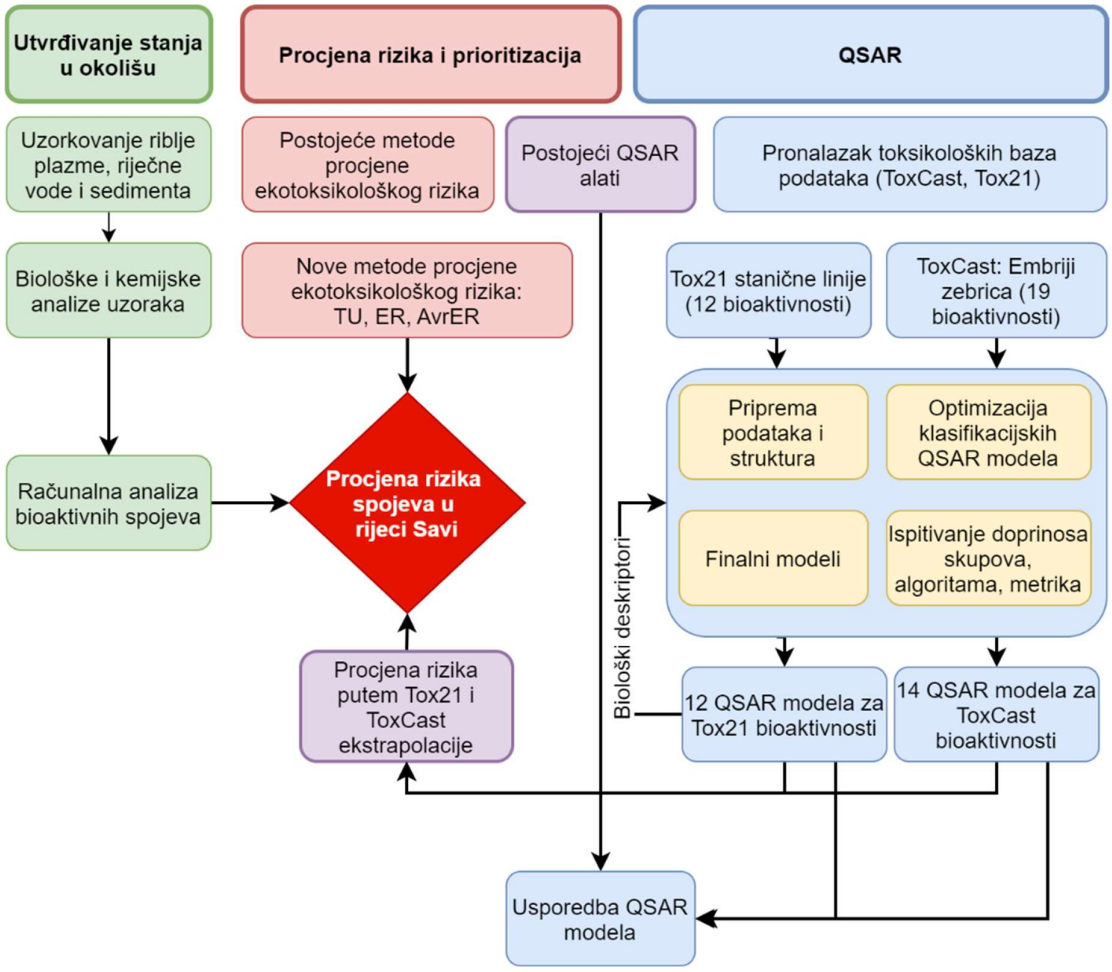
Uzimajući u obzir sve navedeno, vrlo je izvjesno da takve složene smjese tisuća organskih mikroonečišćivala, koje završavaju u europskim rijekama, imaju negativan utjecaj na te ekosustave. [16] Procjena je rizika (engl. *risk assessment*) od onečišćenja kopnenih voda čvrsto povezana s procjenom izloženosti onečišćivalima (KORZ), no sami podaci o koncentracijama nisu do sada bili dovoljni za utvrđivanje načina ili opsega bioloških učinaka onečišćenja. Sveobuhvatna procjena negativnog utjecaja onečišćenja, odnosno prisutnosti i djelovanja smjesa velikog broja KORZ na okoliš, zahtijeva multidisciplinarni pristup i integraciju kemijskih i bioloških *(in vivo, in vitro)* podataka iz online baza podataka ili putem *in silico* analiza. [12] *In vitro* testovi toksičnosti su u posljednjem desetljeću dobili na značaju zbog napretka u razvoju visokoprotočnih testova toksičnosti (engl. *high troughput screening*, HTS).

1.1 Doprinos rada

U ovom su radu analizirani uzorci sedimenta, površinske vode i riblje plazme s područja rijeke Save. Okarakterizirano je i opisano 429 KORZ u rijeci Savi (sediment, voda) te 90 farmakološki aktivnih spojeva nađenih u ribljoj plazmi i vodi. Ekstrakti sedimenta podvrgnuti su toksikološkim ispitivanjima. Na temelju tih podataka predloženo je nekoliko metoda za prioritizaciju KORZ koje pomažu u evaluaciji ekotoksikološkog stanja rijeke Save. U svrhu su prioritizacije korišteni dostupni QSAR (engl. *quantitative structure-activity relationship*) alati kako bi se odredio ekotoksikološki rizik. U drugom su dijelu ovog rada razvijeni novi QSAR modeli koji će poslužiti za razvoj prioritizacijskih strategija kako bi se procijenili učinci koji nisu razvidni iz postojećih alata ili ne pokrivaju dovoljan kemijski prostor. Poznati su modeli razvijani na tradicionalnim toksikološkim bazama koje su koristile tradicionalna kemijska onečišćivala čiji su toksični učinci mjereni pri nestvarno visokim koncentracijama i drastičnim učincima (smrtnost). Stoga njihovi kemijski prostori i domene modela nisu nužno podudarne

sa stvarnim situacijama u okolišu i realističnim koncentracijskim rasponima. Kako bi se prevladali problemi slabih predviđanja i zastupljenosti kemijskog prostora, uvedeni su napredniji pristupi modeliranja kao što su duboke neuronske mreže i algoritam slučajnih šuma, te uravnotežavanje modela. Spomenute metode strojnog učenja (engl. *machine learning*) pokazale su dobra svojstva u višedimenzionalnom deskriptorskom prostoru te i bez izbora varijabli imaju visoku prediktivnu sposobnost (generalizaciju) i razumnu brzinu učenja modela, čak i u primjeni na velikim skupovima. Na slici 2. je vizualno prikazan doprinos ovog rada.

Ovaj rad obuhvaća cjeloviti koncept utvrđivanja stanja i procjene ekotoksikološkog rizika rijeke Save, razvoj QSAR modela sa širokom pokrivenošću kemijskog prostora i korištenje modela za prioritizaciju bioaktivnih spojeva. Iako je istraživanje učinjeno na rijeci Savi, ono je primjenjivo i na druge vodene ekosustave i biotope.



Slika 2. Shematski prikaz postupka cjelovite procjene rizika od onečišćenja prezentiranog u ovom radu.

§ 2. LITERATURNI PREGLED

2.1 Ekotoksikologija i računalna toksikologija

Definiciju je ekotoksikologije dao Truhaut, 1977. godine [17] opisujući je kao granu toksikologije koja se bavi utjecajem kemijskih onečišćivala na sastavne dijelove ekosustava u integriranom kontekstu. U širem se smislu ekotoksikologija bavi (toksičnim i drugim) utjecajima antropogenih kemijskih spojeva na pojedine vrste, zajednice vrsta i populacija, ekosustave i interakcije svega navedenog. [18] "Toksičnost je mjera svih nepoželjnih ili štetnih učinaka kemikalija." [19] Specifična se podvrsta ovih štetnih učinaka opisuje kao toksični učinak (engl. *endpoint*). Toksični učinak je učinak kemijskog spoja na razini stanice ili organizma koji može prekinuti ili promijeniti procese u stanici ili organizmu i tako dovesti do negativnog učinka na zdravlje organizma (definicija preuzeta s poveznice https://pubchem.ncbi.nlm.nih.gov/#query=tox21&tab=substance, 29. travnja 2021. god.). pri određenoj koncentraciji. Toksični učinak u ovom radu odnosi se na kvalitativni učinak kemijskog spoja (aktivno/neaktivno, 1/0) na stanicu ili organizam preuzet iz kasnije opisanih baza podataka kao tzv. "HIT CALL" [20] koji imaju definirane koncentracijske razine u testovima. Za stanice to može biti citotoksičnost, odgovor na stres, aktivnost na nuklearnim receptorima (vidi Tox21 qHTS assays [20]), dok za organizme (embrij) to mogu biti razvojne abnormalnosti, smrt, histopatološke promjene i sl. Toksični učinci mogu biti i kvantitativni (npr. LD50: smrtonosna doza spoja za 50 % jedinki određene vrste u testu toksičnosti u datom vremenskom intervalu, i diskretni (npr. niska, umjerena ili visoka toksičnost). [21] Toksičnim spojem (prema bazi ToxCast) u ovom istraživanju smatra se onaj spoj koji u rasponu koncentracije od 0,0064 µM do 64 µM [22] ima pozitivnu binarnu vrijednost za određeni toksični učinak ili više njih. Testovi toksičnosti nastoje identificirati štetne učinke koje tvari mogu imati na organizmima u onečišćenom okolišu kroz akutno izlaganje (pojedinačna doza) ili višestruko izlaganje (višestruka doza). [11,19,23]

Nekoliko čimbenika određuje toksičnost kemijskih spojeva, poput načina izlaganja ili ulaska u organizam (npr. oralno, dermalno, inhalatorno), doze (količina kemijskog spoja ili smjese), frekvencije izlaganja (npr. pojedinačno ili višestruko izlaganje), trajanja izlaganja (npr. 96 h), ADME svojstava (apsorpcija, distribucija, metabolizam, izlučivanje) [24], bioloških svojstava (npr. dob, spol organizma ili vrsta stanice) i kemijskih svojstava. [19] Animalni se modeli već duže vrijeme koriste u ispitivanju toksičnosti. [25] Kako bi se umanjio broj organizama (*in vivo*)

potrebnih za dobivanje toksikoloških informacija, sve se više nastoje koristiti računalne (*in silico*) i stanične (*in vitro*) metode testiranja toksičnosti. [26] U radu [27] je definirana računalna (*in silico*) toksikologija kao "Svaka radnja s računalom u toksikologiji, a postoji samo nekoliko testova koji ne bi spadali u ovu kategoriju - jer ih većina koristi, barem računalno planiranje i/ili analizu podataka." U užem smislu je to modeliranje toksičnih učinaka iz kemijske strukture. Računalna toksikologija omogućuje procjene toksičnosti uporabom računalnih resursa (tj. metoda, algoritama, softvera, podataka itd.) za organizaciju, analizu, modeliranje, simulacije, vizualizaciju ili predviđanje toksičnosti kemijskih spojeva. [19] Računalne metode imaju za cilj nadopunu *in vitro* i *in vivo* testova toksičnosti kako bi se smanjila potreba za testiranjem toksičnosti na životinjama, smanjio trošak i vrijeme mjerenja/utvrđivanja toksičnosti i poboljšalo njezino predviđanje i procjena sigurnosti uporabe spojeva (tablica 1).

Tablica 1. Usporedba pristupa u procjeni toksikološkog učinka kemijskih spojeva. Preuzeto iz rada. [26]

	***In silico* pristup**	***In vitro* profiliranje**	**Konvencionalno testiranje (*in vivo*)**
Test sustavi	Računalno	Stanični i molekulski testovi	Testiranje na životinjama
Korištenje životinja u testu	Nema	Minimalno	Ekstenzivno
Troškovi testiranja	Mali	Srednji	Veliki
Trajanje testiranja	dani	dani/tjedni	mjeseci/godine
Potrebna količina spoja za testiranje	Nema	μg - mg	g - kg
Kapacitet (brzina)	srednje do visoko	srednje do visoko	nisko
Doziranje	nije primjenjivo	5 do 10 doza	tipično 3 razine

Računalne metode imaju jedinstvenu prednost u tome što mogu procijeniti toksičnost kemijskih spojeva čak i prije nego su sintetizirani. Pojam računalne toksikologije obuhvaća širok izbor računalnih alata kao što su [19]:

- baze podataka za pohranu podataka o kemijskim spojevima, njihovoj toksičnosti i fizikalno-kemijskim svojstvima;
- programi za računanje molekulskih deskriptora i strukturnih otisaka;
- simulacijski alati za biologiju sustava i molekularnu dinamiku;
- metode modeliranja za predikciju toksičnosti;
- alati za modeliranje poput statističkih paketa i softvera za generiranje modela predviđanja;
- ekspertni sustavi koji uključuju unaprijed izgrađene modele na web poslužiteljima ili samostalne aplikacije za predviđanje toksičnosti;
- alati za vizualizaciju.

Ključni je dio računalne/prediktivne toksikologije pronalazak kvantitativnih odnosa strukture i aktivnosti (QSAR - Poglavlje 2.3. ovog rada) koji, uz određenu pouzdanost, na postojećim podacima omogućuju procjenu toksičnosti. Prediktivna toksikologija koja rabi takve metode može pružiti vrijednu pomoć u procjeni negativnog utjecaja onečišćenja na ekosustav odnosno žive organizme koji ga nastanjuju, kao i procjenu prioriteta i negativnog utjecaja relevantnih KORZ koji značajno doprinose toksičnosti otpadnih voda. Kako bi se dobili prediktivni toksikološki modeli, potreban je određeni broj postojećih mjerenja na temelju kojih se modeli optimiraju (uče), odnosno na kojima se uspostavlja kvantitativni odnos između poznate strukture i poznatog toksičnog učinka. S takvim je QSAR modelima moguće, uz određenu pouzdanost, ekstrapolirati uspostavljene odnose (modelima "naučeno znanje") na poznate strukture nepoznatog toksičnog učinka. Međutim, treba imati na umu da modeli ne mogu uvijek zadovoljiti kriterije kvalitete predviđanja podataka o toksičnim učincima, niti nadomjestiti njihovu nedostatnu kvalitetu. [28]

2.2 Procjena ekotoksikološkog rizika onečišćivala prisutnih u kopnenim vodama

Posljednjih je godina postignut veliki napredak u kemijskim analizama onečišćivala od (još uvijek aktualnog) strogo ciljanog pristupa (pretraživanje TKO, vidi Poglavlje 1) do primjene kemijskog probira širokog spektra uporabom visoko osjetljivih analitičkih tehnika koje omogućuju detekciju vrlo niskih koncentracija spojeva u uzorcima uzetima iz okoliša. [29,30] Dostupnost analitičkih metoda kojima je moguće provesti određivanje širokog spektra različitih skupina organskih spojeva prisutnih u niskom (nanogramskom) koncentracijskom području jedan je od osnovnih preduvjeta za kvalitetnu procjenu rizika od onečišćenja kopnenih voda. Skeniranje širokog spektra KORZ potvrdilo je prisutnost farmaceutika, dezinficijensa, droga i

metabolita (FADM) s koncentracijama u rasponu ng/g u uzorcima sedimenta različitih europskih slatkovodnih tijela (npr. jezera, rijeke, močvare, i sl.), kao i onih diljem svijeta. [31–33] Riječni sedimenti predstavljaju velik i važan spremnik za antropogena onečišćivala u vodenom okolišu. [34] Složenost sastava sedimenata zahtijeva razvoj novih standarda kvalitete okoliša i procjene koncentracija koje izazivaju učinke na žive organizme za potencijalnu procjenu rizika kemijskih smjesa [35], a ne samo pojedinačnih spojeva, što je važeća regulatorna praksa u europskim zemljama. [36] Kemijski spojevi iz sedimenta ponovno se otpuštaju u vodeni medij resuspenzijom i trofičkim prijenosom te tako sediment predstavlja dugoročni izvor onečišćenja. [37] Danas je poznat velik broj kemijskih spojeva - njih preko 94 milijuna. [38] Od više desetina tisuća spojeva koji su prisutni na tržištu i koji dospijevaju u okoliš, vrlo je mali broj onih koji su detaljno toksikološki i ekotoksikološki istraženi. Razvidno je da nedostaje velika količina podataka o mogućim toksičnim učincima kemijskih spojeva prisutnih u vodenom okolišu, posebice ne-tradicionalnih onečišćivala (KORZ). Stoga je od velikog značaja procijeniti i karakterizirati ekotoksikološke rizike koje uzrokuju onečišćeni sedimenti. [34] Većina bioloških metoda koje se koriste u procjeni rizika onečišćenja kopnenih voda dugotrajne su i zahtjevne te je stoga od iznimne važnosti uspostavljanje brže i preciznije metodologije procjene negativnog utjecaja onečišćenja na vodeni okoliš. [39,40]

2.2.1 Procjena ekotoksikološkog rizika onečišćenja rijeke Save

Sava je najveća rijeka u jugoistočnoj Europi i najveći pritok Dunava. Njezin sliv pokriva ukupnu površinu od oko 97 700 km^2 na kojoj boravi oko 8 000 000 ljudi. Sava ima 510 km hrvatskog dijela vodotoka. U Savu se ulijevaju komunalne (uglavnom nepročišćene) i industrijske otpadne vode koje, zajedno s vodom koja potječe od ispiranja poljoprivrednih tala, predstavljaju glavne izvore onečišćenja. [41] Do sada su aktivnosti praćenja rijeke Save bile uglavnom ograničene na relativno mali broj onečišćivača, prvenstveno analizom industrijskih i domaćih otpadnih voda. [42–45] Glavni je vodnogospodarski laboratorij (GVL) Hrvatskih voda uspješno implementirao metode masene spektrometrije visoke razlučivosti povezane s tekućinskom kromatografijom (UHPLC-QTOF-MS) za široku spektralnu analizu više stotina KORZ u vodi i sedimentu. [33] To je omogućilo uspješno kvantificiranje KORZ u rijeci Savi tijekom posljednjih nekoliko godina. Hrvatske vode provode redoviti kemijski i biološki nadzor na osamnaest lokacija na rijeci Savi (slika 3). Vrijednosti primijenjenih bioloških indeksa na temelju uzorkovanog makrozoobentosa (makrobeskralježnjaka koji nastanjuju dno vodenih tijela) koje koriste Hrvatske vode ukazale su da 50 % mjernih postaja na rijeci Savi ne zadovoljava kriterije ODV-a. [46] Ekotoksikološki je učinak onečišćenja rijeke Save detaljnije

istraživan u posljednja tri desetljeća u dijelu rijeke nizvodno od Zagreba. U tim su istraživanjima korišteni mnogobrojni biomarkeri na različitim vrstama organizama [47,48] te više biotestova. [49] Nekoliko je studija praćenja prisutnosti kemijskih spojeva usmjereno na opsežnu kemijsku i ekotoksikološku karakterizaciju rijeke Save, gdje je naglašena važnost polarnih KORZ u vodenom okolišu. [44,50] U rijeci je Savi provedena i procjena onečišćenja sedimenta teškim metalima, kao i analiza organskih spojeva [51,52], a toksičnost sedimenta procijenjena je s pomoću nekoliko bioloških testova. [53,54] Imajući u vidu da u rijeku Savu putem komunalnih i industrijskih otpadnih voda i ispiranjem s poljoprivrednih površina svakodnevno ulazi tisuće antropogenih kemijskih spojeva, jasno je da procjena utjecaja onečišćenja na organizme koji nastanjuju taj ekosustav zahtijeva pristup koji će uzeti u obzir (u najmanju ruku) aditivno djelovanje većine ovih spojeva.

Slika 3. Karta s označenim lokacijama postaja na kojima Hrvatske vode provode redovito uzorkovanje i praćenje/nadzor vode u rijeci Savi (preuzeto iz rada [55]).

Razumljivo je da je nemoguće provesti istraživanja utjecaja ovih složenih i promjenjivih mješavina spojeva na sve organizme. Koncentracije kemijskih spojeva u vodi imaju dnevne, mjesečne i sezonske varijacije. Ljudi i ostali organizmi opetovano su izloženi složenim smjesama kemijskih spojeva (engl. *exposome*) čiji se sastav stalno mijenja. Međutim, u velikoj većini strategija procjena rizika uzima se u obzir samo jedan spoj, i nema općenito primjenjivih smjernica kada je i kako potrebno provesti procjenu rizika uzrokovanog kombinacijom (smjesom) spojeva. Tom se problemu može pristupiti korištenjem sve većeg broja izvora

podataka dostupnih online – tj. informacija o kemijsko-molekularnim interakcijama kao vjerodostojnoj osnovi za buduća ekotoksikološka istraživanja. Takav pristup može spojiti podatke dobivene analitičko-kemijskim postupcima s podacima o poznatim biološkim učincima, a sve radi sveobuhvatnije procjene utjecaja smjesa KORZ na onečišćenje voda i ekosustava. [56]

2.2.2 Procjena toksikološkog rizika farmaceutski aktivnih spojeva u ribama

FADM su klasa složenih i multifunkcionalnih kemijskih spojeva s terapijskim svojstvima, koji mahom dolaze iz uporabe u humanoj i veterinarskoj medicini. U vodene ekosustave dospijevaju uglavnom iz komunalnih voda i bolničkog otpada, kao i sa životinjskih farmi. [57] Koncentracije FADM spojeva u vodenom ekosustavu kreću se od vrlo niskih (ng/L) do niskih koncentracija (μg/L do mg/L). [58] Ekološko se stanje u Hrvatskoj, vezano za pojavu FADM, uglavnom procjenjuje u otpadnim vodama. [43,59] Pojedini su podskupovi FADM, kao što su opioidi i antibiotici, potvrđeni u površinskim vodama u više istraživanja. [59,60] U recentnijem se istraživanju [33] pokazalo da riječni sediment može poslužiti i kao spremnik FADM, što ga čini sekundarnim izvorom FADM zbog njihove mogućnosti povratka u vodu. Metaboliti, kao sastavni dio FADM, obuhvaćaju psihotropne kemijske spojeve i njihove metabolite, kao što su kokain, norkokain, kokaetilen, ekgoninski metil ester, kanabis, stimulansi ekstazi i amfetamin. [61]. Zbog njihovog nepotpunog uklanjanja iz otpadnih voda otkriveni su različiti metaboliti u sirovoj ili obrađenoj otpadnoj vodi, površinskoj i vodi iz slavina (ng/L do μg/L [62]). FADM su također otkriveni u otpadnim vodama koje se ispuštaju u rijeku Savu nakon uređaja za pročišćavanje otpadnih voda u Zagrebu (ZOV). [63] Iako je prisutnost FADM u površinskim vodama poznata, nepoznati su rizici za vodene organizme. Razlog tome ograničeno je razumijevanja izloženosti organizama spomenutim podkategorijama kemijskih onečišćivala te nedostatak podataka o njihovom biološkom učinku u koncentracijama značajnim za okoliš. Riblja plazma naznačena je kao posebno poželjno ciljno tkivo za procjenu djelovanja FADM kod riba. [64]

U ovom je radu omjer koncentracije FADM u ribljoj plazmi i humane terapijske koncentracije u plazmi (HTKP) upotrijebljen za predviđanje rizika o farmakološkom učinku FADM na ribe. [65,66] Model se riblje plazme (MRP), koji su predstavili Hugget i sur. [65] temelji na pretpostavkama specifičnosti i konzervacije (očuvanja) ciljanih proteina (engl. *protein target conservation*) između ljudi i riba (tzv. *read-across hypthesis).* [67] MRP je metoda koja se može koristiti i za početnu (*a priori*) procjenu rizika, kada ne postoje alternativni toksikološki podaci za vodene organizme, ali postoje koncentracije djelovanja za ljude. [68] Do danas većina podataka

o FADM u plazmi ribe proizlazi iz izloženosti u kontroliranim laboratorijskim uvjetima [69] ili su usmjereni na manji broj kemikalija poput TKO. [70–73]

2.3 Uvod u QSAR

Početak QSAR modeliranja pripisuje se publikaciji Hanscha i Fujite iz 1962. godine. [74] Cilj je bio povezati odnos aktivnosti regulatora rasta biljaka s njihovom kemijskom strukturom, u vidu koeficijent raspodjele oktanol-voda log*P* (log K_{ow}) koji služi kao procjenitelj lipofilnosti o kojoj ovisi propusnost stanične membrane prema kemijskom spoju. Dvije godine kasnije isti autori koriste računala u svojim analizama [75] i tako otvaraju prostor razvoju modernog oblika ove znanosti, koja je danas posve povezana s računalnim znanostima i matematikom. QSAR je matematički model koji povezuju kvantitativna strukturna svojstva molekula reprezentirana numeričkim varijablama (molekularnim deskriptorima ili strukturnim otiscima) s kemijskom ili biološkom aktivnošću - tj. aktivnost je molekule neka funkcija jednog ili više strukturnih svojstava. Modelirati se može cijeli niz aktivnosti/svojstava poput antiviralne i antitumorske aktivnosti [76] ili fizikalno-kemijskih svojstava molekula poput topljivosti (log*S*) [77] ili log*P*. [78] Potreba za modeliranjem nastaje iz činjenice da za velik broj kemijskih spojeva još nisu izmjerene eksperimentalne vrijednosti njihovih mogućih aktivnosti ili svojstava. Za pronalaženje pouzdane funkcije (QSAR modela) koja strukturne atribute/varijable u fizikalno-kemijsko svojstvo ili toksični učinak, potrebno je imati skup molekula koji već ima eksperimentalno izmjerenu aktivnost/svojstvo (ciljnu varijablu, Y) kako bi se model mogao razviti postupkom optimizacije (učenja, treniranja) iz poznatog skupa strukturnih atributa/varijabli (X). Radi lakšeg razumijevanja i identificiranja pojmova, ovom se radu skup atributa/varijabli X još naziva i skup prediktorskih (ili ponegdje i prediktivnih) varijabli koji se definiraju ponekad i kao kemijska ili molekulska reprezentacija. [79] Kvaliteta modela može provjeriti (ispitati, testirati) u predviđanju na dodatnom (vanjskom) skupu spojeva s izmjerenim aktivnostima/svojstvima. Obavezna je predradnja čišćenje skupa spojeva u smislu pregleda struktura i pretvaranja u potrebne formate prihvatljive računalima. Organizacija za ekonomsku suradnju i razvoj (OECD) definirala je pet načela za razvoj QSAR modela, kako bi se mogli primijeniti u regulatorne svrhe u predviđanju i procjeni toksičnosti: [80]

- definirana toksičnost (toksični učinak);
- jednoznačan algoritam;
- definirana domena primjenjivosti;
- prikladne mjere dobrog uklapanja (engl. *goodness-of-fit*), robusnosti i predviđanja;
- mehanističko tumačenje, ako je moguće.

2.4 Molekulski deskriptori i strukturni otisci

Obvezatna predradnja je priprema skupa spojeva u smislu pregleda struktura i pretvaranja u potrebne oblike zapisa koji su prihvatljivi računalima, kao što su zapisi SMILES, SDF ili MOL. SMILES [81] SMILES (engl. *Simplified Molecular-Input Lne-entry System*) oblik zapisa strukture molekula jednodimenzionalni je ASCII zapis koji prihvaća skoro svaki kemijski softver. MOL oblik tablični je zapis strukture koji sadrži informacije o vrstama atoma, vezama između atoma te njihove koordinate u 2D ili 3D prostoru.

Molekulski deskriptori (**DS**) se računaju iz molekulske strukture koja mora biti jednoznačno definirana. Prema radu [82] molekulski je deskriptor konačni rezultat logičkog i matematičkog postupka koji pretvara strukturne kemijske informacije kodirane simboličkim prikazom molekule u korisni broj. Deskriptori mogu biti jedno- ili više-dimenzionalni. [35] Primjer 1D deskriptora bio bi broj nekih funkcionalnih skupina kao što je npr. skupina −OH, a primjer 2D deskriptora bili bi topološki indeksi izvedeni iz matrica udaljenosti atoma. Još su neki primjeri deskriptora navedeni u tablici 2.

Tablica 2. Primjeri mogućih molekulskih 2D deskriptora podijeljeni u kategorije.

Tip 2D deskriptora	**Primjeri**
Jednostavno prebrojavanje	Broj donora vodikove veze, broj prstenova, molekulska masa
Fizikalno-kemijska svojstva	log*P*, hidrofobnost (aditivna svojstva)
Topološki indeksi	Wiener indeks, Randić indeks
Strukturni otisci	Avalon otisci, Morganovi otisci

Molekulski strukturni otisci korišteni u ovom radu su tzv. otisci proširene konektivnosti (engl. *Extended-connectivity fingerprints*, EFCP) ili „Morganovi strukturni otisci“ [83], za koje će se koristiti kratica FP. Radi se o klasi topoloških strukturnih otisaka koji služe za opisivanje struktura i podstruktura (kemijskih funkcionalnih grupa). U praksi su to binarni vektori koji opisuju molekulske strukture na način da svaki bit u vektoru predstavlja prisutnost/odsutnost određenog atoma u određenom susjedstvu. Ovaj se tip varijable pokazao posebno korisnim u radu s neuronskim mrežama i na složenijim problemima u toksikološkom modeliranju. [84]

Neka svojstva „Morganovih strukturnih otisaka“ su:

- Predstavljaju molekularne strukture s pomoću kružnih susjedstava atoma;

- Mogu se vrlo brzo izračunati;
- Njihove značajke predstavljaju prisutnost/odsutnost određenih podstruktura;
- Nisu definirani *a priori* i mogu sadržavati informaciju o ogromnom broju (novih) različitih strukturnih značajki;
- Dizajnirani su tako da predstavljaju i prisutnost i odsutnost funkcionalnosti, jer su obje ključne za analizu molekularne aktivnosti binarni vektor s vrijednostima 1/0);
- Način se njihovog nastajanja može fleksibilno prilagoditi za stvaranje različitih vrsta kružnih strukturnih otisaka.

2.5 Strojno učenje

Kad je rezultat kvalitativan, problem je klasifikacijski. U klasifikacijskom nadziranom učenju algoritam pokušava naučiti (klasificirati, razvrstati) nove primjere, tj. konstruirati opis svake od klasa koristeći skup već klasificiranih primjera. U jednostavnim klasifikacijskim problemima koristi se primjerice logistička regresija (LogReg). [85] No, u potrazi su za najboljim modelom takve metode inferiorne ako se pokušava modelirati nelinearni problem i kad se koristi veliki broj molekulskih deskriptora. [86] S velikom se količinom podataka povećava mogućnost slučajne korelacije te se prednost daje naprednim metodama strojnog učenja poput umjetnih neuronskih mreža (engl. *neural networks*, NN) [87], strojevima s potpornim vektorima i slučajnim šumama (engl. *random forests*, RF). [88] Za predviđanje biološke aktivnosti kemijskih spojeva (engl. *endpoint*) u ovom radu korištena su tri algoritma strojnog učenja prikladna za problem binarne klasifikacije, a to su RF, NN i LogReg. Ovi klasifikacijski algoritmi mogu rješavati i linearne i nelinearne klasifikacijske probleme, pri čemu se RF i NN smatraju nelinearnim metodama.

2.5.1 Logistička regresija

Logistička regresija je klasifikacijski algoritam [85] koji se pretežito primjenjuje na linearno odjeljive klase (kada je granica između klasa u višedimenzionalnom prostoru linearna) dok je model nelinearan u parametrima (funkcija). Za logističku se regresiju kaže da je probabilistički model, odnosno izlaz je vjerojatnost pripadanja klasi Y = 1 za dani X (prediktorske varijable); Jednadžba 2.1.

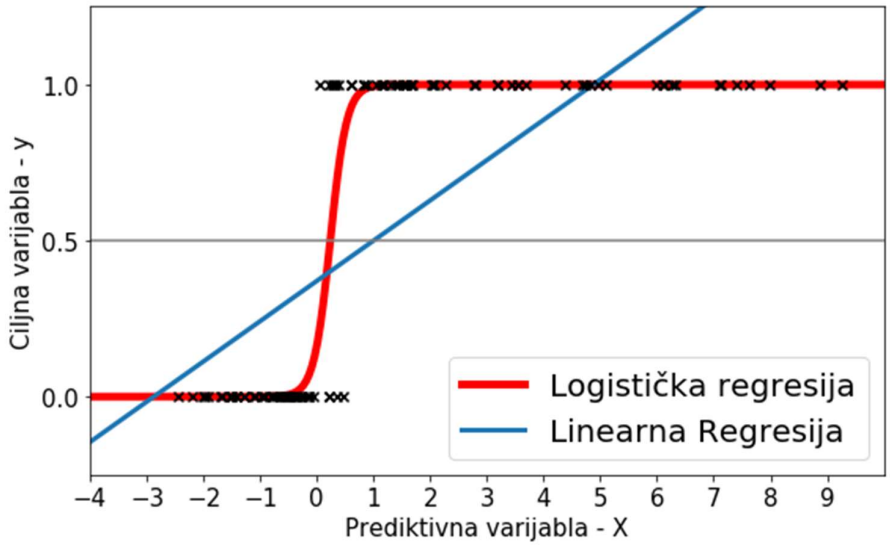
$$p(X) = \Pr(Y = 1|X) \tag{2.1}$$

U slučaju linearne regresije model se može prikazati jednadžbom 2.2, gdje izlaznu vrijednost određuju vrijednosti prediktorske varijable X i njezino uteženje (parametar modela β).

$$p(X) = \beta_0 + \beta_1 X \tag{2.2}$$

Funkcija mora biti takva da p(X) daje vrijednosti 0 ili 1, stoga se koristi logistička funkcija (jednadžba 2.3). Ponašanje funkcije na jednostavnom primjeru za jednu prediktorsku varijablu X prikazano je na slici 4.

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \tag{2.3}$$



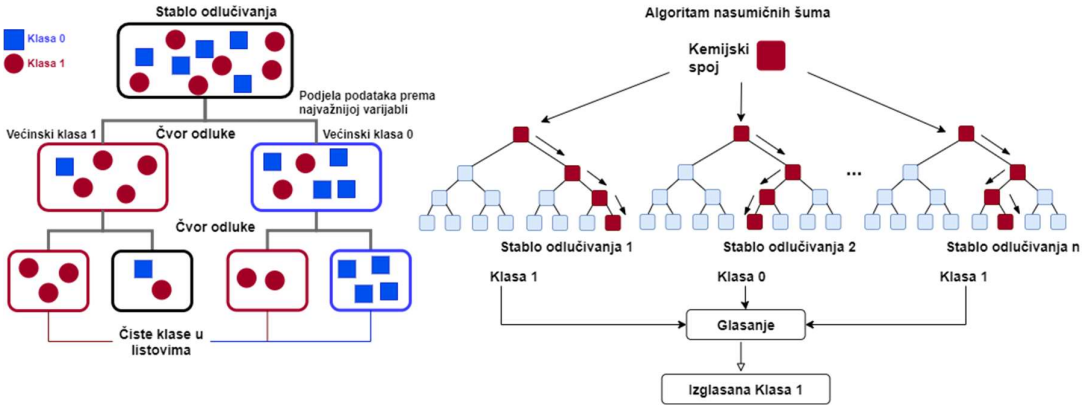
Slika 4. Grafički prikaz primjera binarne klasifikacije varijable Y (vrijednosti 0, 1) na temelju procjene jedne varijable X linearnom regresijom (plava linija) i logističkom regresijom (crvena krivulja).

Dakle, u modelu logističke regresije optimiraju se parametri modela β prema zadanoj funkciji troška (penalizacija pogreške modela). Pregled penalizacijskih funkcija dan je u Poglavlju 2.7.1.

2.5.2 Algoritam nasumičnih šuma (Random Forest)

Algoritam nasumičnih šuma (engl. *Random Forest*, RF) konceptualizirao je L. Breiman. [88] Osnova su algoritma stabla odlučivanja (engl. *Decision trees*) koja predstavlju tzv. „slabe učenike“ (engl *weak learners*). U RF modelu obično se trenira i po više stotina stabala odlučivanja. Izlazi svih stabala agregirani su glasanjem kako bi se dobilo jedno konačno predviđanje; u slučaju binarne klasifikacije to je 1 ili 0. Neovisnost između pojedinačnih stabala (stabla su po algoritmu dekorelirana) smanjuje pristranost u modelima. Kvaliteta se predviđanja na vanjskom skupu (robusnost modela) može kontrolirati pažljivim optimiranjem hiperparametara modela, kao što su dubina stabla, broj stabala, broj uključenih varijabli te

mnogi drugi (detaljan popis parametara dostupan je na poveznici [89]). Dodatna je prednost ovog ansambla način varijacije stabala odlučivanja (tzv. *boostraping aggregation*), odnosno uzorkovanje podskupova unutar skupa za učenje te podskupova prediktorskih varijabli. RF, uz svoju općenito dobru prediktivnost [90,91] prihvaća sve raspodjele numeričkih varijabli te na taj način smanjuje napore za pripremu podataka (primjerice, nije poželjno niti potrebno kategorije varijable pretvarati u binarne jer se algoritam temelji na podjeli varijabli tijekom granjanja odlučivanja). To čini RF pogodnim za uporabu u mnogim aplikacijama, uključujući proizvodnju. Zbog činjenice da se stabla mogu paralelno trenirati, glavna je prednost RF-a njegova paralelizacija - kada se koristi u računalnim infrastrukturama. Shematski prikaz RF-a-a dan je slikom 5.



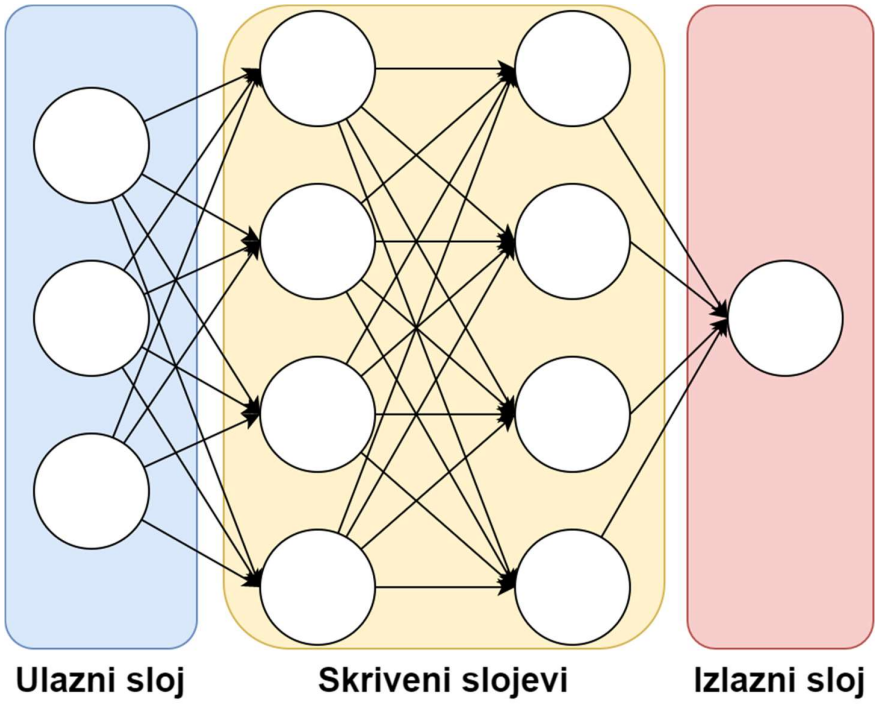
Slika 5. Shematski prikaz klasifikacije putem stabala odlučivanja i algoritma slučajnih šuma .

Slučajna šuma je zajednica (engl. *ensemble*) stabala odluke izgrađena prema sljedećim pravilima: (1) za treniranje svakog stabla korišten je bootstrap uzorak cijelog skupa za učenje koje se koristi, (2) najbolje je razdvajanje u svakom čvoru (engl. *node*) odabrano među definiranim brojevima slučajno odabranih deskriptora ili otisaka, (3) obrezivanje stabala kako bi se izbjeglo prepodešavanje.

Modeli dobiveni metodom RF zadovoljavaju sva prethodno spomenuta OECD načela QSAR modela (Poglavlje 2.3) osim posljednjeg, a to je kriterij mehanističkog modela koji nije neophodan. RF modeli imaju sve uobičajene statističke karakteristike i dodatno imaju vlastiti pouzdani postupak za procjenu robusnosti i predvidljivosti na temelju predviđanja za skupinu spojeva izvan tog područja (engl. *out-of-bag set compounds*).

2.5.3 Neuronske mreže

Neuronske mreže se često koriste u QSAR problemima za rješavanje nelinearne klasifikacije i regresije. [92] U neuronskoj su mreži neuroni organizirani u međusobno povezane slojeve. Tipična se umjetna neuronska mreža sastoji od ulaznog i izlaznog sloja te ponekad skrivenih slojeva (engl. *hidden layers*) između njih (slika 6).



Slika 6. Shematski prikaz rada neuronskih mreža.

Duboke neuronske mreže (engl. *deep neural networks,* DNN) koje sadrže više skrivenih slojeva doživljaju uspon zadnjih godina zbog rastućih količina podataka i potrebe za složenijim predviđanjima na skupovima širokog kemijskog prostora. [84] Neuronske su mreže u ovom radu upotrijebljene za klasifikaciju toksičnosti (kao binarih zapisa). Zbog složenosti postojećeg problema korišten je višeslojni perceptron (engl. *multilayer perceptron*). [87] Nadzirano učenje u NN-u odvija se algoritmom „povratnog rasprostiranja pogreške“ (engl. *backpropagation of error*) postupkom gradijentnog spuštanja. Signal u mreži putuje prema naprijed (engl. *feed forward*) kako bi minimizirao pogrešku ugađanja modela te se nakon računanja pogreška vraća kroz mrežu i korigira uteženja povratnim rasprostiranjem. Iterativnim se procesom uteženja poboljšavaju sve dok se pogreška učenja modela ne smanji. Modeli izgrađeni s pomoću MLP-a prethodno su pokazali kompetitivne rezultate na skupu Tox21. [93]

2.5.4 Bayesova optimizacija

Hiperparametri modela u strojnom učenju imaju snažan učinak na njegovu izvedbu, stoga je nužno prilagoditi hiperparametre modela i naći optimum u parametarskom prostoru s kojima će model dati najbolju generalizaciju. [94] Odabrani hiperparametri trebaju dati najbolje rezultate metričke procjene modela izmjerene na skupu za provjeru valjanosti (engl. *test set*), odnosno minimalnu pogrešku. Ovisno o broju hiperparametara i složenosti modela, ručno isprobavanje različitih hiperparametara može biti izrazito dugotrajno i računalno skupo, budući da se model svaki put iznova trenira i validira s novim skupom hiperparametara. Jedan od automatskih načina ispitivanja prostora hiperparametara je Bayesova optimizacija [95] koja na temelju vrijednosti koje su se u prošlosti (*a priori*) pokazale dobrima (npr. tijekom postupka učenja modela) odabire podprostore parametarskog prostora koji daju manju grešku u križnoj validaciji. Cilj je Bayesove optimizacije postaviti model vjerojatnosti (MV) ciljne funkcije kojim će se odabrati najbolji hiperparametri za zadanu ciljnu funkciju (u ovom slučaju minimizacija pogreške u QSAR modelu). Prvi pronađeni hiperparametri s najboljom izvedbom u MV-u primjenjuju se na QSAR model koji se želi optimirati. Rezultatima iz QSAR modela ažurira se MV te se s ažuriranim hiperparametrima dalje minimizira ciljna funkcija. Iterativni postupak ponavlja se do postignuća optimalnih hiperparametara QSAR modela. Budući da se Bayesovom optimizacijom hiperparametri ne prilagođavaju slučajno već na temelju znanih informacija iz prošlosti (tijekom treniranja), optimalni se hiperparametri pronalaze u manje iteracija. Nadalje, tim se postupkom smanjuje broj pokretanja ciljne funkcije, čime su smanjeni troškovi računanja [96] u odnosu na algoritme poput pretraživanja mreže (engl. *grid-search*). [97] Za optimizaciju hiperparametara Bayesovom optimizacijom primijenjena je Python biblioteka „bayesian-optimization“ u kojoj je implementirana Bayesova optimizacija s Gaussovim procesima koji su korišteni u modelu vjerojatnosti. [98] Uporaba Bayesijanske optimizacije u QSAR klasifikacijskim problemima pokazala je dobre rezultate u modeliranju toksičnosti i mutagenosti na dva velika skupa podataka. [99]

2.6 Mjere kvalitete modela

Za procjenu kvalitete prediktivnih klasifikacijskih modela definirane su mjere koje ocjenjuju koliko je dobro model predvidio ishod. Mjere se izračunavaju iz elemenata konfuzijske matrice (tablica 3). Za potrebe razumijevanja mjera kvalitete dviju klasa definirane su: klasa *P* (pozitivna, 1) i klasa *N* (negativna, 0). U binarnim je klasifikacijskim problemima cilj povećati broj točno procijenjenih spojeva obje klase P i N (TP i TN) u odnosu na pogrešno procijenjene FP i FN.

Tablica 3. Konfuzijska matrica, gdje je TP (engl. *True Positive*, točno pozitivno) ukupan broj pozitivnih točnih predviđanja klase P promatrane/eksperimentalne klase P; FP (engl. *False Positive*, lažno pozitivno) je ukupan broj podcjenjivanja klase N (kada je eksperimentalnu klasa N model predviđa klasu P); FN (engl. *False Negative*, lažno negativno) je ukupan broj slučajeva kada je model/metoda pozitivnu klasu P krivo predvidio kao spoj negativne klase N; TN (engl. *True Negative*, točno negativno) ukupan je zbroj slučajeva kada je eksperimentalna klasa N točno previđena kao klasa N.

	Mjereni	
Predviđeni	Pozitivni (1 ili P)	Negativni (0 ili N)
Pozitivni (1 ili P)	TP, P = P	FP
Negativni (0 ili N)	FN	TN, N = N

Stoga je vrlo intuitivna i uobičajena mjera točnosti Q_2 (engl. *accuracy/quality of 2-class model*) prikazana formulom gdje se računa ukupan broj točno procijenjenih instanci u odnosu na zbroju svih instanci (jednadžba 2.4). Još neke poznate mjere za ocjenjivanje kvalitete klasifikacije su: osjetljivost (engl. *sensitivity*, Sn*)* (jednadžba 2.5) i specifičnost (engl. *specificity*, Sp) (jednadžba 2.6). Osjetljivost je broj istinitih pozitivnih stavki koje su pravilno klasificirane. Može se smatrati mjerom potpunosti (tj. broj spojeva ispravno identificiran kao pozitivan od ukupnog broja stvarnih pozitivnih pacijenata). Specifičnosti predstavljaju negativnu stopu stvarno negativnih. Specifičnost se koristi za bolju procjenu prediktivne sposobnosti mjerenih spojeva negativne klase.

$$Q_2 = \ Accuracy \ = Točnost = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{2.4}$$

$$\mathrm{Sn} \ = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2.5}$$

$$\mathrm{Sp} \ = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{2.6}$$

Sve tri često korištene klasifikacijske mjere imaju jednu zajedničku osobinu, a ta je da koriste isključivo elemente konfuzijske matrice jednog retka ili jednog stupca. Svaka će procjena

metričke vrijednosti koja koristi vrijednosti iz samo jednog stupca ili retka biti neosjetljiva na neuravnoteženu klasifikaciju. Dakle, točnost je relativno neosjetljiva mjera za neuravnotežene klasifikacijske probleme [100,101] Nekoliko metoda procjene kvalitete klasifikacijskih modela osjetljivije su na neuravnotežene podatke kada su spojevi jedne klase u skupu podataka brojniji (većinska klasa *V)*, od uzoraka druge (manjinske) klase *M*. [102] Mjere koje koriste vrijednosti iz oba stupca mogu se koristiti s uravnoteženim i neuravnoteženim podacima. To se može protumačiti kao da mjerni podaci koji koriste vrijednosti iz jednog stupca poništavaju promjene u raspodjeli klase. Međutim, neki mjerni podaci koji koriste vrijednosti iz oba stupca nisu osjetljivi na neuravnotežene podatke budući da se promjene u raspodjeli klase međusobno ponište. Matthewsov je koeficijent korelacije ili MCC (jednadžba 2.7) koeficijent korelacije za binarne varijable [103,104], koji za razliku od točnosti daje jednaku težinu krivo klasificiranim spojevima. Za savršenu je klasifikaciju (FP = FN = 0) vrijednost MCC = 1, što znači savršenu pozitivnu korelaciju. Ako je klasifikator sve spojeve pogrešno klasificirao (TP = TN = 0), dobit će se vrijednost MCC od -1, što predstavlja savršenu negativnu korelaciju. Kao takva, vrijednost MCC se nalazi između -1 i 1, pri čemu 0 označava slučajni rezultat ili korelaciju. MCC se u ovom radu koristi za bolju procjenu ukupne kvalitete modela ako je jedna od klasa manje ili rijetko zastupljena. [105]

$$\mathrm{MCC} = \frac{\mathrm{TP}\cdot\mathrm{TN} - \mathrm{FP}\cdot\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}} \tag{2.7}$$

Bitna je stavka u procjeni kvalitete modela s neuravnoteženom raspodjelom klasa procjena slučajne klasifikacije. Nasumična se ili slučajna točnost ($Q_{2,\mathrm{rnd}}$) može prikazati jednadžbom 2.8. Ova je vrijednost $Q_{2,\mathrm{rnd}}$ uvijek između vrijednosti minimalne točnosti (tj. maksimalnog neslaganja) i maksimalne točnosti. Maksimalni raspon vrijednosti $Q_{2,\mathrm{rnd}}$ vrijednosti je između 0 i 1. I Q_2 i $Q_{2,rnd}$ mogu se iskazati u postocima.

$$Q_{2,\mathrm{rnd}} = \frac{(\mathrm{TP}+\mathrm{FN})(\mathrm{TP}+\mathrm{FP}) + (\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}{(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN})^2} \tag{2.8}$$

Uz to, razlika (u %) između točnosti Q_2 dobivene modelom i odgovarajućom najvjerojatnijom slučajna točnost $Q_{2,\mathrm{rnd}}$ (jednadžba 2.9) može se jednostavno izračunati.

$$\Delta Q_2 = Q_2 - Q_{2,rnd} \tag{2.9}$$

Vrijednost ΔQ_2 može biti najviše max = 0,5 u slučajevima kad je:

(1) potpuno jednak broj elemenata u obje klase (50 : 50 %) u skupu podataka i

(2) savršena procjena ili predviđanje modela (FN = FP = 0).

Stoga se ΔQ_2 može smatrati mjerom vrijednosti doprinosa modela stvarnoj točnosti procjene. U analizi međusobne kvalitete različite klasifikacije modela, parametar ΔQ_2 može poslužiti za rangiranje modela. Veća vrijednost parametra ΔQ_2znači da model doprinosi većoj količini korisnih informacija. Jedna od često korištenih mjera kvalitete klasifikacijskog modela je Cohenova Kappa. [106] Cohenova Kappa (skraćeno Kappa) definirana je kao razlika točnosti i nasumične točnosti normalizirana s razlikom maksimalne točnosti (1) i nasumične točnosti. Kappa je mjera na skali 0 - 1, gdje je vrijednost 0 nausmični model, a vrijednost 1 model koji je točno procijenio sve instance. Kappa je opisana jednadžbom 2.10.

$$\text{Kappa} = \frac{Q_2 - Q_{2,\text{rnd}}}{1 - Q_{2,\text{rnd}}} \tag{2.10}$$

Važna mjera u QSAR klasifikacijskim modelima je uravnotežena točnost (engl. *Balanced Accuracy*, BA), posebice u neuravnoteženim problemima i konzensus modelima. [107,108] BA je geometrijska sredina specifičnosti i osjetljivosti [109], jednadžba 2.11.

$$\text{BA} = \frac{\text{Sn} + \text{Sp}}{2} \tag{2.11}$$

Na temelju uravnotežene točnosti, specifičnosti i osjetljivosti modela računa se kvaliteta podešavanja modela (*Godness of fit,* GOF) [107] na skupu za učenje ili kvaliteta predviđanja modela ako se ista jednadžba (jednadžba 2.12) primjeni na skupu za vrednovanje modela.

$$\text{GOF} = 0{,}7 \cdot \text{BA} + 0{,}3 \cdot (1 - |\text{Sn} - \text{Sp}|) \tag{2.12}$$

Robusnost je modela mjera koja predstavlja generalizaciju modela, jer uspoređuje GOF na skupu za učenje i skupu za vrednovanje [107], jednadžba 2.13.

$$\text{ROB} = 1 - |\text{BA}_{\text{Tr}} - \text{BA}_{\text{Te}}| \tag{2.13}$$

Za sveobuhvatnu je procjenu kvalitete i generalizacije modela predložena sljedeća metoda bodovanja (S) prema sveobuhvatnoj istrazi na konzensus modelima [107], jednadžba 2.14.

$$S = 0{,}3 \cdot GOF_{Tr} + 0{,}45 \cdot GOF_{Te} + 0{,}25 \cdot ROB \quad (2.14)$$

Na osnovi jednadžbi 2.12 i 2.13 predlažu se strože mjere za procjenu robusnosti modela ROBM i kvalitete (Scoring MCC ili SM) na osnovi MCC, a opisane su u jednadžbama 2.15 i 2.16:

$$ROBM = 1 - |MCC_{CV} - MCC_{Te}| \quad (2.15)$$

gdje su MCC_{cv} Matthewsov koeficijent korelacije koji se računa tijekom križne validacija i MCC_{Te} Matthewsov koeficijent korelacije koji se računa na skupu za vrednovanje (test set).

$$SM = 0{,}2 \cdot MCC_{CV} + 0{,}5 \cdot MCC_{Te} + 0{,}3 \cdot ROBM \quad (2.16)$$

Mjera za procjenu robusnosti modela ROBM i SM empirijske su mjere po uzoru na novija istraživanja [90,107] koja će se koristiti kao cjelovita mjere za odlučivanje o najboljim modelima za pojedinačni toksični učinak.

2.7 Klasifikacija neuravnoteženih skupova

Većina problema s klasifikacijom u stvarnosti pokazuje određenu razinu neravnoteže klase, odnosno svaka klasa ne čini jednak dio skupa podataka. [105] Optimalni način rješavanja neuravnoteženih skupova za klasifikaciju (NSK u nastavku) bio bi skupljanje dodatnih podataka, međutim u praksi nije uvijek lako prikupiti više podataka. Pretpostavka klasifikacije na NSK jest da je cilj točna identifikacija manjinske klase, jer u suprotnom ove tehnike zapravo nisu potrebne. Još veći izazov u klasifikaciji su tzv. rijetki slučajevi ili rijetke klase, gdje konvencije strojnog učenja ulaze u sasvim nova rješenja ne bi li se unaprijedila kvaliteta predviđanja. [105] Između ostalog je važno pravilno prilagoditi klasifikacijske metrike i metode strojnog učenja koje se bolje ophode s problemima rijetkih klasa. Popularan pristup za NSK opisan u ovom radu je penalizirano učenje.

2.7.1 Penalizirano učenje

Dok metode naduzorkovanja i poduzorkovanja pokušavaju uravnotežiti raspodjelu po klasama uzimajući u obzir proporcije instanci (elemenata) po klasama, penaliziranje učenja uzima u obzir troškove povezane s pogrešnim klasificiranjem. [102,110] Nedavna istraživanja upućuju na

postojanje veze između penaliziranog učenja i učenja iz neuravnoteženih podataka; dakle, teorijski se temelji i algoritmi sustava penaliziranja mogu prirodno primijeniti na neuravnotežene probleme učenja. [105,111] Temelj je penaliziranog učenja koncept matrice troškova (tablica 4). U redovnom učenju modela sve pogreške klasifikacije tretiramo podjednako, tj. s jednakim uteženjem/troškom (C = 1). Takav pristup uzrokuje probleme kod neuravnoteženih klasifikacijskih problema, jer ne postoji dodatna nagrada za identificiranje manjinske klase u odnosu na većinsku klasu. [102] Manjinskom se klasom u ovom radu smatraju toksični spojevi kojih je u visokoprotočnim testiranju uglavnom manji broj stoga što se prati široki spektar spojeva prisutnih u okolišu, od koji su većina farmaceutici. [33,66] Penalizirano učenje to mijenja i koristi funkciju C (V,M) (obično predstavljenu kao matrica) koja određuje trošak pogrešnog klasificiranja spojeva klase P i klase N. To omogućuje kažnjavanje/penalizaciju pogrešne klasifikacije manjinske klase mnogo više nego s pogrešnu klasifikaciju većinske klase, u nadi da će to povećati točnu procjenu mjerenih vrijednosti što je u konfuzijskoj matrici predstavljeno elementima TN i TP.

Tablica 4. Konfuzijska matrica troškova gdje je C_{TP} trošak/uteženje točnog predviđanja klase P; C_{FP} je trošak pogrešnog predviđanja klase P, C_{FN} je trošak pogrešnog predviđanja klase N, C_{TN} je trošak točnog predviđanja klase N.

	Mjereni	
Predviđeni	Pozitivni (1 ili P)	Negativni (0 ili N)
Pozitivni (1 ili P)	C_{TP}	C_{FP}
Negativni (0 ili N)	C_{FN}	C_{TN}

Ideja je da trošak bude jednak obrnutoj proporciji skupa podataka koji klasa čini. To povećava penalizaciju, jer se veličina klase smanjuje za manjinsku klasu N. Ova se implementacija može podesiti u ciljnoj funkciji/ funkciji gubitka klasifikatora, što je korišteno u ovom radu. U scenariju se binarne klasifikacije definira dakle C(N → P) kao trošak pogrešnog klasificiranja manjinske klase i obrnuto, C(P → N) predstavlja trošak suprotnog slučaja. U pravilu nema troškova za ispravno razvrstavanje klase i troškovi pogrešne klasifikacije manjinskih primjera su veći od obrnutog, tj. C(N → P) >>C(P → N). [102] U ovom radu korištena je penalizacija u obliku metrike kojom se optimira model u postupku razvoja - a to je bio MCC. Standardna postavka u algoritmima je točnost, što znači da se penalizacija provodi neravnomjerno između

elemenata konfuzijske matrice pa je trošak pogrešne predikcije u vidu C(N → P) i C(P → N) relativno nizak. Koristeći MCC algoritam provodi penalizaciju tako da uzima u obzir trošak krivo procjenjenih predviđanja FN i FP.

2.8 *Post-hoc* interpretacija QSAR modela

Peti princip smjernica OECD-a traži "mehanističko tumačenje/interpretaciju QSAR modela, ukoliko je moguće". [112–114] Međutim, ovo načelo je opcionalno. Mehanistički pristup u modeliranju QSAR zahtijeva da su modeli izgrađeni koristeći samo interpretabilne deskriptore koji imaju jasno strukturno ili fizičko-kemijsko značenje. [115] Dodatak ili korištenje deskriptora koji se ne mogu ili teško interpretiraju može poboljšati predviđanje modela. [116] Interpretacija se kod takvih jednostavnih algoritama, poput linearne i logističke regresije. određuje putem regresijskih koeficijenata. U nastojanju postizanja više točnosti modela u predviđanju aktivnosti ili svojstava, modeli se značajno usložnjavaju u pogledu broja deskriptora u modelima, a i u pogledu njihove finkcionalne (matematičke, logičke) strukture. [107,108] Moderni pristupi strojnog učenja koji se uspješno primjenjuju za izradu visoko prediktivnih modela, poput metoda potpornih vektora (engl. *Support Vector Machine*, SVM), neuronske mreže i algoritma slučajnih šuma nisu jednostavni za tumačenje. [115,117] Kod složenijih modela ne može se podastrijeti njihovo mehanističko objašnjenje kako se to može učiniti s modelima temeljnim na linearnoj ili logističkoj regresiji u kojima je funkcionalna ovisnost jednostavna i pregledan. Ako model nije prediktivan, to znači da ne može obuhvatiti međuovisnost između strukture i svojstva skupa molekula, a razlozi za to mogu biti različiti. [118] Ako su slabi modeli kombinirani u gradnji konsenzus modela, potonji se može tumačiti jer će proći statistički korak validacije. Tumačenje može biti korisno u dijagnosticiranju problema s modeliranim skupom podataka koji vodi lošim modelima. To je stvorilo potrebu za kompromisom između predviđanja i interpretabilnosti QSAR modela. [119–121] Budući je kvaliteta prediktivnih modela primarni cilj QSAR-a, interpretacija modela smatra se nekad sporednom i često je zanemarena, a mnogi radovi ne uključuju interpretaciju. Nedavni napredak u tumačenju modela učinili su predvidljivost i interpretabilnost jednostavnima u složenim metodama strojnog učenja. [120,122,123] Korišteni modeli u ovom radu osim logističke regresije, smatraju se crnim kutijama (engl. *black box*) budući da njihova interpretacija nije jednostavna za razliku od linearnog modela. [115] Kod takvih složeniji modela teži se ka *post-hoc* interpretaciji. [124] Jedan široko prihvaćeni takav postupak je permutacijska važnost varijabli.

2.8.1 Permutacijska važnost varijabli

Permutacijska važnost varijabli (PVV) [122,125,126] je *post-hoc* metoda koja mjeri važnost varijable izračunavanjem povećanja pogreške predviđanja modela nakon permutacija varijable. Varijabla je "važna" ako nasumična izmjena njezinih vrijednosti (permutacija) značajnije povećava pogrešku modela. Varijabla je "nevažna" ako izmjena njezinih vrijednosti pogrešku modela ostavlja nepromijenjenom ili ju minimalno promijeni. Isti postupak može se ponoviti više puta, ne bi li se dobile srednje vrijednosti i interval pouzdanosti promjene pogreške modela usljed izmjene (permutacije) vrijednosti pojedinog deskriptora. Metoda je dobila na popularnosti zbog svoje univerzalne primjenjivosti u mnogim složenim modela i jednostavnog pristupa. [127] U QSAR-u procjena važnosti varijable daje informacije o relativnoj važnosti deskriptora korištenih za izradu modela, ali ne daje informaciju o smjeru njihovog utjecaja (pozitivan ili negativan) kao što je slučaj s regresijskim koeficijentima. U ovom radu PVV će se koristiti za interpretaciju modela i selekciju varijabli.

2.9 Kemijski prostor i UMAP

UMAP (engl. *Uniform Manifold Approximation and Projection*) [128] algoritam je koji se koristi za smanjenje broja dimenzija višedimenzionalnih skupova podataka. Algoritam je zasnovan na višestrukim metodama učenja i idejama iz topološke analize. Jedna od pretpostavki za korištenje metode UMAP za smanjenje broja dimenzija je da se manje informacija izgubi u odnosu na linearne metode poput PCA koje se pokazuju suboptimalnima za rad s, primjerice, binarnim podacima (kao što je to slučaj s molekulskim otiscima korištenim u ovoj disertaciji). UMAP se koristi za vizualizaciju velikog broja kemijskih spojeva u manjem broju dimenzija. [129] U ovom radu UMAP će se koristiti za računanje novog oblika kemijske reprezentacije smanjenih dimenzija na osnovi vektorske reprezentacije duljine 5120 molekulskih otisaka za pojedini spoj. U novom prostoru je svaki kemijski spoj označen s tri vektora (3D prostor), gdje je očuvana informacija o strukturnom susjedstvu (strukturnoj sličnosti) spojeva.

§ 3. MATERIJALI I METODE

3.1 Uzorkovanje sedimenta, riječne vode i riblje plazme

3.1.1 Uzorkovanje riječnog sedimenta i vode

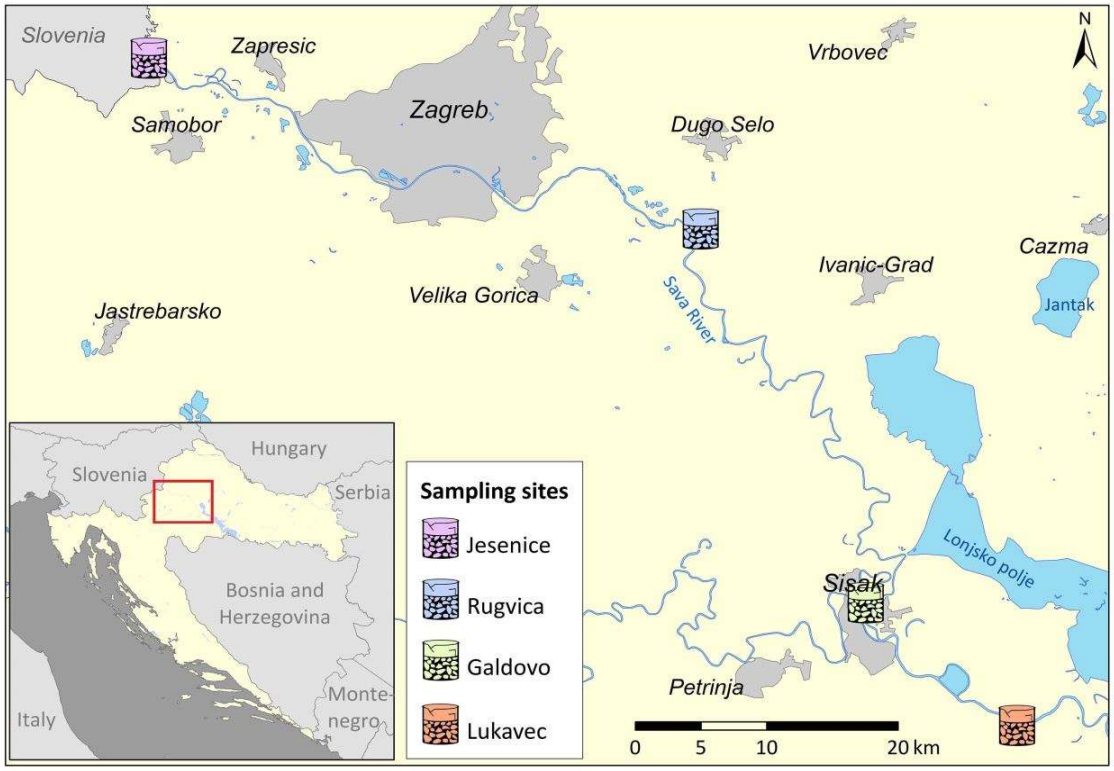
Uzorci sedimenta u ovom radu dolaze iz četiri segmenta rijeke Save i opisani su u tablici 5.

Tablica 5. Opis četiri nalazišta riječnog sedimenta u rijeci Savi.

Lokacija	Opis lokacije	Status onečišćenja
Jesenice	13 km uzvodno od grada Zagreba, u blizini slovensko-hrvatske granice	umjereno onečišćenje
Rugvica	~ 10 km nizvodno od izlaza odvoda otpadnih voda iz postrojenja za pročišćavanje otpadnih voda	umjereno do visoko onečišćenje; prima komunalnu i industrijsku otpadnu vodu iz šireg područja grada Zagreba (~1 milijun stanovnika)
Galdovo	2,4 km uzvodno od ušća rijeke Save i Kupe na ulazu rijeke Save u grad Sisak	nisko ili umjereno onečišćenje
Lukavec	10 km nizvodno od grada Siska (50.000 stanovnika)	prima industrijske otpadne vode iz proizvodnje pesticida, objekata, željeza, rafinerije nafte i urbanog otjecanja

Lokacije su vidljive na slici 7. Uzorke sedimenta rijeke prikupili su djelatnici Hrvatskih voda 2014. godine. Na svakom je mjestu prikupljen 1 kg površinskog sedimenta (gornji sloj 5 – 10 cm) u posudama od smeđeg stakla prikladnima za uzorkovanje i skladištenje krutih uzoraka, koje su zatim zamrznute pri -20 °C prije ekstrakcije. Prije uzimanja uzoraka za analizu uzorci su odmrznuti i temeljito izmiješani. Zatim je odvagnuto približno 5 g uzoraka mokrog sedimenta u posude za ekstrakciju. Suha je masa svakog uzorka sedimenta određena sušenjem u pećnici na 100 °C u trajanju od oko 4 h (tj. do postizanja konstantne mase), a sve vrijednosti

koncentracije prikazane su u odnosu na suhu masu. [130] Alikvot sedimenta od 5 g ekstrahiran je nakon mehaničkog miješanja s organskim otapalima. Postupak stupnjevite ekstrakcije pojačan je potresanjem pri 30 °C, 200 rpm tijekom 30 min. Za ekstrakciju su korištena otapala mješljiva s vodom (metanol-aceton, 1:1, v/v). [131] Nakon ekstrakcije su matrica uzorka i ekstrakt odvojeni centrifugiranjem. Naknadno je čišćenje ekstrakata izvršeno preko stupca silikagela adsorpcijskom kromatografijom. Dobiveni ekstrakt pročišćen je i uparen do suha u blagoj struji dušika. Ekstrakti su ponovno otopljeni u 100 mL HPLC ultra čiste vode. [131] Na nalazištima su također obavljene analize masenog udjela organskog ugljika u vodi.

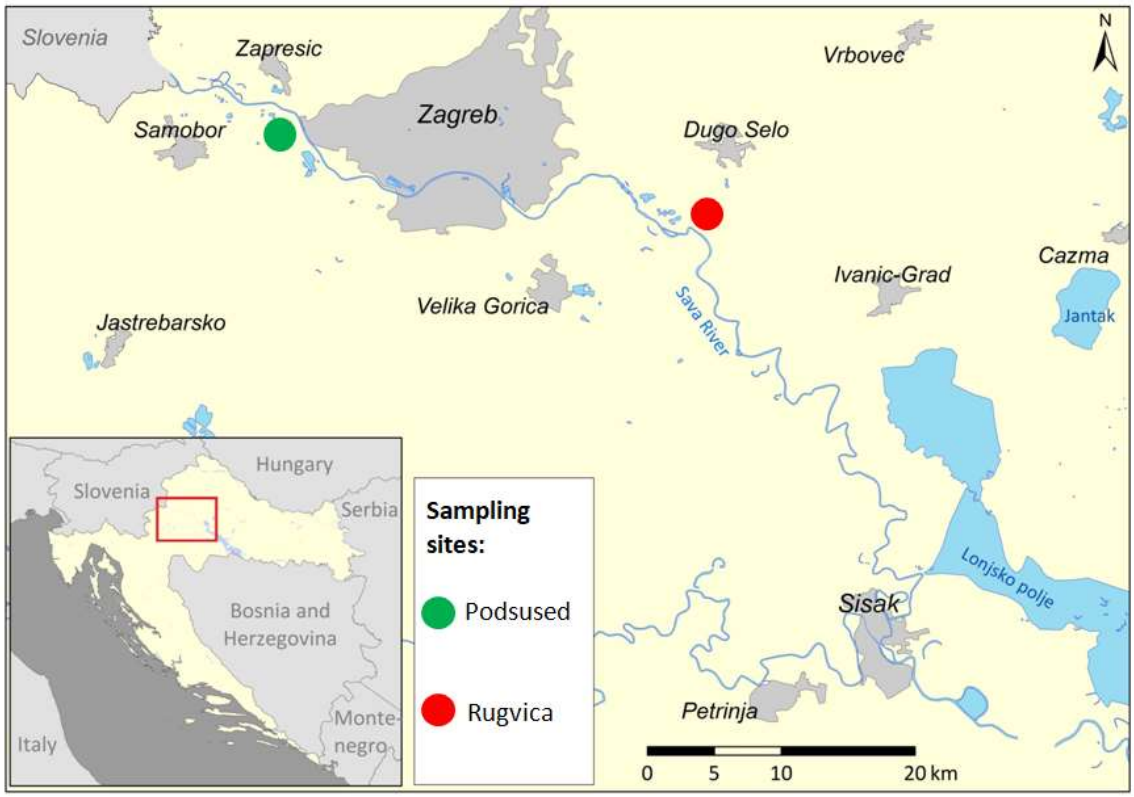


Slika 7. Karta Hrvatske i rijeke Save s mjestima uzorkovanja (ArcGIS 10.1 program). Sivom bojom su na karti označena naseljena područja, preuzeto iz [33].

3.1.2 Uzorkovanje riblje plazme

Uzorke riblje plazme uzorkovali su znanstvenici Biološkog odsjeka, PMF te Instituta Ruđer Bošković tijekom travnja 2016., na dvije lokacije u rijeci Savi u Hrvatskoj. Uzorkovane su i površinske vode s istih postaja. Prvo mjesto uzorkovanja nalazi se uzvodno od Zagreba (mjesto Podsused), a drugo nizvodno od Zagreba (mjesto Rugvica). Podsused je nisko do umjereno zagađeno mjesto uzorkovanja. Rugvica se nalazi ~ 10 km nizvodno od izlaza odvoda otpadnih

voda iz postrojenja za pročišćavanje otpadnih voda (tablica 5). Lokacije su prikazane kartografski na slici 8.



Slika 8. Karta Hrvatske i položaja sliva rijeke Save s područjima uzorkovanja površinskih voda i riblje plazme opisanih u studiji. Siva područja predstavljaju gradske zone. Karta je preuzeta iz rada [66].

Uzorci površinske vode sakupljeni su i analizirani u laboratoriju Hrvatskih voda prema opisu u radu [132]. Ribe su uzorkovane prema opisu u radu [133] te obrađene prema opisu do dobivanja uzoraka plazme [66]. S lokacije Podsused uzeta je plazma od 10 pojedinačnih primjeraka klena (*Squalius cephalus*) kako bi se dobilo pet kombiniranih uzoraka za kemijsku analizu. Nadalje, s lokacije Rugvica prikupljeno je pet pojedinačnih primjeraka klena kako bi se dobila tri kombinirana uzorka klena. Uzorci plazme pet pojedinačnih primjeraka mrene (*Barbus barbus*) prikupljeni su kako bi se dobila tri kombinirana uzorka za analizu kemijskih spojeva s lokacije Podsused. Jedan je kombinirani uzorak plazme koji pripada dvoprugastoj ukliji (*Alburnoides bipunctatus*) (od četiri pojedinačne jedinke) s lokacije Podsused također uključen u analizu. Ukupno je 12 kombiniranih uzoraka plazme. Analize su u ovom radu predstavljene kao neovisne spram ribljih vrsta s obzirom na ograničen broj ulovljenih živih primjeraka. Jedan je

jedini kombinirani uzorak dvoprugaste uklije isključen iz statističkih analiza. Pridržavane su sve primjenjive međunarodne, nacionalne i/ili institucionalne smjernice za njegu i uporabu životinja.

3.2 Obrada podataka i računalna analiza prikupljenih podataka

Cjelokupna obrada prikupljenih podataka i sve računalne analize u ovom radu napravljene su algoritmima napisanim u programskom jeziku Python. [134] Python je skriptni objektno-orijentirani programski jezik visoke razine. Njegove su velike prednosti: otvoren kod, iznimno čitka sintaksa, cijena (besplatan je), velika zajednica korisnika i time razvijene programske biblioteke. Bitan koncept u Pythonu je da je sve objekt. Objekti imaju metode (funkcije koje se vrše na objektima) i atribute (svojstva). U programskom će se kodu u ovom radu često pozivati metode za pojedine objekte. Sve su skripte pisane u programu PyCharm (https://www.jetbrains.com/pycharm , preuzeto 29. travnja 2021. god.). Korištene su sljedeće Python biblioteke: StatsModels [135], Scikit-learn [136], Scipy i NumPy [137]. Neke su od Pythonovih biblioteka, kao što je Pandas [138] posebno namijenjene obradi podataka i statističkoj analizi. Grafovi i vizualizacije su pripremljeni koristeći Pythonove biblioteke Seaborn [139] i Matplotlib [140]. Tamo gdje Pythonove biblioteke nisu mogle biti korištene, koristili su se programi GraphPad Prism 6.01 (GraphPad Software Inc., USA) i MS Excel (Microsoft, USA). Koncentracije nađenih KORZu uzorcima sedimenta, riblje plazme i površinske vode analizirane su na moguće obrasce zajedničke pojavnosti (engl. *co-occurence*), grupiranje u vizualnim analizama te međusobne korelacije (Pearson, Spearman).

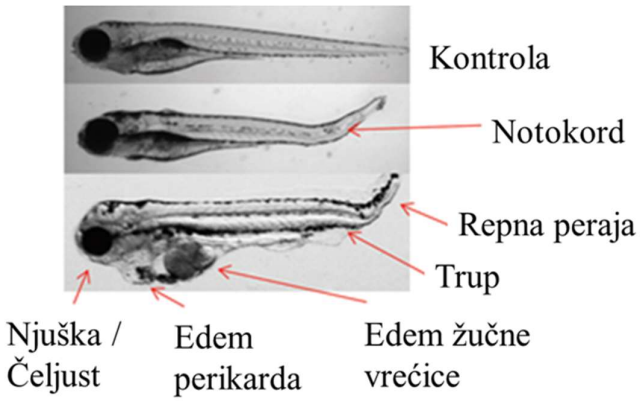
3.3 Ekotoksikološki skupovi i baze podataka

Podaci potrebni za *in silico* modeliranje dostupni su u on-line bazama toksikoloških i ekotoksikoloških podataka kao što su ECOTOX [141] i PAN [142]. Podaci u takvim bazama su često vrlo varijabilni zbog različitih trajanja izlaganja, višestrukih mjerenih toksičnosti i tipa formulacije testiranih kemijskih spojeva. U posljednjih se nekoliko godina svjedoči velikom porastu dostupnosti podataka o vrijednostima akutne i kronične toksičnosti za veliki broj kemikalija, kao i podataka o biološkim učincima spojeva, uključujući i KORZ. [143] Važan izvor podataka o toksičnim učincima velikog broja spojeva, koji se kontinuirano generira, a kojim upravlja Američka agencija za zaštitu okoliša (US Environmental Protection Agency, EPA), je Toxicity ForeCaster (ToxCast ili CompTox) projekt [144] i s njime povezani Tox21 programi. [143] Spomenuti programi procijenili su u posljednjih deset godina toksično djelovanje više od 8000 spojeva [145] koristeći 1200 *in vitro* testova, kojima su obuhvatili više od 300 signalnih puteva

(on-line baza podataka: ToxCast dashboard (https://actor.epa.gov/dashboard/#Chemicals , preuzeto 29. travnja 2021. god.).

3.3.1 Opis skupa ToxCast

ToxCast baza podataka ima niz povoljnih karakteristika za ekotoksikološku evaluaciju. Jedna od tih prednosti je veliki broj spojeva ispitanih na zebricama (N ~ 1060 jedinstvenih US EPA ToxCast faza 1 i 2). [22,146] Baza također pokriva veliku kemijsku raznolikost te relativno nove podatke prikupljene visokoprotočnim metodama. Predmet je ovog rada 18 istovremeno izmjerenih toksičnosti (toksični učinci) na embrijima ribe zebrice (engl. *Zebra fish*, *lat*. *Danio rerio*). Transparentni embriji koji se razvijaju nakon oplodnje (engl. *post fertilization*) čine toksikološku procjenu jednostavnom uz pomoć automatiziranih sustava opremljenih senzorima. Velika je većina genetskog koda kod embrija izražena i aktivna tijekom ranih životnih faza [22], stoga procjena na embrijima daje značaj i za druge kralježnjake i sisavce uključujući i čovjeka. 18 je toksičnih učinaka opisanih u bazi podataka ribe zebrice s vremenom izlaganja od 120 hpf (engl. *hours post fertilization*): mortalitet (MORT), plivanje (SWIM), edem žučnevrećice (YSE), devijacija notokorda (NC), devijacija tjelesne osi (AXIS), deformacija oka (EYE), deformacija njuške (SNOU), deformacija čeljusti (JAW), deformacija ušnog mjehurića (OTIC), perikardijalni edem (PE), deformacija mozga (BRAI), skraćeno tijelo (TRUN), deformacija somita (SOMI) (praorgan u razvoju embrija) , deformacija prsne peraje (PFIN), deformacija repne peraje (CFIN), promjena u pigmentaciji (PIG), promjene cirkulacije (CIRC) i reakcija na dodir (TR). Uz ovih 18 toksičnih učinaka iz baze je preuzet i kumulativni učinak ActivityScore koji predstavlja bilo koji oblik deformacije. Neki od navedenih organa prikazani na slici 9.



Slika 9. Prikaz kontrolnog embrija i dvaju embrija s vidljivim pojedinim promjenama na organima. [22]

Kemijska je raznolikost 1060 spojeva od velikog značaja i uključuje: aditive (kao što su plastika, pesticidi, insekticidi itd.), međuprodukte, reaktante, reagense, otapala, tenzide, antioksidanse, biocide, bojila, industrijske kemikalije i FADM. [22] Podaci su preuzeti iz baze ToxCast s poveznice (https://actor.epa.gov/dashboard/ preuzeto 29. travnja 2021. god.) Podatkovne su datoteke iz ove baze preuzete u formatu .csv i združene horizontalno prema indeksu spojeva iz baze, tzv. DTXSID. Uz ID datoteke sadrže SMILES strukture mapirane prema indeksu DTXSID. Aktivnost spojeva u podacima dana je kao HIT CALL (binarni zapis, aktivno/neaktivno, 1/0, gdje 'aktivno' znači 'toksično'). Potpuna tablica s toksičnim učincima iz skupa ToxCast dana je u elektroničkom dodatku E1.

3.3.2 Opis skupa Tox21

Baza podataka Tox21 skup je od ~8000 jedinstvenih spojeva [145] čije su toksičnosti ispitane spram stanica (staničnih linija) u formatu visoko-protočnih testova (engl. *high throughput screening*, HTS) kao serija titracija od 15 točaka u tri ponavljanja (https://ntp.niehs.nih.gov/whatwestudy/tox21/index.html, preuzeto 29. travnja 2021. god.). Titracijske serije predstavljaju krivulju koncentracija-odgovora za svaki testirani kemijski spoj u triplikatima, što uvelike smanjuje učestalost lažnih pozitivnih i negativnih rezultata. Istraživanje je usmjereno na aktivnost sedam nuklearnih receptora (engl. *Nuclear Receptor,* NR*)* i pet testova staničnih odgovora na stres (engl. *stress response*, SR). [147] Opisi toksičnih učinaka navedeni su niže u tekstu, a pregled istih dan je u tablici 6.

P53 → Protein, supresor rasta tumora, aktivira se nakon staničnog oštećenja, uključujući oštećenje DNA i druge stanične stresove. Aktivacija p53 regulira staničnu sudbinu potičući popravak DNA, zaustavljanje staničnog ciklusa, apoptozu ili stanično starenje. Stoga je aktivacija p53 dobar pokazatelj oštećenja DNA i drugih staničnih stresova.

ER-BLA i ER-BG1 agonist → Estrogenski receptor (ER); stanični hormonski receptor koji igra važnu ulogu u razvoju, metaboličkoj homeostazi i reprodukciji. Postoje dvije podvrste ER, ER-alfa i ER-beta koje imaju slične obrasce ekspresije. Kemijski spojevi koji ometaju endokrini sustav, tj. koji su disruptori endokrinog sustava (DES), i njihove interakcije s receptorima steroidnih hormona poput ER uzrokuju poremećaj normalne endokrine funkcije.

Aromatase → DES ometaju biosintezu i normalne funkcije steroidnih hormona, uključujući estrogen i androgen. Aromataza katalizira pretvorbu androgena u estrogen i ima ključnu ulogu u održavanju ravnoteže androgena i estrogena u mnogim organima osjetljivim na DES.

AhR → Arilni ugljikovodični receptor (AhR), član obitelji osnovnih transkripcijskih čimbenika zavojnica-petlja-zavojnica, presudan je za prilagodljive reakcije na promjene u okolišu. AhR posreduje staničnim odgovorima na onečišćivala okoliša poput aromatskih ugljikovodika

indukcijom enzima faze I i II detoksikacije, ali također komunicira s drugim signalnim putovima nuklearnih receptora.

AR-MDA i AR-BLA agonist → Androgeni receptor (AR), stanični hormonski receptor, ima presudnu ulogu u raku prostate koji je ovisan o AR. DES i njihove interakcije s receptorima steroidnih hormona poput AR mogu uzrokovati poremećaj normalne endokrine funkcije, kao i ometati metaboličku homeostazu, reprodukciju, razvoj i ponašanje.

ARE → Oksidativni stres uključen je u patogenezu raznih bolesti, od raka do neurodegeneracije. Signalni put elementa antioksidativnog odgovora (ARE) igra važnu ulogu u ublažavanju oksidacijskog stresa. Stanična linija CellSensor ARE-bla HepG2 (Invitrogen) može se koristiti za analizu signalnog puta Nrf2 / antioksidativnog odgovora.

ATAD5 → Stanice raka se brzo dijele i tijekom svake diobe dupliciraju svoj genom. Ako ne uspiju, stanica raka umire umire. Na temelju ovog koncepta razvijena su mnoga kemoterapeutska sredstva. S time u vezi je razvijen novi stanični test kako bi se pronašli spojevi koji učinkovito blokiraju replikaciju DNA izravnim oštećivanjem DNA ili inhibiranjem drugih staničnih mehanizama.

HSE-BLA → Različiti kemijski spojevi, uvjeti okoliša i fiziološki stres mogu dovesti do aktiviranja reakcije toplinskog šoka (odgovora raskloplјenog proteina). Postoje tri čimbenika transkripcije pri toplinskom šoku koji posreduju u transkripcijskoj regulaciji. Kod ovog receptora se radi o disrupciji transkripcije.

Mitochondria toxicity → Potencijal mitohondrijske membrane (PMM), jedan od parametara funkcije mitohondrija, generira se protokom elektrona kroz mitohondrijski lanac za prijenos elektrona koji stvara elektrokemijski gradijent nizom redoks reakcija. Ovaj gradijent pokreće sintezu ATP-a, ključne molekule za različite stanične procese. Mjerenje PMM u živim stanicama obično se koristi za procjenu učinka kemijskih spojeva na funkciju mitohondrija; smanjenje PMM-a može se otkriti upotrebom lipofilnih kationskih fluorescentnih boja.

PPAR-gamma agonist → Receptori aktivirani proliferatorom peroksisoma (PPAR) su lipidno aktivirani transkripcijski čimbenici natporodice staničnih receptora s tri različita podtipa, naime PPAR alfa, PPAR delta (također nazvana PPAR beta) i PPAR gama (PPARg). Svi navedeni podtipovi heterodimeriziraju se s receptorima Retinoid X (RXR) i ti heterodimeri reguliraju transkripciju različitih gena. PPAR-gama receptor (glitazonski receptor) sudjeluje u regulaciji metabolizma glukoze i lipida.

Tablica 6. Toksičnosti skupa Tox21 s brojem aktivnih/toksičnih (N = 1) i neaktivnih/netoksičnih spojeva (N = 0) s obzirom na ishod nakon pročišćavanja struktura spojeva. Toksičnosti koje počinju sa SR receptori su odgovora na stres (engl. *Stress Response*), a NR su nuklearni receptori (engl. *Nuclear Receptor*).

Toksični učinak	N = 0	N = 1	N/A	Stanična linija	Stanice
SR-HSE	6491	362	1291	HeLa	Rak grlića maternice
NR-AR	7238	321	594	MDA-MB-453	Rak dojke
SR-ARE	5160	1023	1961	HepG2	Jetra
NR-Aromatase	5741	333	2070	MCF-7	Rak dojke
NR-ER-LBD	6931	355	858	BG1	Jajnici
NR-AhR	6053	849	1242	HepG2	Jetra
SR-MMP	5172	956	2016	HepG2	Jetra
NR-ER	5655	846	1643	BG1	Jajnici
NR-PPAR-gamma	6596	213	1335	HEK293	Bubreg
SR-p53	6644	467	1033	HCT-116	Rak crijeva
SR-ATAD5	7082	310	752	HEK293	Bubreg
NR-AR-LBD	6812	235	1097	MDA-MB-453	Rak dojke

3.4 Računalni alati za prioritizaciju spojeva

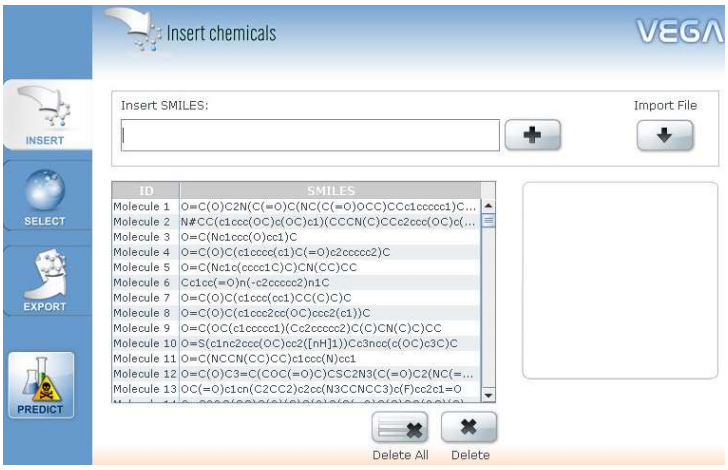
Budući da su toksikološki podaci rijetki, posebice za KORZ, koja se pojavljuju u niskim koncentracijama i tragovima, teži se alatima koji mogu dati procjenu toksičnosti dobro poznatim QSAR modelima. Primjer takvih alata su VEGA-QSAR [148] i Prometheus. [149] Slikoviti prikaz korištenja VEGA-QSAR dan je na slici 10. To su alati za prediktivnu toksikologiju, razvijeni mahom od strane istraživača Instituta Mario Negri (*Istituto di Ricerche Farmacologiche Mario Negri)* kao posljedica niza europskih istraživačkih projekata pod pokroviteljstvom OECD-a. VEGA-QSAR može predvidjeti niz toksikoloških ishoda, što humanih što životinjskih, na temelju unutarnjih modelnih parametara za pojedini ishod, za koje

postoje podaci u bazama na kojima su ti modeli razvijeni, optimirani i vrednovani. Svi se unutarnji modeli temelje na poznatim softverskim alatima razvijenim uz potporu EU i OECD kao što su CAESAR [150,151] i DEMETRA [152]. Prometheus je orijentiran na prioritizaciju računanjem PBT bodova što označava perzistenost, biološku koncentraciju i toksičnost (engl. *Persistence*, *Bioconcentration*, *Toxicity*) za predani popis SMILES struktura. VEGA ima širok spektar uporabe i računa predviđanja za nekoliko toksikoloških modela. Model od značaja iz alata VEGA-QSAR je Fathead Minnow 96 h LC50 (FatMin_96h) [153], tj. akvatična toksičnost koja odgovara ribljoj vrsti velikoglavi klen (*Pimephales promelas*), a predstavlja koncentraciju onečišćivala pri kojoj 50 % tretiranih/ispitivanih jedinki ugine). Model je treniran na 726 spoja pomoću kNN algoritma i referentan je model za akutnu toksičnost (96 h izlaganja) kod riba.

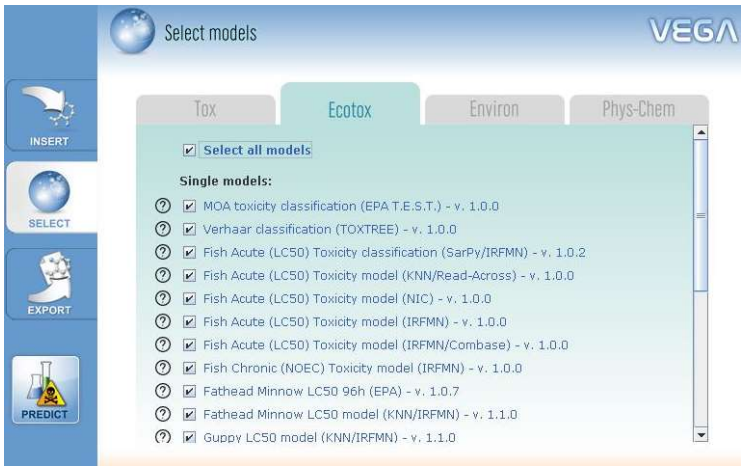
3.5 Softverski alati za izradu QSAR modela

Za rukovanje je podacima, pretvorbe i filtriranja korištena biblioteka Pandas. [138] Numpy i Scipy [137] biblioteke su razvijene za rad s matricama i algebarskim operacijama. Za razvijanje i podešavanje modela korištene su Python biblioteke scikit-learn [136], statsmodels [154], imblearn [155] i bayes_opt [98]. Za računanje PVV korištena je biblioteka eli5 [156] pomoću koje se računa permutacijska važnost varijabli. Vrijednosti log*P* ($\log K_{ow}$) izračunate su programom ChemAxon Marvin [157]. ChemAxon je komercijalni vrlo sofisticirani program za crtanje, konverziju i ispis kemijskih struktura s modulom za standardizaciju kemijskih struktura (Standardizer). Standardizer je vrlo intuitivan alat za standardizaciju kemijskih struktura koji uz više vrsta mogućih ulaznih formata podržava i spremanje standardizacijskih procedura u .xml podatkovnom formatu za ponovnu uporabu. U ovom je radu ChemAxon korišten i za konverzije formata, pregledavanje struktura te standardizaciju i geometrijsku optimizaciju spojeva u dvodimenzionalnom i trodimenzionalnom prostoru. [157] Za obradu se molekulskih zapisa, konverziju formata te računanje molekulskih deskriptora (DS) i otisaka (FP) koristila besplatna Python biblioteka RDKit, [158] koju je lako integrirati u servere i ostale softvere.

a)



b)



c)

Slika 10. Sučelje programa VEGA-QSAR. Pomoću sučelja preuzimaju se kemijske strukture iz datoteke sa SMILES zapisima (a), zatim se odaberu traženi modeli iz baze modela (b), iz zadanih modela ekstrapoliraju se vrijednosti za zadani skup spojeva (c).

§ 4. EKSPERIMENTALNI DIO

4.1 Kemijske analize i test embriotoksičnosti

Kemijska onečišćivala od rastućeg značaja za okoliš **(**KORZ) karakterizirana su u uzorcima sedimentnih ekstrakata i riblje plazme tekućinskom kromatografijom s masenom spektrometrijom s TOF detektorom (UHPLC-QTOF-MS) u laboratoriju Hrvatskih Voda. Detalji su pripreme uzoraka i analitičke metode sustava opisani u [132]. Spojevi su kvantificirani korištenjem MS skenirajućeg moda (maseni raspon od 50 do 1000 m/z) s pomoću točno mjerene mase i masene podudarnosti uzorka izotopa, kako je definirano u [132]. Podaci o izmjerenim masama obrađeni su dalje s pomoću programa Agilent MassHunter (verzija B.07.00/ Build7.0.457.0, Agilent Technologies, SAD). Popis je svih analiziranih kemijskih spojeva iz uzoraka sedimenta, vode i riblje plazme prikazan u elektroničkom dodatku E2. Svi su kvantificirani KORZ raspoređeni u 12 klasa spojeva prema terapeutskom učinku, kako je predloženo u recentnim publikacijama [33,66,132], tablica s kategorijama priložena je u tablici sa spojevima u dodatku E2. Kao nadopuna analitičkim metodama, a u svrhu određivanja toksičnosti ekstrakata sedimenta uzrokovanih s lokacija Jesenice, Rugvica, Lukavec i Galdovo, korišten je test embriotoksičnosti na zebricama (*D. rerio)*. Test je embriotoksičnosti proveden prema standardiziranom OECD 236 [159] protokolu, s nekoliko modifikacija. [33] Testiranje embriotoksičnosti (engl. *zebrafish embriotoxicity*, ZET), opisano u ovom radu napravljeno je na Institutu Ruđer Bošković. Tom je eksperimentu prethodilo preliminarno istraživanje provedeno na 10 embrija na širokom rasponu koncentracija ekstrakata sedimenta (100×, 1000×, 10 000×, 25 000×, 50 000× i 100 000× razrjeđenje) kako bi se utvrdio koncentracijski raspon od interesa. U skladu s dobivenim rezultatima, embriji *D. rerio* izlagani su trima koncentracijama ekstrakata sedimenta: 25 000×, 50 000×, 100 000× razrjeđenje. Dvanaest je embrija u duplikatu pojedinačno raspoređeno u mikroploče s 24 jažice (2 ml po jažici), što iznosi ukupno 24 embrija po svakoj ispitivanoj koncentraciji. Kao negativna kontrola i otopina za razrjeđivanje uzoraka korištena je umjetna voda. [159] Kao pozitivna je kontrola korišten 3,4-dikloroanilin. Ploče su inkubirane (26 ± 1 °C) pod reguliranim fotoperiodima dana i noći (14/10 h). Ispitivani su uzorci i kontrole svakodnevno zamijenjeni (50 % ukupnog volumena po jažici) prethodno aeriranom i ugrijanom otopinom. Nakon 24 i 48 sati po oplodnji embriji su pregledani, tijekom čega su zabilježeni letalni i subletalni učinci ispitivanih uzoraka (invertni mikroskop Olympus CKX41) [160] te je obavljen histopatološki pregled. Letalnim se učinkom

smatrala koagulacija embrija, neformiranje somita, neodvajanje repa od žumanjčane vreće i prestanak rada srca. [159]

4.2 Postupak prioritizacije spojeva u sedimentu i vodi

4.2.1 Računanje toksikoloških jedinica (TU)

Toksikološka jedinica (engl. *toxic unit*, TU) definirana je kao omjer koncentracije kemijskog spoja i odabrane vrijednosti toksičnosti (EC50 ili LC50). Koristi se za karakterizaciju toksičnog učinka na vodene organizme u vodenim staništima [161] za pojedini spoj, jednadžba 4.1. Temelj je TU da je spoj rizičniji za organizam što mu je koncentracija viša a LC50 niži (viša mortalitet ili učinak pri što nižoj koncentraciji).

$$\mathrm{TU_i} = \frac{\mathrm{c_i}}{\mathrm{LC50_i}} \tag{4.1}$$

Kvantificirani kemijski spoj imat će vrijednost TU jedan (i logTU nula) ako je njegova koncentracija u vodi jednaka njegovoj LC50 vrijednosti. Za izračunavanje ekotoksikološkog rizika smjese kemijskih spojeva (KORZ) specifičnih za neku lokaciju pretpostavlja se da je valjano primijeniti aditivni model (engl. *concentration addition*, CA) [35,162] doprinosa pojedinačnih spojeva u smjesi. Do ukupne procjene rizika u tom modelu dolazi se zbrajanjem TU vrijednosti (jednadžba 4.2) pojedinačnih izmjerenih spojeva prisutnih u smjesi u određenoj mjerljivoj koncentraciji, procjenjenom toksičnošću koja je dobivena eksperimentalno ili predviđanjem s pomoću QSAR modela [163,164]:

$$\mathrm{TU_{SITE}} = \log \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{TU_i} \tag{4.2}$$

Uz pretpostavku postojanja ravnoteže između organskog ugljika u sedimentu i vodenom stupu iznad sedimenta [165]; $\mathrm{TU_{pw}}$ (engl. *pore water*) može se izračunati za KORZ izmjerene u sedimentnim ekstraktima na osnovi njihovog koeficijenta raspodjele između vode i sedimenta, $\mathrm{K_p}$, (jednadžba 4.3).

$$\mathrm{K_p} = \frac{\mathrm{c_s}}{\mathrm{c_d}} = \mathrm{f_{oc}K_{oc}} \tag{4.3}$$

gdje je c_s koncentracija KORZu sedimentu [μg/kg], $\mathrm{c_d}$ njihova koncentracija u vodenom stupcu [μg/L], $\mathrm{K_{oc}}$ je koeficijent raspodjele između organskog ugljika u sedimentu i vode, a $\mathrm{f_{oc}}$ maseni

udio organskog ugljika u sedimentu koji se mjerio za sve četiri lokacije uzorkovanja. K_{oc} računa se s pomoću K_{ow} prema [166] (jednadžba 4.4). Računanje K_{ow} opisano je u Poglavlju 3.4.

$$\log K_{oc} = 1{,}03 * \log K_{ow} - 0{,}61 \quad (4.4)$$

Finalno se TU_{pw} (za ribe) može izračunati iz izvedene jednadžbe 4.5.

$$TU_{pw} = \frac{c_s}{f_{oc} K_{oc} LC_{50}} \quad (4.5)$$

TU_{zet} (jednadžba 4.6.) koristi se za procjenu toksičnosti otopina ekstrakta u testu embriotoksičnosti, a računa se iz koncentracija KORZ u ekstraktu i njihove LC50 vrijednosti za FatMin_96h dobivene pomoću programa VEGA-QSAR (Poglavlje 3.4):

$$TU_{zet} = \frac{c_{ekstrakt}}{LC50} \quad (4.6)$$

gdje je $c_{ekstrakt}$ koncentracija KORZ u ekstraktima.

4.2.2 Rangiranje PBT faktora

PBT (perzistencija, biokoncentracija, toksičnost) rangiranje prioritizacijska je metoda [149] kojom se procjenjuje opasnost KORZ na temelju poznate ili izračunate perzistencije (P), biokoncentracije (B), toksičnosti (T). Kako procjena rizika KORZ ne ovisi samo o njihovim pojedinačnim PBT bodovima, nego i o izmjerenoj koncentraciji, predlaže se upotrijebiti rangiranje uz koncentraciju KORZ. U ovom će se radu koristiti PBT rangiranje (PBTr), koje je neparametarska metoda temeljena na uteženoj aritmetičkoj sredini pojedinačno dodijeljenih PBT rangova za svaki KORZ i njihove koncentracije u sedimentu (jednadžba 4.7). PBTr se koristi u ovom radu za određivanje prioritetnih KORZ s najvišim rangom rizika za svako mjesto uzorkovanja. PBT bodovi (engl. *scores*) izračunati su na temelju vrijednosti predviđenih programom Prometheus [149] i koncentracija (maseni udio) KORZ u sedimentu, c_s. Prosječne su vrijednosti PBT izračunate za svaki spoj, gdje je R_i rang PBT boda i koncentracije, w_i je njihovo uteženje. Svim uteženjima pridijeljena je vrijednost 1 jer su svi parametri promatrani kao jednako važni čimbenici koji doprinose potencijalnom riziku za vodene organizme.

$$PBTr = \frac{\sum_{i=1}^{n} R_{i,PBT} w_i}{\sum_{i=1}^{n} w_i} \quad (4.7)$$

4.3 Postupak prioritizacije spojeva u plazmi

4.3.1 Računanje omjera učinka

Izmjerene koncentracije za kvantificirane FADM u ribljoj plazmi uspoređene su s humanim terapijskim koncentracijama u plazmi (HKTP), tj. koncentracijama potrebnima za terapijsko djelovanje. Metoda usporedbe naziva se omjer učinka (engl. *effect ratio,* ER). [65] Omjer učinka ER izračunava se dijeljenjem maksimalne koncentracije u ljudskoj plazmi (C_{max}) na najmanjoj terapijskoj dozi lijeka s predviđenim koncentracijama u ribljoj plazmi. Umjesto da se za izračunavanje ER koriste Cmax [65] ili 1 % vrijednosti C_{max} [167], u ovom je radu korišten manje konzervativan pristup temeljen na razmatranjima iz literature. [167] Unutar raspona terapijskih koncentracija odabrana je najniža vrijednost kao minimalna koncentracija (C_{min}) lijeka potrebna za provođenje farmakološkog odgovora ili terapijskog učinka kod ljudi. C_{min} vrijednosti preuzete su iz literaturnih izvora [168,169] i sažete u radu [66]. Vrijednosti C_{min} podijeljene su sa stvarnim koncentracijama mjerenima u ribljoj plazmi, kako bi se dobile ER vrijednosti za kvantificirane FADM. ER vrijednosti korištene su u ovom radu za prioritizaciju potencijalno rizičnih kemijskih spojeva pronađenih u slatkovodnim ribama.

4.3.2 Prioritizacija FADM temeljena na ER

Za potrebe analize uzeti su u obzir samo spojevi u kojima je omjer učinka ER bio ispod 1000 - ukupno njih 50. ER vrijednosti raspodijeljene su u 5 kategorija kako bi se rangirale prema riziku koji predstavljaju za okoliš: ER < 1, ER1 - 10, ER10 - 100, ER100 - 1000 te ER > 1000. Vrijednosti ER < 1000 (uključen čimbenik sigurnosti/nesigurnosti) za pojedine FADM ukazuju na potencijalni dugoročni rizik za ribe. [65] Za prioritizaciju FADM u plazmi predlaže se neparametarski model prosječnog rangiranja ERa (R_{ER}), predložen u jednadžbi 4.8. R_{ER} je korišten za određivanje prioriteta FADM s najvišim rangom rizika na temelju vrijednosti ER izračunatih preko modela riblje plazme (MRP). Dodijeljeni su rangovi FADM za ER vrijednosti u svakom uzorku. Ako su dvije ili više ER vrijednosti identične, za sve takve se FADM izračunava i pridružuje prosječni rang.

$$\bar{R}_{ER,j} = \frac{1}{m} \sum_{i=1}^{m} R_{ER,j_i} \tag{4.8}$$

gdje je $\bar{R}_{ER,j}$ prosječni rang za spoj *j* u svim skupovima uzoraka *m*. R_{ER} rang je svakog FADM u uzorku *i*.

4.4 Priprema QSAR modela

4.4.1 Elementi potrebni za izradu QSAR modela

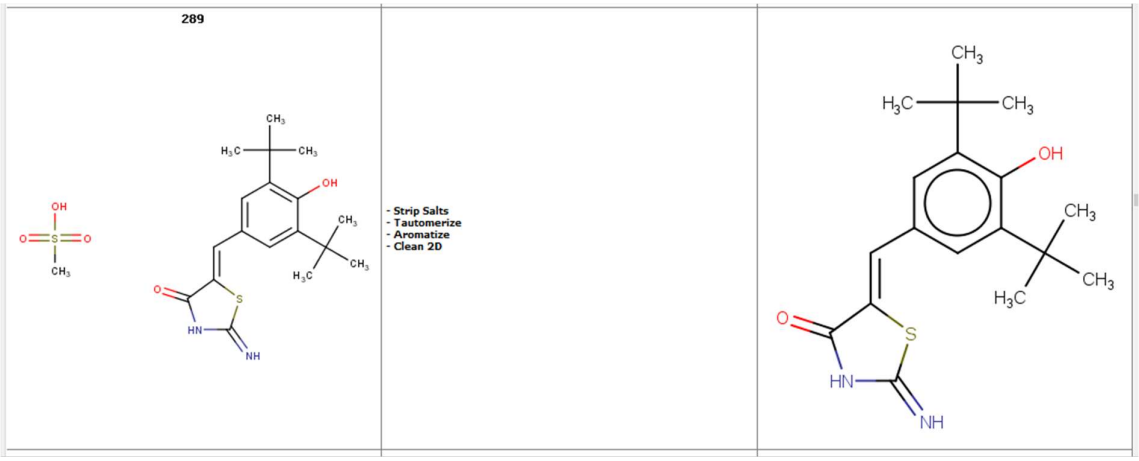
Razvoj, ugađanje, optimizacija i validacija (često nazvano i „učenje“) QSAR modela sastoji se od niza koraka koji su standardizirani, vremenom nadograđivani te izloženi u literaturi. [118,170,171] Uobičajeni se postupak QSAR modeliranja može svesti na:

1. Priprema podataka
    a. Odabir toksičnosti (ili kemijske) (y)
    b. Priprema prediktorskih varijabli (X)
        i. Provjera i priprema kemijskih struktura
        ii. Računanje molekulskih strukturnih otisaka ili deskriptora
2. Odabir prikladnog algoritma ili funkcije za rješavanje problema oblika y = *f*(X)
    a. Strojno učenje
    b. Statističke/kemometrijske metode
3. Podjela skupa spojeva na skup za učenje i skup za vanjsku provjeru modela
4. Učenje (ugađanje i optimizacija) modela na skupu za učenje (engl. *training set*)
5. Ispitivanje kvalitete modela na skupu za vanjsku provjeru (engl. *test set*)

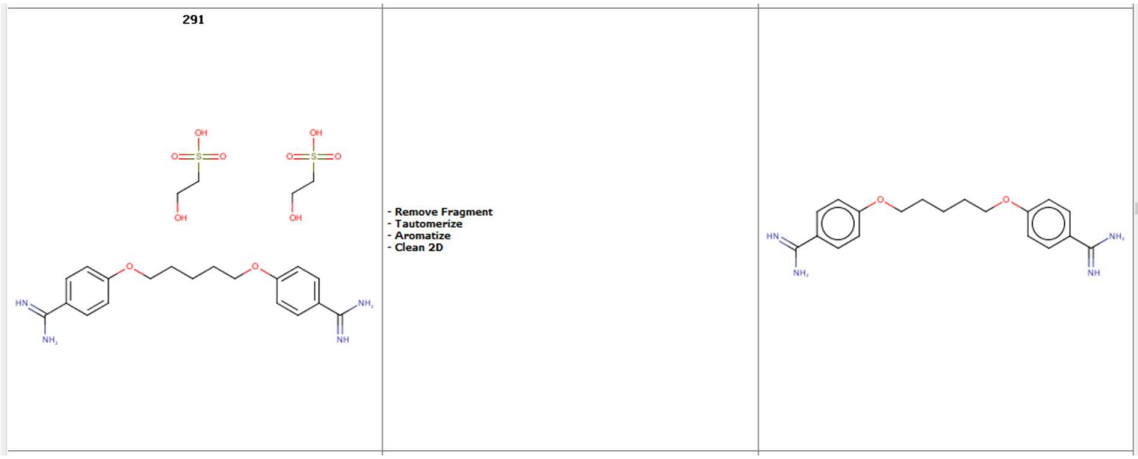
Zbog složenosti zadatka i velikog broja podataka – što toksičnih učinaka, što kemijskih spojeva - ovaj je rad usmjeren na automatizaciju svih navedenih postupaka. Modeli su optimirani Bayesovom optimizacijom [99] koja je opisana u Poglavlju 4.4.4, s deseterostrukom križnom validacijom na temelju tri algoritma (Logistička regresija, Neuronske mreže i Random Forest) koji su opisani u Poglavlju 2.5.4.

4.4.2 Automatizirana provjera struktura

Strukture su molekula u SMILES obliku (Elektronički dodatak E3) analizirane aplikacijom Marvin *Standardizer* u svrhu njihovih korekcija i optimizacija. Konfiguracija za standardizaciju obuhvaća sljedeće radnje: uklanjanje eksplicitnih vodikovih atoma, dearomatizaciju, pretvorbu π-metalnih veza, odvajanje metalnih atoma, skidanje soli, uklanjanje fragmenta, neutralizaciju, čišćenje valovitih veza, mezomerizaciju, tautomerizaciju, aromatizaciju, čišćenje 2D strukture, čišćenje 3D strukture. Standardizacije struktura provedene u ovom radu slijede preporuke iz literature. [170,172] Konfiguracija korištena u *Standardizer* aplikaciji u ovom radu dana je u XML konfiguracijskoj datoteci koja se nalazi u dodatku (Elektronički dodatak E4). Svaka struktura korigira samo one komponente iz procedure koje su na nju primjenjive. Primjeri standardizacije dani su u slikama 11. - 13. na kojima je vidljivo da su izbori standardizacijskih procedura individualni za svaku molekulu.

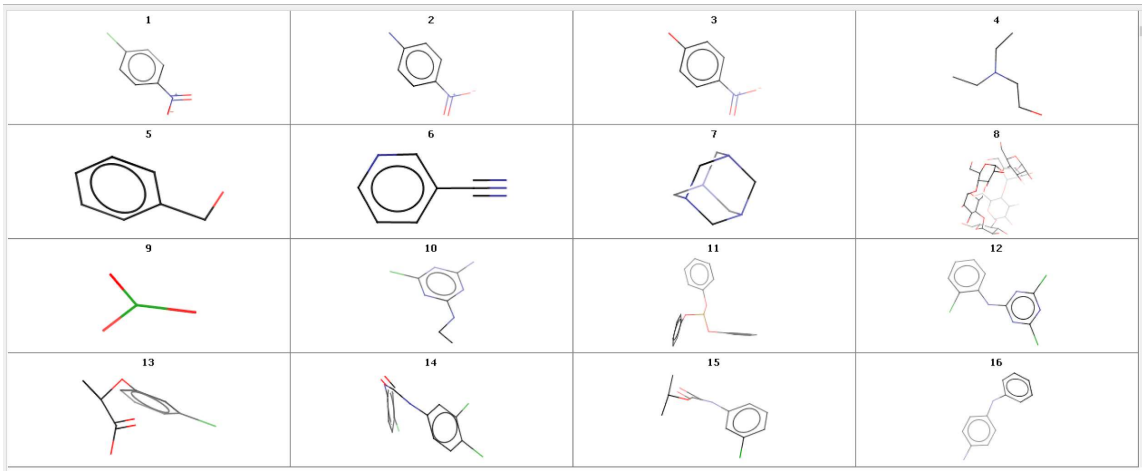


Slika 11. Primjer standardizacije strukture. Struktura je neutralizirana, odvojen je manji fragment.



Slika 12. Prikaz standardizacija strukture. Odvojeni su manji fragmenti i provedena je aromatizacija prstenova.

Prvi korak u pripremi podataka bilo je uklanjanje 15 spojeva koji nisu imali valjane SMILES strukture te je preostalo 1077 spojeva. Spojevi su indeksirani prema identifikatoru DTXSID. Zatim je uklonjeno 19 spojeva koji su identificirani kao duplikati na temelju istih identifikacijskih brojeva u bazi (ID). Nadalje, pregledane su SMILES strukture te je pronađeno i uklonjeno iz skupa 26 duplikata koji su nastali (uglavnom) nakon uklanjanja fragmenata.



Slika 13. Prikaz trodimenzionalnih struktura prvih 16 spojeva (iz skupa ToxCast).

Pomoću algoritma danog u tablici 7. provjerena je valjanost preostalih kemijskih struktura u SMILES zapisu. Ta se provjera temelji na uspješnosti pretvorbe strukture iz SMILES u *.mol* oblik. Ako pretvorba uspije, struktura se smatra valjanom, a u suprotnom algoritam javlja pogrešku ('error'). Svi preostali spojevi prošli su opisanu provjeru.

Tablica 7. Algoritam u Pythonu za provjeravanje valjanosti kemijske strukture. U algoritmu se u petlji provjeravaju SMILES strukture s pomoću RDKit modula Chem.MolFromSmiles koji pokušava pretvoriti SMILES u MOL zapis. U slučajevima kada pretvorbu nije moguće provesti, algoritam javlja pogrešku.

```
for i in [SMILES]:
    try:
        mol = Chem.MolFromSmiles(i)
        Chem.MolToSmiles(mol, isomericSmiles=True)
    except:
        print('error')
```

Potom je provedena provjera spojeva iz skupa prema sljedeća četiri koraka:

- Provjera je li struktura spoja organska?
- Provjera sadrži li strutura spoja metalne atome?
- Sadrži li struktura molekulske fragmente?
- Ručna provjera ispravnosti strukture.

U prvom su koraku identificirani anorganski spojevi, odnosno spojevi koji ne sadrže ugljične atome označene s ‘c’ ili ‘C’ u SMILES zapisu pomoću pretraživačke funkcije na SMILES zapisima. Identificirano je sedam anorganskih spojeva (tablica 8.) koji su uklonjeni iz skupa sukladno uputama iz literature. [170]

Tablica 8. Identificirani anorganski spojevi u skupu ToxCast izuzeti iz QSAR modeliranja.

DTXSID	**Naziv spoja**	**SMILES**
DTXSID0020941	natrijev nitrit	ON=O
DTXSID1020194	borna kiselina	OB(O)O
DTXSID1029677	silicijev dioksid	O=[Si]=O
DTXSID5020811	živin klorid	[Hg++]
DTXSID5023825	indij arsenid	[AsH3]
DTXSID8020121	natrijev azid	[N-]=[N+]=[N-]
DTXSID8021272	natrijev klorit	O[Cl]=O

U narednom su koraku strukture provjerene na prisutnost metalnih atoma. Budući programi ne mogu računati deskriptore za strukture koje sadrže soli i metalne komplekse, potrebno je takve spojeve ili u cijelosti izuzeti, ili izdvojiti za modeliranje samo aktivne fragmente. U tu je svrhu napisan programski kod kojim se pretražuju SMILES zapisi i provjerava prisutnost ključnih oznaka izlistanih u tablici 9. Rezultat pretrage prisutnosti metalnih atoma u spojevima izlistani su u tablici 10, a zbog toga je sedam spojeva izuzeto iz skupa za modeliranje (1025 – 7 = 1018).

Tablica 9. Popis SMILES kodiranih atoma za pretraživanje kemijskih struktura.

Popis metalnih atoma	['Li','Be','.*B[^.r].*', 'Al' , 'Si' , 'Sc' , 'Ti' , 'V' , 'Cr' , 'Co' , 'Ni' , 'Cu' , 'Ga' , 'Ge' , 'As' , 'Se' , 'Rb' , 'Sr' , 'Y' , 'Zr' , 'Nb' , 'Mo' ,'Tc' , 'Ru' , 'Rh' , 'Pd' , 'Ag' , 'Cd' , 'In' , 'Sn' , 'Sb' , 'Te' , 'Cs' , 'Ba' , 'Hf' , 'Ta' ,'W' , 'Re' , 'Os' , 'Ir' , 'Pt' , 'Au' , 'Hg' , 'Tl' , 'Pb' , 'Bi' , 'Po' , 'At' , 'Rn' , 'Fr' , 'Ra' ,'Rf' , 'Db' , 'Sg' , 'Bh' , 'Hs' , 'Mt' , 'Ds' , 'Rg' , 'Cn' , 'Fl' , 'Lv' , 'La' , 'Ce' , 'Pr' , 'Nd' , 'Pm' , 'Sm' , 'Eu' , 'Gd' , 'Tb' , 'Dy' , 'Ho' , 'Er' , 'Tm' , 'Yb' , 'Lu' , 'Ac' , 'Th' , 'Pa' , 'U' , 'Np' , 'Pu' , 'Am' , 'Cm' , 'Bk' , 'Cf' , 'Es' , 'Fm' , 'Md' , 'No' , 'Lr']
Popis metalnih atoma koji često stvaraju soli	['Na' , 'Mg' , 'K' , 'Ca' ,'Zn', 'Mn' , 'Fe']
Popis halogenih atoma	['F','Cl' ,'Br','I','S','P']

Pročišćene SMILES strukture 1018 spojeva zapisane su u datoteku .sdf formata iz koje su, nakon prethodne optimizacije 3D struktura programom Marvin Standardizer, računati molekulski deskriptori i molekulski otisci programom RDKit. [173]

Tablica 10. Spojevi koji sadrže teške metalne atome, kao i anorganski spojevi, izuzeti su iz skupa i daljnje analize.

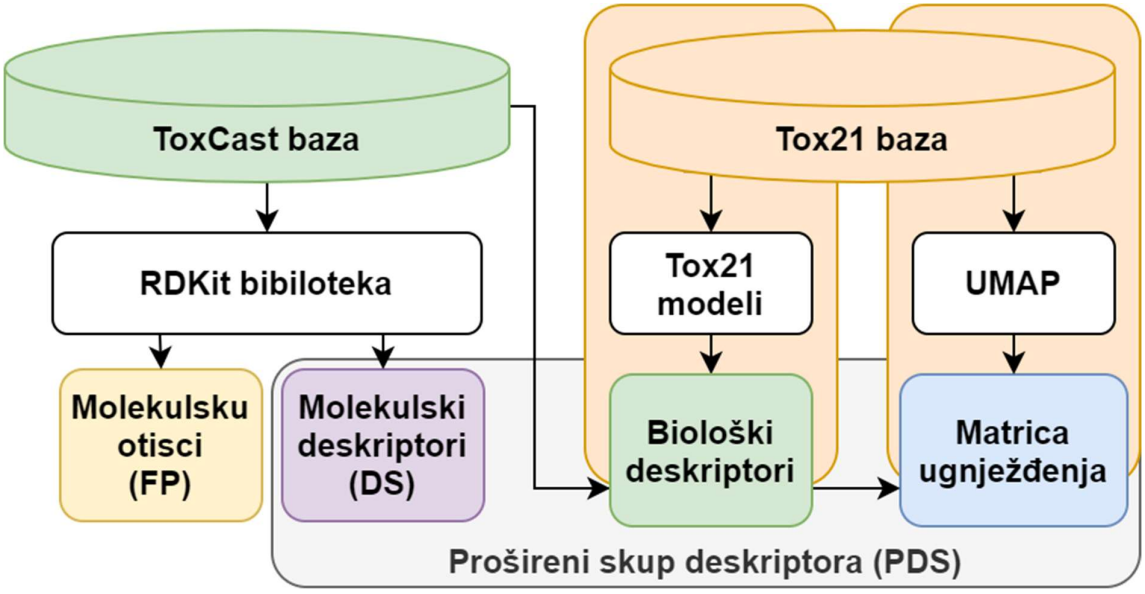
DTXSID	**Naziv spoja**	**Standardizirani SMILES**
DTXSID1021409	Trifeniltin-hidroksid	c1ccc(cc1)[SnH](c1ccccc1)c1ccccc1
DTXSID2029246	trietoksi-oktilsilan	CCCCCCCC[Si](OCC)(OCC)OCC
DTXSID3024235	flusilazol	C[Si](Cn1cncn1)(c1ccc(F)cc1)c1ccc(F)cc1
DTXSID5044493	1-[3-(trietoksisilil)propil]urea	CCO[Si](CCCNC(N)=O)(OCC)OCC
DTXSID7020508	dimetil-arsinska kiselina	C[As](C)(O)=O
DTXSID7021150	fenil-živin acetat	[Hg+]c1ccccc1
DTXSID7027205	oktametilciklotetra-siloksan	C[Si]1(C)O[Si](C)(C)O[Si](C)(C)O[Si](C)(C)O1

Izračunati deskriptori dani su u podatkovnoj tablici u elektroničkom dodatku E5. Identični postupak primijenjen je na spojeve iz baze Tox21 (poglavlje 3.3.2). Konačni broj spojeva u bazi Tox21 je 8144.

4.4.3 Molekulski deskriptori i molekulski otisci

Biblioteka RDKit računa 203 molekulska (fizikalno-kemijska) deskriptora (DS). Izračunati se deskriptori mogu podijeliti na skupine 1D, 2D i 3D deskriptora (popis deskriptora na engleskom jeziku nalazi se u dodatku D1). Biblioteka RDKit [173] također se koristila za računanje molekulskih otisaka (FP): duljina vektora 5120 bitova (binarni vektori), koji opisuju susjedstva atoma radijusa tri (tri susjedne pozicije). U pregledu različitosti kemijskih struktura napravljena je analiza skupa spojeva na temelju izračunatih deskriptora AMR (molekulska masa) i ALOGP [174] (particijski koeficijent) izračunatih programom RDKit. Uz opisane molekulske deskriptore i molekularne otiske izračunate programom RDKit u ovom radu bit će izračunati, razmatrani i rabljeni u modeliranju aktivnosti molekula iz skupa ToxCast i biološki deskriptori te matrica ugnježđenja. Biološki deskriptori dobiveni su kao predviđanje 12 bioloških svojstava spojeva iz baze ToxCast modelima koji su razvijeni na spojevima iz baze Tox21. Među spojevima iz te dvije baze može biti i preklapanja. Međutim, bitno je napomenuti da eksperimentalne biološke aktivnosti iz baze Tox21 nisu iste kao i aktivnosti iz baze ToxCast. One se razlikuju u vrsti organizma na kojem su mjereni toksični učinci spojeva a i u vrsti mjerenih aktivnosti. Podaci o toksičnim učincima Toksičnosti u bazi Tox21 dobiveni su kvantifikacijom i kvalifikacijom toksičnih učinaka na staničnim linijama, dok su podaci o toksičnosti u bazi ToxCast dobiveni kvantifikacijom i kvalifikacijom učinaka na embrijima zebrica (živi organizam). Baza Tox21 opisana je u Poglavlju 3.3.2.

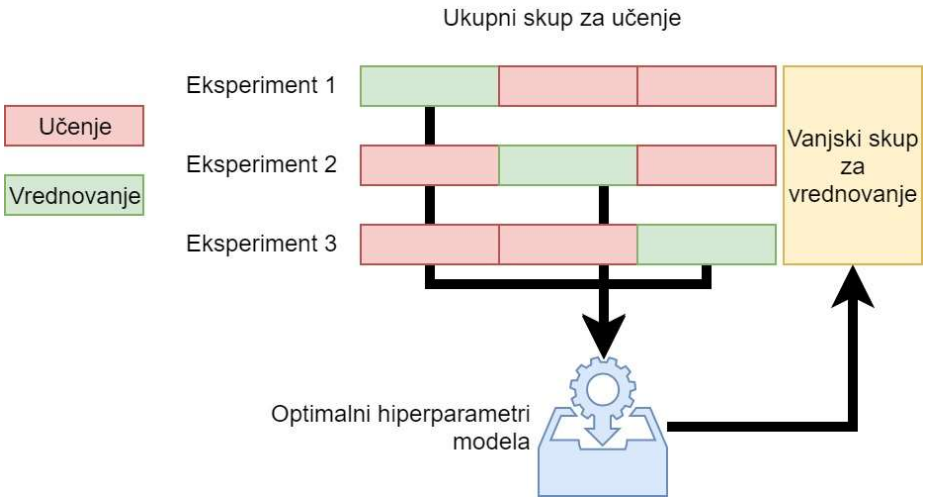
Matrica ugnježđenja koja se koristi u razvoju i optimizaciji modela objašnjena je u Poglavlju 2.9. U ovom radu ona se koristi kao dodatni skup deskriptora na način da su vektori matrice koordinate spojeva u trodimenzionalnom latentnom prostoru (tj. latentna reprezentacija skupa). Na taj način vrijednosti u ta tri vektora daju informacije o susjedstvima u kojima se nalazi (položaj na koordinatama) svake molekula. Reprezentacija je dobivena preslikavanjem skupa ToxCast na skup Tox21 i, stoga što preslikavanje ne uključuje informacije o susjedima iz istog skupa, takav postupak nije pristran. Matrica se sastoji od tri vektora koji nose informacije o kemijskom (strukturalnom) susjedstvu pojedinog spoja. Pregled prediktorskih varijabli korištenih u ovom radu dan je na slici 14. Skup prediktorskih varijabli (X) koji se sastoji od molekulskih deskriptora (DS), molekulskih otisaka (FP), bioloških deskriptora (12 modela) i matrice ugnježđenja naziva se prošireni skup deskriptora (PDS).



Slika 14. Shematski prikaz prediktorskih skupova (X) korištenih u ovom radu. Uz molekulske otiske (FP) i fizikalno-kemijske deskriptore (DS), izračunati su biološki deskriptori i matrica ugnježđenja koji s DS čine proširenih skup deskriptora (PDS). Tox21 baza poslužila je za izračun bioloških deskriptora i matrice ugnježđenja.

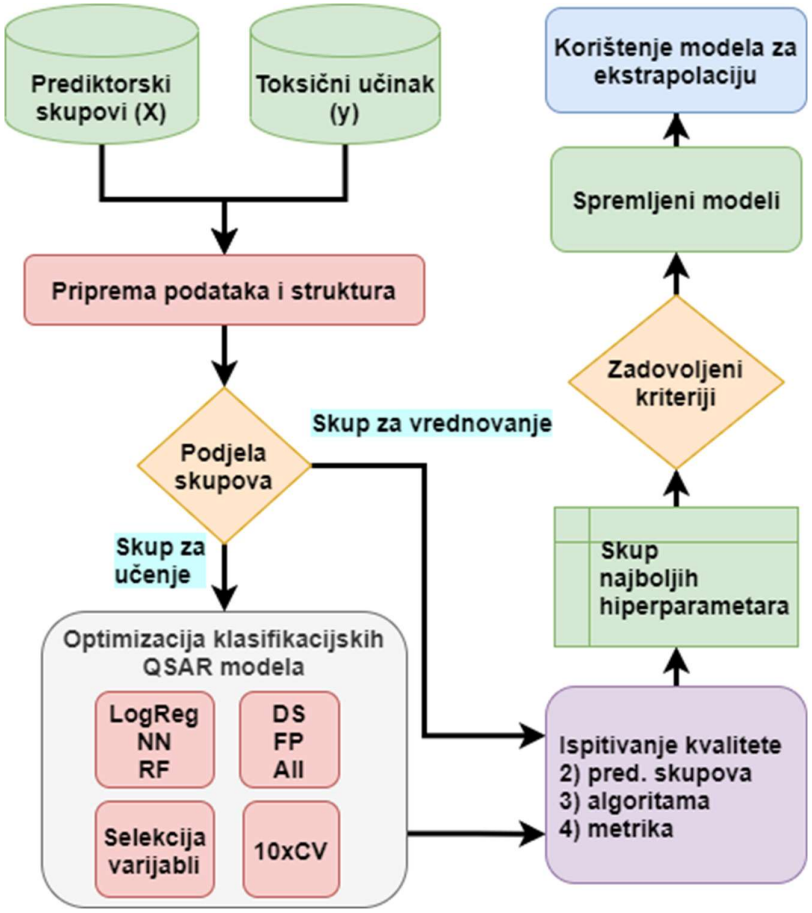
4.4.4 Optimizacija modela

Sve metode korištene u modeliranju opisane su u Poglavlju 2.5. Korištena su tri tipa klasifikacijskih algoritama: RF, NN i LogReg. Prije učenja modela svaki skup nasumično je podijeljen na spojeve za učenje (75 % spojeva, skup *X_train*) i skup za vrednovanje (25 % nasumičnih instanci, skup *X_test*). U podjeli skupa spojeva vodilo se računa o omjerima zastupljenosti manjinske i većinske klase (stratifikacija ciljne varijable, *y*). Budući da dva algoritma: RF i LogReg, imaju u sebi uključenu selekciju varijabli, stvarni će broj varijabli korištenih u ugađanju modela biti uvijek manji od ukupnog raspoloživih varijabli u skupu ToxCast. Tijekom učenja, svaki od modela optimiran je na skupu za učenje u postupku deseterostruke križne validacije (10xCV). Pritom, skup za učenje podijeljen je na deset podskupova s jednakim udjelima manjinske klase (slika 15).



Slika 15. Trostruka križna validacija prikazana kao primjer za razumijevanje postupka unakrsne validacije. U ovom radu korištena je deseterostruka unakrsna validacija (10xCV).

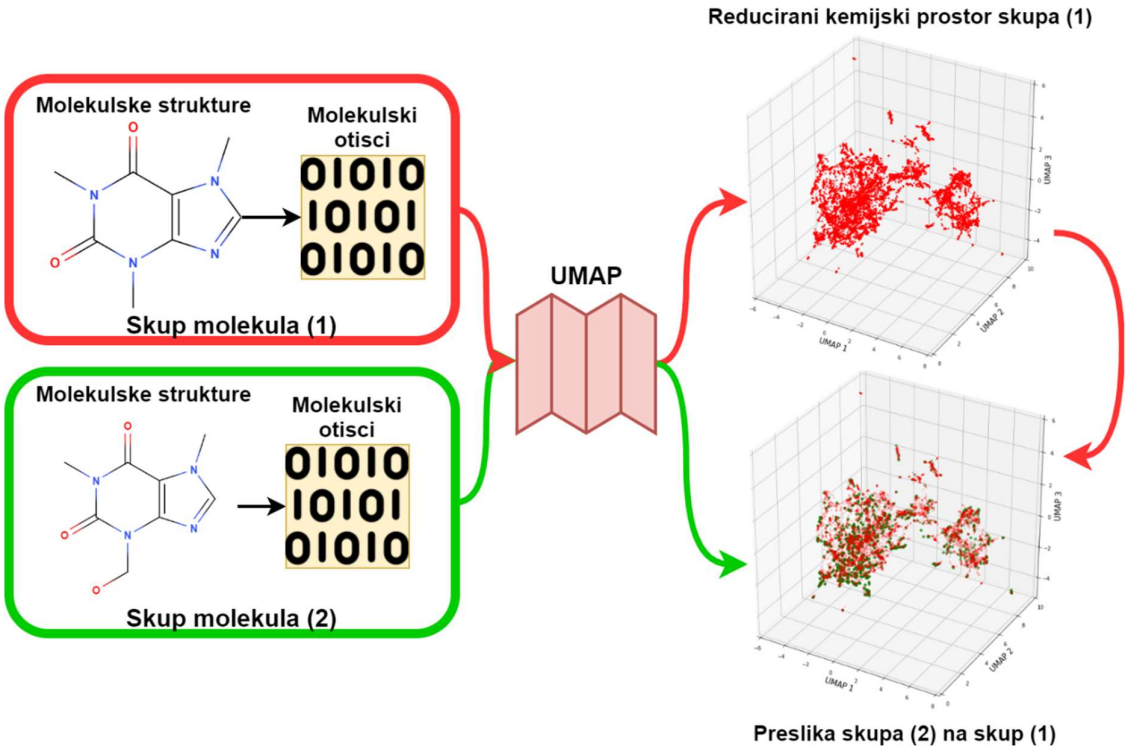
Algoritam iterativno optimira parametre modela na devet podskupova (9xCV), a kvaliteta se modela provjerava na preostalom (desetom) podskupu. Za optimizaciju hiperparametara pojedinog modela korištena je Bayesova optimizacija opisana u Poglavlju 2.5.4. Kvaliteta ugađanja modela tijekom optimizacije provjeravana je (i vođena) u svakom koraku vrijednošću koeficijenta korelacije (MCC) prilagođenom za binarne klasifikacijske varijable. Nakon završetka optimizacije izračunati su svi parametri kvalitete (mjere kvalitete, metrike) navedeni u Poglavlju 2.6 na skupu za učenje i na skupu za vanjsku provjeru. Izrađen je program za izbor optimalnih modela za sva tri korištena algoritma za modeliranje (RF, NN i LogReg). Program optimira modele za tri vrste algoritama za kreiranje modela, za različite skupove ulaznih prediktorskih varijabli (FP, DS, PDS) i sve uz opciju bez i sa selekcijom varijabli. Za logističku regresiju i neuronske mreže potrebno je skalirati (normirati) varijable. Kako je to tehnički zahtjevno rješenje za ponovnu uporabu razvijenih modela na serverskoj aplikaciji u korištenju za vanjskim skupovima molekula, za modele izgrađene i na molekulskim deskriptorima (DS) korišten je samo RF algoritam (stoga što za njega skaliranje deskriptora nije potrebno). Skripta za optimizaciju modela automatski sprema rezultate modeliranja u tri zasebne datoteke, i to: skup optimalnih hiperparametara, predviđanja i sve izračunate metrike. S obzirom na dozvoljene/nedozvoljene kombinacije, konačni skup sadrži 19 kombinacija modelnih parametara (klasifikatori, skupovi prediktorskih varijabli, selekcija varijabli) po modeliranom toksičnom učinku. Na slici 16 dan je shematski prikaz tijeka modeliranja.



Slika 16. Shematski prikaz postupka modeliranja i optimiranje modela. Za svaku od 19 ciljnih varijabli (toksičnih učinaka) izračunato je 19 modela, što ukupno daje 361 model.

4.5 Nelinearno projiciranje kemijskog prostora na manji broj dimenzija - matrica ugnježđenja

U ovom se radu UMAP, tehnika uvedena u Poglavlju 2.9, koristi za vizualizaciju kemijskog prostora, računanje prediktivnih svojstava i razumijevanje nelinearne domene primjenjivosti. Za to je potrebno imati skup molekula s izračunatim svojstvima ili molekulskim otiscima, koji se zatim nelineranim postupkom (algoritmom) UMAP transformira u prostor niže dimenzionalnosti, tj. u trodimenzionalni latentni prostor. Definirani se transformator može spremiti i ponovno koristiti na drugim skupovima s istim varijablama. Ovdje su molekularni otisci skupa Tox21 (N = 8144) pretvoreni putem UMAP-transformatora u 3D prostor, a shema postupka nalazi se na slici 17. S pomoću skupa (1) na toj slici trenira se UMAP-transformator, zatim se može neki drugi skup (2) transformirati na postojeći latentni prostor (latentna reprezentacija) te dobiti novu matricu ugnježđenja.



Slika 17. Shematski prikaz UMAP transformacije dvaju skupova molekula. Iz kemijskih se struktura računaju molekulski otisci. Zatim se skup (1) transformira na latentni prostor (reducirani kemijski prostor skupa (1), crvena boja). Istim se postupkom može preslikati svaki dodatni skup molekulskih otisaka (ovdje skup 2, zelena boja). UMAP-transformator može se spremiti i ponovno upotrijebiti.

Osim za vizualizaciju, matrica ugnježđenja korištena je za proširenje skupa deskriptora. Naime, na postojeći skup kemijskih deskriptora dodana su još tri vektora matrice ugnježđenja kao tri dodatna deskriptora. Stvorena su dva skupa latentnih vektora:

a) Preslika skupa ToxCast (N = 1018) na latentni prostor skupa Tox21 (UMAP ToxCast);
b) Preslika skupa molekula Sava (N = 357) na latentni prostor skupa Tox21 (UMAP Sava), i dobivena tri deskriptora korištena su (zajedno s drugim deskriptorima na kojima je model razvijen i optimiran) za predviđanje toksičnosti svih krajnjih točaka mjerenim na embrijima zebrice baze ToxCast. Skup Sava sadrži 358 spojeva, ali na Flusilazolu programski paketi zbog Si atoma ne mogu vršiti računalne operacije.

§ 5. REZULTATI I RASPRAVA

5.1 Rezultati kemijske analize uzoraka iz rijeke Save

5.1.1 Sediment i površinska voda

Prisustvo 429 KORZ, za koje su postojali kromatografski standardi, ispitano je u uzorcima sedimenata i vode rijeke Save. Otkriveno je prisustvo njih 313 (u koncentracijama iznad minimalnog praga detekcije) i kvantificirano na jednom ili više mjesta uzorkovanja (Elektronički dodatak E2). Analizirani su spojevi klasificirani u tri glavne kategorije prema prijedlogu u radu [132]: pesticidi, industrijske kemikalije i FADM. Otkriveni FADM uključuju analgetike, antibiotike, antikolesteremike, antidepresive, antiepileptike, kardiovaskularne lijekove, hipnotike/ antikonvulzive/ anestetike, opioide, halucinogene/stimulanse. Pesticidi se dijele na fungicide, herbicide i insekticide (tablica 11).

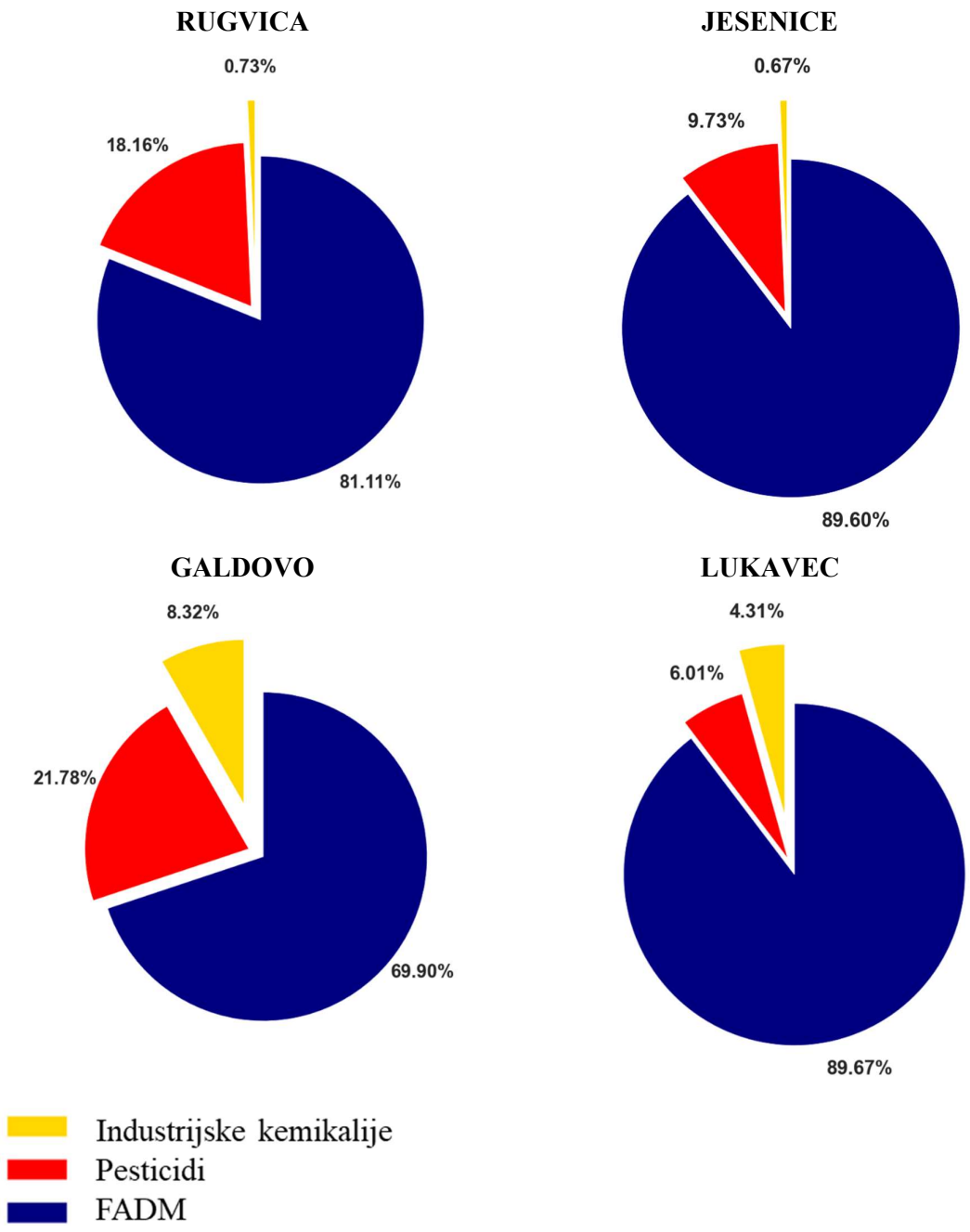
Tablica 11. Brojnost KORZu ekstraktima sedimenta (f – fungicidi, h – herbicidi, i – insekticidi).

	Jesenice	**Rugvica**	**Galdovo**	**Lukavec**
Pesticidi	109 (f:31, h:43, i:34)	132 (f:33, h:55, i:43)	65 (f:20, h:27, i:16)	73 (f:14, h:31, i:26)
FADM	101	109	74	89
Industrijske kemikalije	2	2	2	2
Ukupno	212	243	141	164

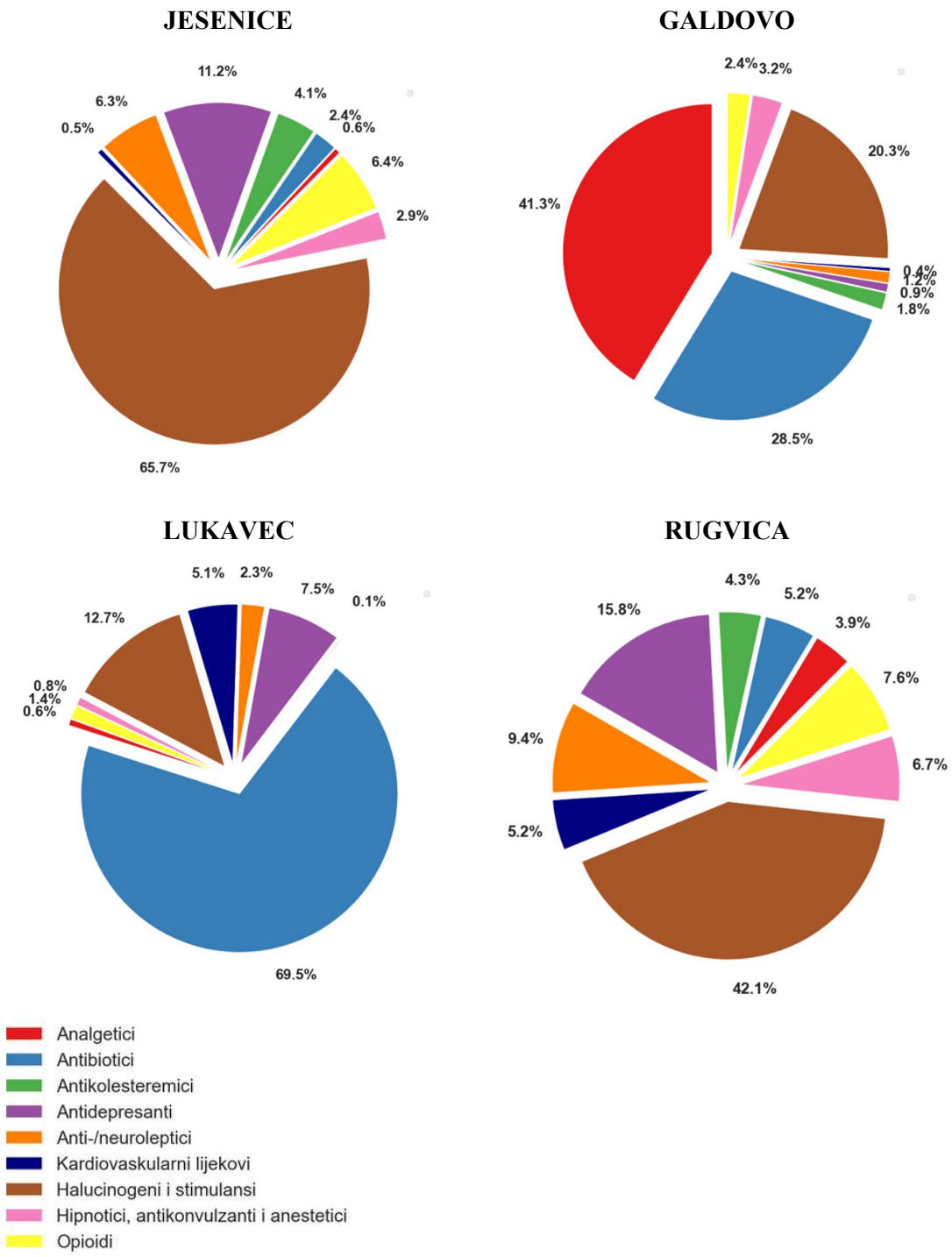
Među kvantificiranim su spojevima 132 različita FADM, 179 pesticida i dvije industrijske kemikalije (tj. 1H-benzotriazol i perfluoroktanoična kiselina - PFOA). Najčešće detektirani FADM su analgetici (3 od 4), antidepresivi (15 od 22), antibiotici (5 od 10), antiseptici/dezinficijensi (1 od 2), kardiovaskularni lijekovi (3 od 7) i halucinogeni/stimulansi (8 od 19). Ostali spojevi imaju znatno niži omjer otkrivenih predstavnika u odnosu na ukupni broj raspoloživih kromatografskih standarda među 429 KORZ. Ukupno su 73 spoja bila prisutna u svakom od uzoraka sedimenta. Na temelju 313 kvantificiranih spojeva, koncentracija zbroja u uzorcima sedimenta pokazala je silazni trend od 15 222,8 ng/g spojeva na lokaciji

Rugvica, 10 048,6 ng/g u Jesenicama i 9130,5 ng/g u Lukavcu, nakon čega slijedi koncentracija od 1247,6 ng/g na mjestu Galdovo.

Na svim su mjestima uzorkovanja dominantna kategorija (od spomenutih) bili FADM od 69,90 % (mjesto Galdovo) do 89,67 % (mjesto Lukavec) ukupne koncentracije spojeva po lokaciji. Na mjestima su Jesenice i Rugvica najzastupljenije podskupine FADM bili halucinogeni/stimulansi, odnosno 65,7 % i 42,1 %, najviše zbog koncentracija kofeina (3628,23 ng/g i 2085,08 ng/g) i kotinina, nikotinskog metabolita (2218,99 ng/g i 2964,36 ng/g). Najveće su koncentracije, gotovo polovica svih otkrivenih spojeva, mjerene na Rugvici. U Galdovu su najizraženiji analgetici (41,3 %), uglavnom zbog visoke koncentracije ibuprofena (350,6 ng/g), dok su antibiotici dominirali na mjestu Lukavec (69,5 %), gdje je azitromicin imao najvišu mjerenu koncentraciju (4813 ng/g). Pojava azitromicina u rijeci Savi od prije je poznata. [43] Na slikama 18 i 19 prikazane su raspodjele prema grupama i podgrupama. Neki od spojeva koji su u velikoj mjeri pridonijeli opterećenju bili su (abecednim redom): alfa-hidroksitriazolam (hipnotik), citalopram (antidepresiv), diaveridin (antibiotik) dietiltoluamid (pesticid/insekticid), gemfibrozil (antikoleresterem), ibuprofen (analgetik) i norpropoksifen (opioid), primidon (antiepileptik), tebutiuron (herbicid) i trimipramin (antidepresiv) na mjestima Jesenice i Rugvica. Na manje zagađenim mjestima Galdovo i Lukavec nekoliko spojeva s višim koncentracijama uključuju: ciprofloksacin (antibiotik), glifosat (herbicid), ibuprofen (analgetik), sulfametazin (antibiotik), venlafaksin (antidepresiv), verapamil (kardiovaskularni lijek) i 1H-benzotriazol.



Slika 18. Maseni udjeli podgrupa po mjestima nalaska u ukupnoj masi kvantificiranih KORZ. Slika preuzeta iz rada [33].



Slika 19. Maseni udjeli podgrupa po mjestima nalaska u ukupnoj masi kvantificiranih FADM. Slika preuzeta iz rada [33].

Usporedba lokacija na temelju koncentracija FADM i pesticida prikazana je u tablici 12. Koncentracije KORZ s pojedinih lokacija uspoređene su Spearmanovim koeficijentom korelacije. Rezultati pokazuju višu pozitivnu korelaciju između lokacija Jesenice i Rugvica od 64 % te pozitivnu korelaciju od 38 % između lokacija Galdovo i Lukavec. Sve ostale korelacije ne prelaze vrijednost 0,20.

Tablica 12. Spearmanova korelacija između koncentracija KORZ kvantificiranih na četiri lokacije rijeke Save.

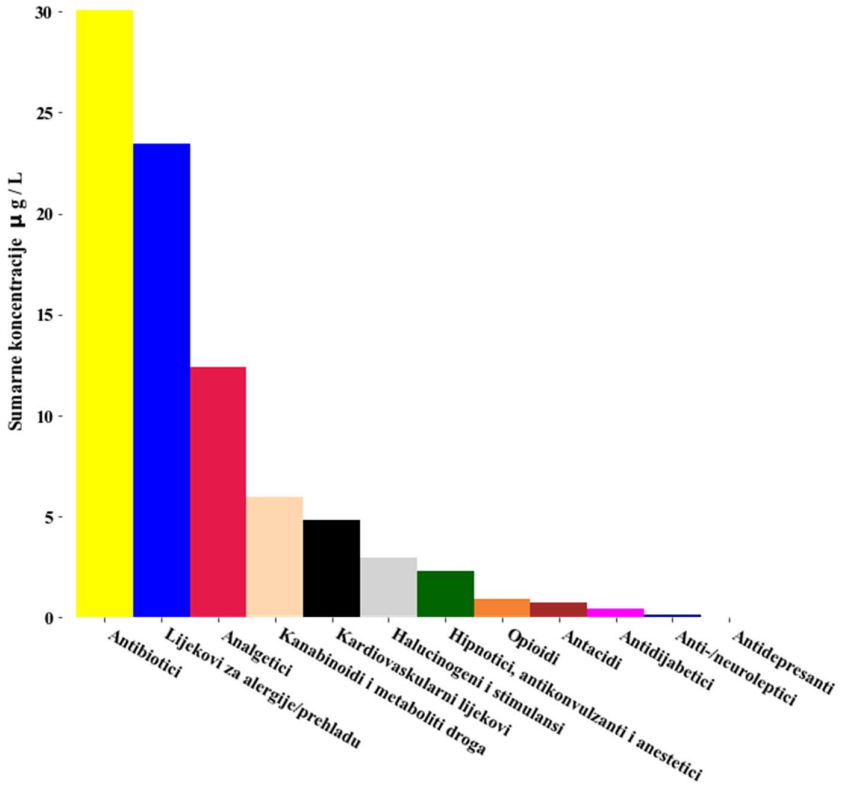
	JESENICE	**RUGVICA**	**GALDOVO**	**LUKAVEC**
JESENICE	-	***0,64**	0,15	0,20
RUGVICA	***0,64**	-	0,10	0,20
GALDOVO	0,15	0,10	-	***0,38**
LUKAVEC	0,20	0,20	**-**	-

Ovakav je odnos korelacija očekivan s obzirom na to da su Jesenice i Rugvica urbanizirane regije s relativnom blizinom farmaceutske industrije. Galdovo i Lukavec su manje urbanizirana mjesta koja se nalaze u okolici grada Siska, i više su pod utjecajem poljoprivrednog sektora i naftne industrije.

5.1.2 Riblja plazma i površinska voda

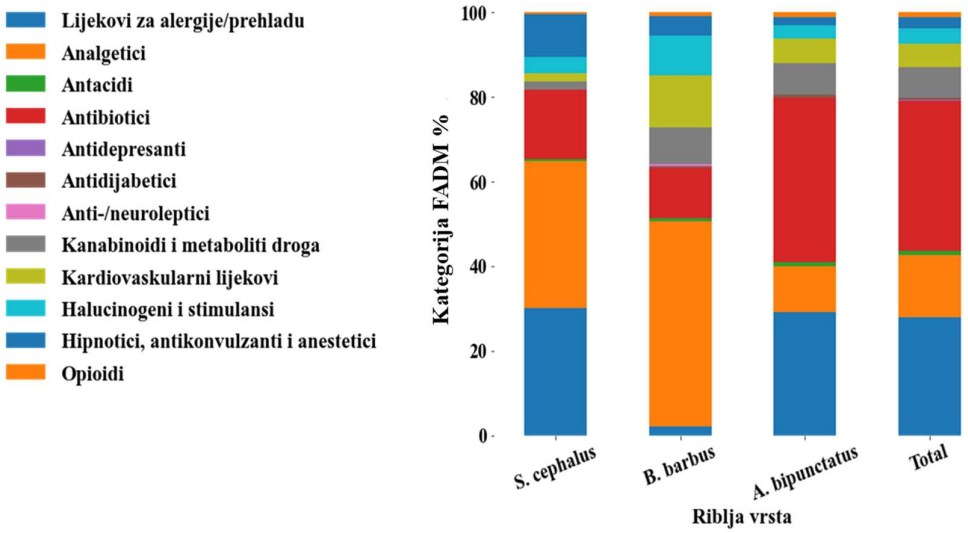
U ovom radu provedena je kvalifikacija i kvantifikacija FADM u ribljoj plazmi i površinskoj vodi na mjestima ulova ribe. S pomoću masene spektrometrije potvrđena je prisutnost 90 spojeva koji spadaju u kategoriju FADM. Među otkrivenim i kvantificiranim FADM bilo je 48 psihoaktivnih kemijskih spojeva, podijeljenih po sljedećim skupinama: 10 droga i metabolita droga (7 kanabinoida), 15 antibiotika, 9 analgetika, 8 kardiovaskularnih lijekova, 7 protuupalnih lijekova, dva antacida i jedan antidijabetik. Zbirni prikaz spojeva po kategorijama nalazi se u slici 20. Skoro svi analizirani FADM (87/90) detektirani su u plazmi klena, a zatim mrene (60/90) i uklije (54/90) što je razumljivo s obzirom na broj analiziranih jedinki svake vrste. Iz usporedbe nađenih spojeva u plazmi razvidno je da od analiziranih 90 FADM, 45 je pronađeno u površinskoj vodi (na mjestima ulova ribe) i plazmi ribe, dok je 45 bilo prisutno isključivo u ribljoj plazmi i nisu otkriveni u vodi ni na jednom mjestu uzorkovanja. Sveukupno je zbrojena koncentracija bila niža u uzorkovanoj riječnoj vodi nego u uzorcima plazme. Međutim, antibiotici azitromicin i penicilin, antidijabetički metformin, kao i analgetski

propoksifen te nikotin bili su prisutni u sličnim ili malo višim koncentracijama u vodi u usporedbi s koncentracijama izmjerenima u plazmi.



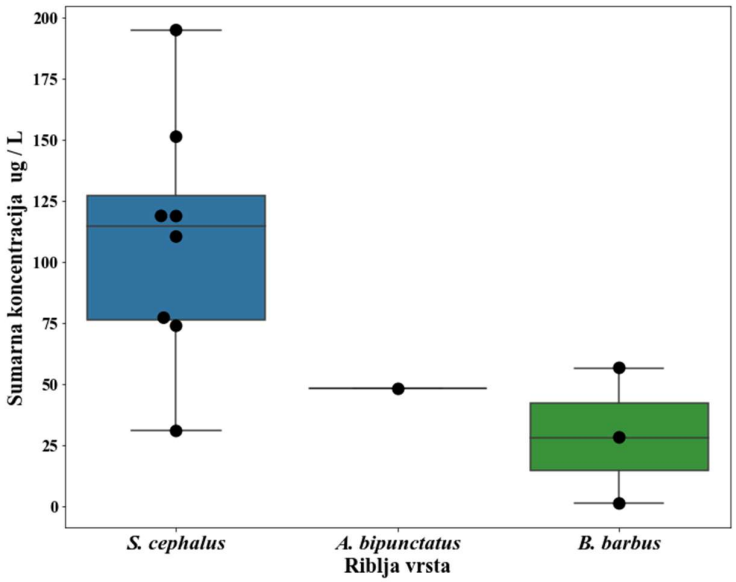
Slika 20. Stupčani dijagram raspodjele grupa spojeva nađenih u ribljoj plazmi.

Omjeri koncentracija svih FADM po kategorijama prikazani su u slici 21 prema ribljoj vrsti.



Slika 21. Trakasti dijagram raspodjele FADM prema ribljoj vrsti.

Ispitana je i razlika među vrstama ANOVA testom. ANOVA nije pokazala statistički značajne razlike u koncentracijama u plazmi među vrstama (klen i mrena; p* = 0,11) na uzvodnom mjestu (Podsused). Jedan je skupljeni uzorak dvoprugaste uklije (uzorak 2) isključen iz testa ANOVA zbog znatno niže kumulativne koncentracije. Koncentracije po vrstama prikazane su pravokutnim dijagramom u slici 22. Iznesena kvantitativna kemijska analiza pokazala je da plazma ribe iz rijeke Save akumulira brojne FADM s najvišim koncentracijama spojeva iz kategorija antibiotika, kortikosteroida i analgetika. Izmjerene koncentracije spojeva veće su (do 1000 puta) u plazmi nego u riječnoj vodi. Rezultati ove studije potkrepljuju prethodna istraživanja FADM u sedimentu rijeke Save (Poglavlje 5.1.1), a također prate svjetski trend u kojem su analgetici i antibiotici najzastupljeniji FADM u slatkovodnim ekosustavima (Hughes i sur., 2013).



Slika 22. Pravokutni dijagram sumarnih koncentracija organskih spojeva u plazmi analiziranih ribljih vrsta.

Konzumiranje FADM ima sezonski obrazac, s većim opterećenjem antibiotika u rijekama tijekom zime [43] te protuupalnim lijekovima tijekom ranog proljeća. [175] Pojedinačni spojevi unutar kategorije kortikosteroida s najvišom razinom u plazmi ribe bili su prednizolon (do 162,5 µg/L) i deksametazon (do 5,5 µg/L). Među analgeticima, acetaminofen je imao najveću otkrivenu koncentraciju u plazmi (do 28,7 µg/L), koja je približno šest puta veća od najveće koncentracije prethodno izmjerene u plazmi ribe. [70] Podaci pokazuju i da su koncentracije antibiotika u ribljoj plazmi zabrinjavajućih razina. Antibiotici se koriste u velikim količinama u terapeutske svrhe kod ljudi i životinja. Uzorci plazmi ribe u Savi akumulirali su uglavnom

trimetoprim i makrolidne antibiotike. To je uvelike povezano s visokim koncentracijama makrolidnog antibiotika azitromicina u rijeci Savi, koje uglavnom potječu iz postrojenja za proizvodnju antibiotika u sjevernoj Hrvatskoj. [43] FADM su dominantno polarni spojevi i umjereno hidrofobni, što dovodi do lakše cirkulacije u vodenoj tjelesnoj tekućini u odnosu na nakupljanje istih u organima s visokim sadržajem lipida. [72] U ovoj su studiji utvrđene i visoke koncentracije psihoaktivnih spojeva u plazmi, poput kofeina (1,44 μg/L), antiepileptičkog karbamazepina (0,17 μg/L), opioidnog buprenorfina (2,38 μg/L) i stimulansa metilfenidata (0,19 μg/L).

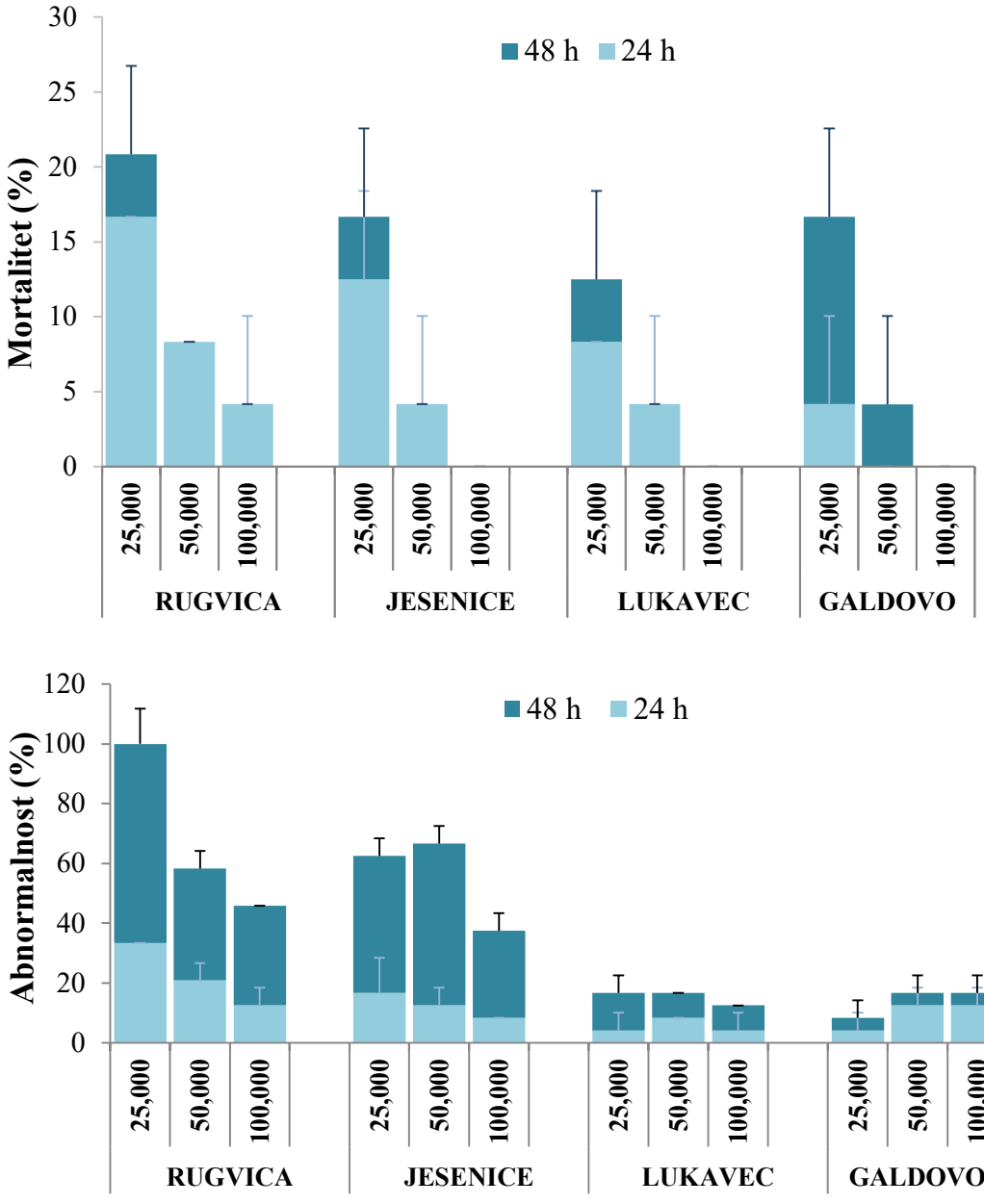
5.2 Rezultati prioritizacije spojeva u riječnom sedimentu

U ovom je radu prioritizacija spojeva u okolišu (sediment i površinska voda) napravljena trima metodama (TU_{pw}, TU_{zet} i PBTr) koje su opisane u poglavlju 4.2. Rezultati triju metoda opisani su u narednim poglavljima.

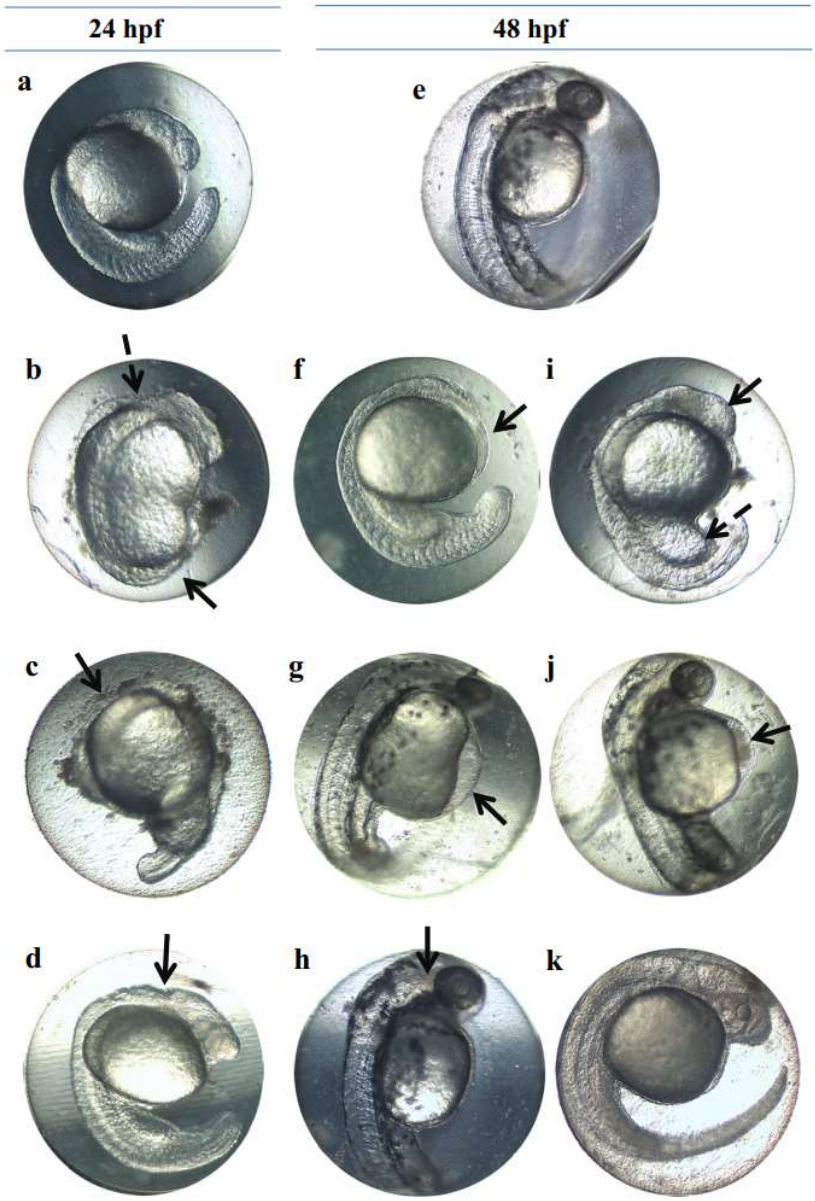
5.2.1 Rezultati testova embriotoksičnosti na zebricama

Tijekom izlaganja embrija *D. rerio* ekstraktima riječnog sedimenta u tri razrjeđenja (25 000x, 50 000x, 100 000x) zabilježen je negativan utjecaj na njihov razvitak, što se očitovalo povišenim mortalitetom i pojavom razvojnih abnormalnosti (RA) (slika 23, a i b). Povećanje mortaliteta prati povećanje ukupne koncentracije spojeva te vremena izlaganja embrija ekstraktima sedimenta. Nakon 24 h izlaganja, najveći je mortalitet zabilježen tijekom izloženosti najvišoj koncentraciji (razrjeđenje 25 000x) sedimentnih ekstrakata Rugvice (16,7 %) i Jesenica (12,5 %). Unatoč niskom mortalitetu zabilježenom nakon 24 h izlaganja, 48-satno izlaganje na postaji Galdovo rezultiralo je porastom mortaliteta na 12,5 %. Mortalitet je unutar kontrolne skupine bio ispod 5 %. Učestalost pojave abnormalnosti među embrijima (slika 23 b) rasla je s porastom koncentracije ekstrakta sedimenta i razvojnom progresijom, sukladno i razinom smrtnosti. Embriji koji su preživjeli izloženost najvišoj ispitnoj koncentraciji sedimentnih ekstrakata pokazali su najveću abnormalnost, što je istaknuto u slučaju Rugvice (33,3 % nakon 24 hpf i 66,7 % nakon 48 hpf). Povećanje razvojnih abnormalnosti tijekom trajanja eksperimenta bilo je najizraženije kod embrija nakon 48 hpf izloženosti uzorku iz Jesenica u usporedbi s vrijednostima od 24 hpf. Najčešće zabilježene morfološke abnormalnosti prikazane su na slici 24. Embriji zebrica u kontrolnoj skupini razvijali su se normalno, a tijekom eksperimenta nisu zabilježene razvojne abnormalnosti. Razvojne su abnormalnosti zbrojene prema mjestima uzorkovanja sedimenta. Tako je Rugvica (R) pokazala najveći broj razvojnih abnormalnosti (26), zatim Jesenice (J) (19) pa Lukavec (L)

(16) i Galdovo (G) (13). Rezultati su podrobno prikazani u tablici 3 u objavljenom radu [33]. Stoga, ukupni brojevi Razvojnih abnormalnosti (RA) u biotestu prema mjestu uzorkovanja sedimenta pokazuju ovaj poredak R > J > L > G. Mortalitet embrija zebrice mjeren kroz 24 h pri najvišim koncentracijama (najnižim razrjeđenjima, 25 000 x) pokazuju sljedeći poredak prema mjestu uzorkovanja R > J > L > G. Za mjerenja provedena kroz 48 h, poredak je R > G > J > L.



Slika 23. Prikaz a) mortaliteta i b) abnormalnost embrija *D. rerio* nakon 24-satnog i 48-satnog izlaganja ekstraktima sedimenta rijeke Save. Rezultati su prikazani kao srednja vrijednost ± SD (preuzeto iz rada [33], zarez na osima označava razmak).



Slika 24. Prikaz razvojnih abnormalnosti nakon izlaganja ekstraktima: a) i e) normalno razvijeni kontrolni embriji nakon 24 i 48 h; b) 24 h - razvojna deformacija, neodvajanje repa od žumanjčane vreće (strelica), deformacija u području glave (isprekidana strelica); c) 24 h - deformacija cijelog embrija, nerazvijeno područje glave (strelica); d) 24 h – deformacija u području glave (strelica); f) 48 h – nedostatak pigmentacije i nerazvijeno područje glave (strelica); g) edem u području žumanjčane vrećice (strelica); h) 48 h – nakupljanje krvi u području glave (strelica); i) 48 h -razvojne deformacije, nerazvijene oči (strelica), nedostatak

pigmentacije; j) 48 h - nakupljanje krvi u području žumanjčane vrećice (strelica); k) 48 h - nedostatak pigmentacije. Slika preuzeta iz rada [33].

Rezultati histopatoloških analiza objavljeni su u tablici 4 u radu [33]. Ovdje su kumulativno prikazani zbog kasnije usporedbe s računalnim metodama za procjenu rizika (tablica 13).

Tablica 13. Zbroj histopatoloških promjena u biotestovima prema mjestima uzorkovanja sedimenta.

	Jesenice	**Rugvica**	**Lukavec**	**Galdovo**
Poremećaji optjecajnog sustava (POS) (edemi, nakupljanje krvi)	4	5	20	26
Oštećenja tkiva (OT)	16	27	13	12
Razvojne deformacije tkiva (RDT) (oko, mozak)	15	25	11	9

Iz rezultata histopatoloških analiza danih u tablici 13 razvidno je da za zbroj poremećaja **POS** vrijedi ovaj poredak prema mjestu uzorkovanja G > L > R > J, za zbroj oštećenja **OT** poredak je R > J > L > G, dok je za razvojne deformacije tkiva **RDT** poredak R > J > L > G.

5.2.2 Prioritizacija uzoraka sedimentu prema procijenjenoj toksičnosti: metoda TU

Toksičnost uzoraka sedimenata izračunata je s pomoću TU_{pw} pristupa opisanog u poglavlju 4.2.1 (jednadžba 4.5). TU_{pw} vrijednosti zbrojene su za sve kvantificirane spojeve na pojedinim lokacijama. Tablica izračunatih TU_{pw} vrijednosti za sve spojeve nalazi se u elektroničkom dodatku E6. Zbroj vrijednosti TU_{pw} prema lokacijama uzorkovanja daje sljedeći redoslijed: G > R > L > J (3,63> 3,62> 3,39> 3,03). Samo je šest spojeva (KORZ) na svim lokacijama imalo logTU_{pw} iznad 0, dakle za veći dio kumulativnih lokacija na mjestu: beta glukuronid (J, R, L), ciprofloksacin (G, L), glifosat (J, R, G, L) glukuronid (R), ritalinska kiselina (S), sulfametazin (L).

Metoda TU_{pw} izračunava toksične jedinice na temelju ravnoteže sedimenta i vodenog stupa. Stoga se očekuje da će se u obzir uzeti uglavnom spojevi koji lako difundiraju u vodu. To je ograničavajući čimbenik ove metode i proizlazi iz definicije metode koja koristi particijski koeficijent spojeva (Poglavlje 4.2.1). Stoga TU_{pw} i dalje ne pruža relevantne informacije o opasnim učincima ekstrakta sedimenta za ispitivanje *in vivo* embriotoksičnosti. Zbog

spomenutih je nedostataka izračunata i TU vrijednost za KORZ u ekstraktu sedimenta, definirana kao TU_{zet} (Poglavlje 4.2), koja uzima u obzir koncentraciju ekstrakta KORZ iz kojeg su pripremljene otopine za testove embriotoksičnosti (ZET). Tablica izračunatih TU_{zet} vrijednosti nalazi se u elektroničkom dodatku E6.

Rezultati pokazuju da samo amiodaron ima logTU_{zet} vrijednost iznad 0 (J, R, L) %. Redoslijed TU_{zet} za ekstrakte sedimenta je kako slijedi R > J > L > G (1,43 > 0,77 > 0,65 > -1,5). Izračunate vrijednosti *TU* uspoređene su s rezultatima ZET Zbroj svih zabilježenih razvojnih abnormalnosti za sva razrjeđenja s uzorkom određene lokacije ima isti gradijent zagađenja kao i za *TU* i kumulativnu koncentraciju KORZ (R > J > L > G; 26> 19> 16> 13).

5.2.3 Prioritizacija na sedimentu: metoda PBTr

Iz tablice PBT rangova (PBTr) izdvojeno je za svako mjesto uzorkovanja prvih 20 prioritetnih spojeva za daljnju analizu (tablica 14).

Tablica 14. Lista izabranih 20 prioritetnih spojeva poredanih prema abecednom redoslijedu.

Jesenice	Rugvica	Galdovo	Lukavec
α-Hidroksi Triazolam	10-hidroksikarbazepin	Ciprofloksacin	amiodaron
alprazolam	alfa-hidroksi triazolam	CP 47,497	fenciklidin
amiodaron	alprazolam	CP 47,497-C8 homolog	fenitoin
butokarboksim	amiodaron	desalkilflurazepam	flonikamid
diaveridin	amitriptilin	desmedifam	karbamazepin
dietofenkarb	butokarboksim	diklofenak	karbosulfan
dinoseb	diaveridin	heksakonazol	klomipramin
fenpiroksimat	fentanil	imazalil	kvetiapin
fentanil	fluvoksamin	karbamazepin	lamotrigin
flumioksazin	ipkonazol	klomipramin	maprotilin
klomipramin	klomipramin	lamotrigin	norbuprenorfin
lorazepam	lorazepam	o-desmetiltramadol	o-desmetiltramadol
metkonazol	norbuprenorfin	o-desmetilvenlafaksin	o-desmetilvenlafaksin
norpropoksifen	norpropoksifen	PFOA	oksazepam
pentazokin	pentazokin	propoksifen	PFOA
PFOA	PFOA	prosulfokarb	propoksifen
propikonazol	sertralin	sulfametoksazol	sertralin
sertralin	tiabendazol	triadimefon	spiroksamin
sulfentrazon	triazolam	triklozan	venlafaksin

tiabendazol	trimipramin	tritikonazol	zolpidem

Rezultati ukazuju na nekoliko spojeva koji su visoko rangirani na svim lokacijama, tj. PFOA, amiodaron, sertalin i klomipramin, i svi oni imaju visoke log*P* vrijednosti (redno: 3,81; 7,8; 5,07; 5,19). Daljnja analiza za navedena četiri spoja s višestrukim pojavama u PBT rangiranju vodi ka zaključku da koncentracija nije jedini čimbenik koji utječe na rizik, nego i njihova hidrofobnost, pokazujući stoga različite rezultate u odnosu na *TU* metode. Pojedini spojevi pokazuju visok rizik prema obje metode (*TU* i PBTr), kao što su amiodaron i ciprofloksacin. Razlog tome su iznimno visoke vrijednosti toksičnosti ta dva spoja. Niz ostalih spojeva koji su prema metodi PBTr rangirani među prvih 20 pokazuju da je potrebno razmatrati sve dostupne metode za procjenu rizika.

5.3 Rezultati prioritizacija spojeva u ribljoj plazmi

Spojevi u uzorcima plazme u ovom radu prioritizirani su modelom riblje plazme (MRP), tj. usporedbom s humanom terapijskom koncentracijom u plazmi, a taj postupak opisan je u Poglavljima 2.2.2 i 4.3. MRP se temelji na hipotezama o konzerviranosti bioloških ciljeva/ciljanih proteina [65,68], tj. da će biološki aktivni spojevi stoga vrlo vjerojatno pokazati slične biološke učinke na ribama i ljudima pri minimalnoj djelotvornoj koncentracije lijeka koja se postiže u plazmi. Za one spojeve koji djeluju na iste ciljne proteine mogu se očekivati aditivni ili sinergijski učinci. Kod sinergijskog učinka doprinos pojedinih spojeva ukupnoj toksičnosti veća je od zbroja njihovih pojedinačnih toksičnosti. Međutim, prioritizacije toksičnosti smjesa u disertaciji temeljile su se na aditivnom modelu toksičnih učinaka pojedinih spojeva čije je prisustvo utvrđeno u smjesi, što je i dominantni toksikološki model. Na primjer, devet opioida otkrivenih u ovoj studiji (buprenorfin, oksimorfon, oksikodon i hidrokodon) ciljaju iste opioidne receptore u organizmu. Opioidni je sustav, koji se sastoji od ova tri G-proteinski povezana receptora, od centralnog značaja u ponašanju za bol, nagradu i ovisnost. [176] Za većinu opioida otkrivenih u plazmi ribe iz rijeke Save vrlo je vjerojatna interakcija (ali ne nužno jedina) s mutijskim opioidnim receptorom (MOR) kao primarnim ciljem. [177] Pretpostavka je da se MOR-ovi kontinuirano aktiviraju pod kontinuiranim izlaganjem opioidima koji izazivaju molekularne modifikacije unutar opioidnog sustava i prilagodbe signalnih receptora, koje utječu na ponašanje organizama i izazivaju ovisnost. [178] Može se dakle pretpostaviti da i drugi FADM u riječnoj vodi ciljaju iste proteine i mogu također aditivno doprinijeti ukupnom štetnom utjecaju na ribe.

Za računanje MRP uzete su humane terapijske vrijednosti (HTKP) kvantificiranih spojeva u ribljoj plazmi. Za dva spoja, THCA-A i metil ester ekgonin, te vrijednosti nisu pronađene u dostupnoj literaturi. Izračunate su terapijske vrijednosti za riblju plazmu kako bi se odredili omjeri učinka (ER) prema opisu u Poglavlju 4.3.1. ER vrijednosti podijeljene su u pet kategorija prema veličini: ER < 1, ER 1 - 10, ER 10 - 100, ER 100 - 1000, ER > 1000. Sve izračunate ER vrijednosti navedene su u tablici 15. Pojedini kemijski spojevi analizirani u ribljoj plazmi - kao što su kotinin, buprenorfin, nikotin, azitromicin i oksimorfon - imali su ER vrijednosti manje od 1, što upućuje na to da su djelovali u koncentraciji koja izaziva učinak. U kategoriji spojeva s ER manjim od 10 nađena su dva kortikosteroida (prednizolon i deksametazon), jedan kardiovaskularni lijek (ramipril), kanabinoidi (CBD i THC) te jedan opioid (hidrokodon). Među visokorizičnim FADM može se utvrditi opći trend koji diskriminira visoko zastupljenu klasu tvari koje djeluju prvenstveno na središnji živčani sustav - psihoaktivne farmaceutike [179] od sustavno nižih vrijednosti ER (ER < 1000; N = 27/50). Izračunate ER vrijednosti ukazuju na ukupno 50 spojeva potencijalno opasnih za ribu, od čega je 27 psihoaktivnih spojeva, šest protuupalnih lijekova, šest antibiotika, pet kardiovaskularnih lijekova, dva antacida, dva analgetika i jedan antidijabetik. Halucinogeni i stimulansi, opioidi, antiepileptici i neuroleptici te kanabinoidi i metaboliti droga podskupine su koje pokazuju najveći omjer zabrinjavajućih vrijednosti prema ER. Opioidi pronađeni u ribama iz rijeke Save temelje se na 4,5- epoksimorfinskom prstenu i podvrgnuti su procesu O-dealkilacije. To je put kojim se hidrokodon metabolizira u hidromorfon, a oksikodon u oksimorfon. [180] Treba naglasiti da metaboliti mogu biti toksičniji od svojih matičnih spojeva. [177] Sva su četiri opioida otkrivena u plazmi ribe, a ER pristup uvrštava THC i CBD među prioritetne FADM u rijeci Savi (tablica 15). Svjetski su trendovi upotrebe kanabisa u velikoj mjeri stabilni, dok je u porastu razina THC-a kanabisa. [181] THC je glavni psihoaktivni spoj kanabisa, koji djeluje prvenstveno na kanabinoidni receptor 1 (CB1-R) izražen u CNS-u, ali djeluje i na opioidne i benzodiazepinske receptore, što indicira sinergistički učinak kanabinoida i opioida. [182] Embriji zebrice izloženi THC-u ili CBD-u pokazuju promjene srčanog ritma, morfologiju motoričkog neurona, lokomotorne reakcije, obrazac sinaptičke aktivnosti i ličinke. [183] Kronično mikrodoziranje riba s FADM, posebno u njihovim najosjetljivijim fazama (zametak i ličinke), može imati negativne učinke na njihov razvoj i ponašanje, smanjujući njihov reproduktivni uspjeh i mogućnost preživljavanja. [184]

Tablica 15. ER vrijednosti izračunate za FADM izmjerene u ribljoj plazmi.

FADM	ER					ukupno
	< 1	1 - 10	10 - 100	100 - 1000	> 1000	
kotinin	4	7				11
buprenorfin	2	7	1			10
nikotin	1	9	2			12
azitromicin	1	7	3	1		12
oksimorfon	1	2	4			7
CBD		5	4	2		11
prednisolon		2	2	4	4	12
ramipril		2	1			3
THC		2	1			3
deksametazon		1	1	5	5	12
hidrokodon		1	1	5		7
oksikodon			3	7		10
klaritromicin			3	4	3	10
kodein			2	4	4	10
trimetoprim			2	2	6	10
amlodipin			2	2		4
metformin			2	1	6	9
bisoprolol			2	1		3
fentermin			1	7	3	11
amfetamin			1	5		6
cimetidin			1	4	7	12
alprazolam			1	4		5
simvastatin			1	3	5	9
cetirizin			1	3	4	8
morfin			1	3	1	5
risperidon			1	3		4
loratadin			1	2		3
nalokson			1			1
metilfenidat				8	3	11
acetaminofen				7	4	11
kokaetilen				7	2	9
mdma				7	1	8
hidromorfon				6	4	10
prokainamid				5	7	12
eritromicin				5	6	11
fenilefrin				5	5	10
citalopram				4	2	6
CBN				4		4
olanzapin				4		4
dekstrometorfan				3	5	8
bupivakain				2	9	11
norfloksacin				2	8	10
pseudoefedrin				2	7	9
naltrekson				2	2	4
metamfetamin				2		2
propoksifen				1	10	11
ciprofloksacin				1	9	10
omeprazol				1	4	5
haloperidol				1		1
verapamil				1		1

5.4 Analiza uravnoteženosti skupa ToxCast

Analiza podataka izrađena je s pomoću skripte napisane u programskom jeziku Python. Cilj je bio prikazati broj toksičnih spojeva u svakom biološkom testu. Rezultat te pregledne analize prikazan je u tablici 16.

Tablica 16. Broj toksičnih i netoksičnih spojeva prema toksičnom učinku iz skupa ToxCast. Retci tablice ciljne su varijable poredane prema udjelu aktivnih spojeva (tj. broja spojeva koji su u mjerenjima pojedinog učinka ispoljili toksično djelovanje).

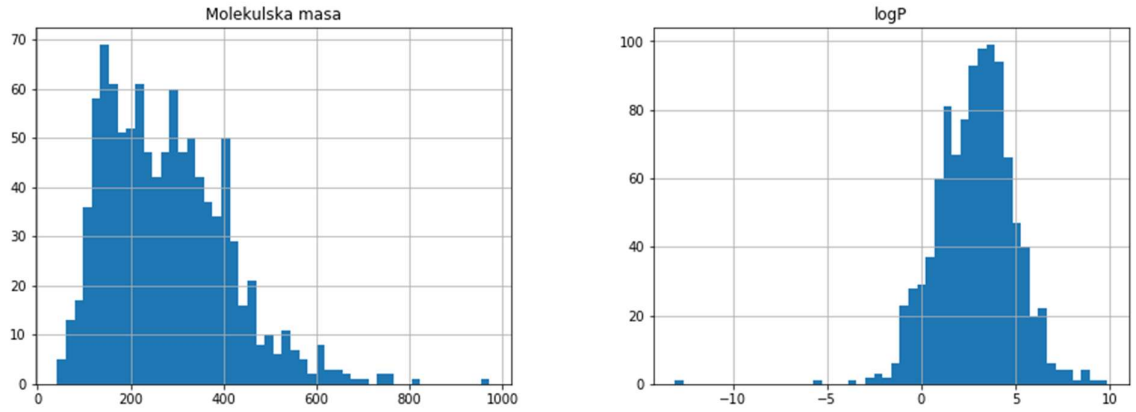
Toksični učinak	Neaktivnih (0)	Aktivnih (1)	Nedostaje vrijednost	% aktivnih
ActivityScore	812	187	19	18,7
YSE	867	123	28	12,4
PE	874	116	28	11,7
MORT	884	115	19	11,5
JAW	881	109	28	11,0
AXIS	882	108	28	10,9
SNOU	883	107	28	10,8
TR	912	78	28	7,9
EYE	913	77	28	7,8
BRAI	930	60	28	6,1
TRUN	934	56	28	5,7
PFIN	936	54	28	5,5
CFIN	942	48	28	4,8
PIG	945	45	28	4,5
OTIC	949	41	28	4,1
SOMI	952	38	28	3,8
SWIM	958	32	28	3,2
CIRC	972	18	28	1,8
NC	977	13	28	1,3

Toksičnosti iz tablice 16 pokazuju udio aktivnih spojeva kroz sve toksične učinke od 1,3 % (NC) do 18,7 % (ActivityScore). Stoga se može pristupiti modeliranju s pretpostavkom da se radi o neuravnoteženom klasifikacijskom problemu, čije je rješavanje razloženo u Poglavljima 2.5 i 2.7.

5.5 Usporedba skupova ToxCast, Sava i Tox21

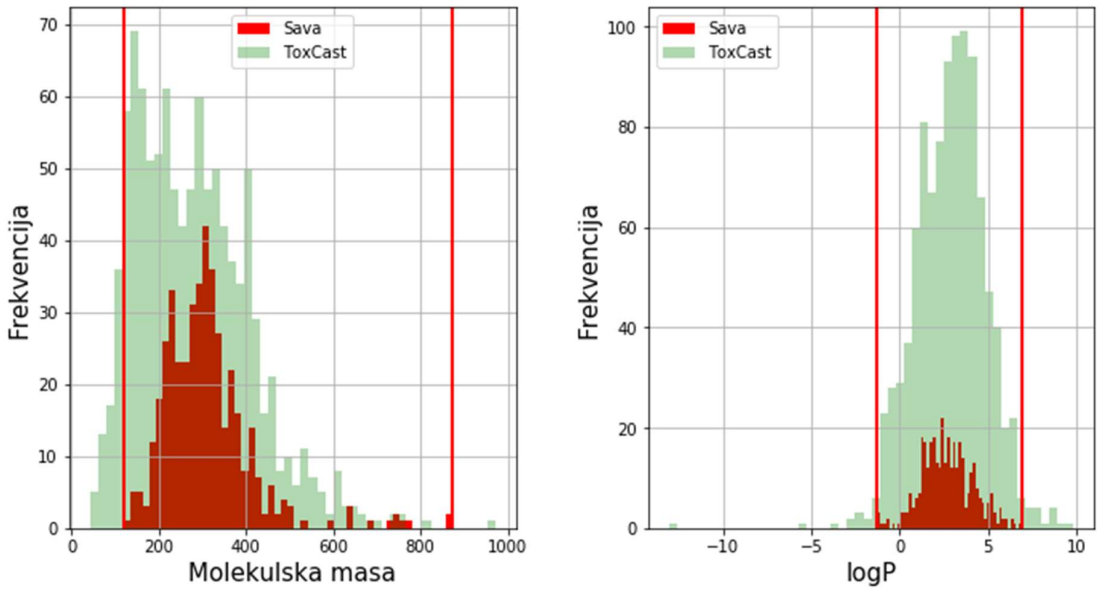
5.5.1 Usporedba kemijskih svojstava skupova Sava i ToxCast

Ovdje su deskriptivno analizirani skupovi ToxCast (N = 1018), Tox21 (N = 8144) i Sava (N = 357). Prije pripremanja modela korisno je razumjeti podatke s kojima se modelira. Važna su kemijska svojstva u ekotoksikologiji vrlo često relativna molekulska masa i log*P* zbog pretpostavke da imaju važnu ulogu u prolasku kroz membrane i ulasku u stanicu. Iz slike 25 vidi se da su spojevi izmjereni u rijeci Savi pretežito pozitivnog particijskog koeficijenta, s medijalnom vrijednošću log*P* = 2,62 te s 25 i 75 percentilnim vrijednostima log*P* = 1,73 i log*P* = 3,59. Te vrijednosti ukazuju na blagu hidrofobnost prisutnih spojeva, što je očekivano s obzirom da je većina njih pronađena u organizmu i/ili sedimentu (riječnom dnu). Hidrofobnost odnosno lipofilnost omogućuje spojevima prolazak kroz fosfolipidnu staničnu membranu i time utječe na njihovu toksičnost. Također, ti su spojevi pretežito nižih relativnih molekulskih masa (*Mr*) (medijan 300,01, a 25 i 75 percentilne vrijednosti su 240,31 i 348,68). Niže molekulske mase omogućuju lakši prolazak spojeva kroz stanične membrane i time doprinose njihovoj većoj toksičnosti. Najveća relativna masa nađena je među spojevima u skupu Sava i iznosi 872,4.



Slika 25. Raspodjele relativnih molekulskih masa i log*P* spojeva iz skupa Sava

Slika 26 daje usporedbu skupova Sava i ToxCast, koji služi kao osnova za razvoj QSAR modela u disertaciji. Usporedbe vrijednosti log*P* i relativnih molekulskih masa pokazuju da se kemijski prostori tih dvaju skupova dobro preklapaju.



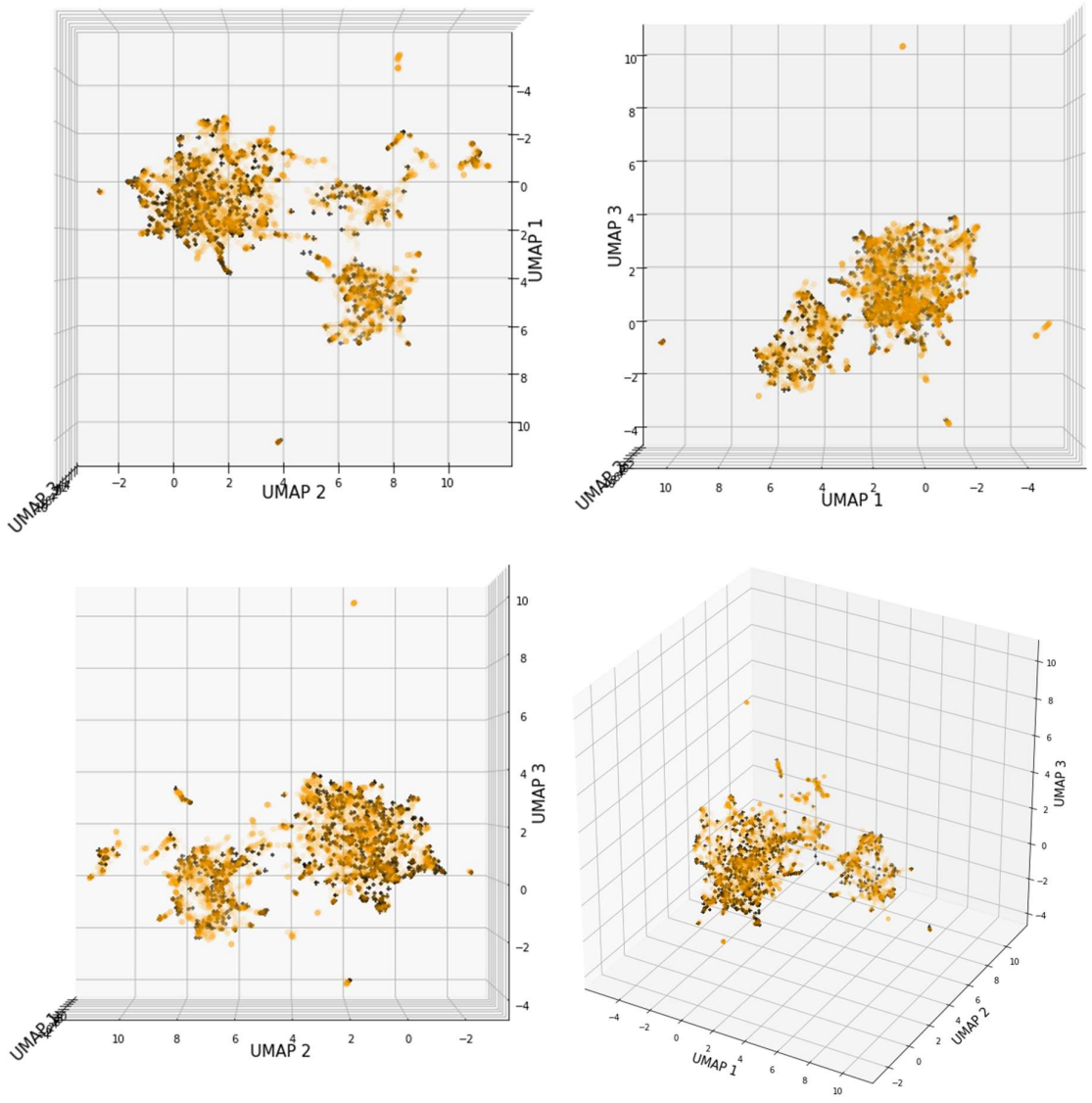
Slika 26. Usporedba raspodjela vrijednosti molekulskih masa (MM) i particijskog koeficijenta oktanol-voda (log*P*) molekula iz skupova Sava i ToxCast. Crvene vertikalne crte označavaju rubne vrijednosti Mr i log*P* za skup Sava

Medijan relativne molekulske mase skupa ToxCast je 263,42, s 25 i 75 percentilnim vrijednostima 170,09 i 36,102. Najveća relativna molekulska masa u skupu ToxCast je 972,31. Vrijednosti log*P* u skupu ToxCast imaju medijan 2,97 te 25 i 75 percentilne vrijednosti 1,51 i 4,15. Iz priloženih je vrijednosti razvidno da veći skup (ToxCast) zauzima i širi kemijski prostor, što ukazuje na to da bi eventualni QSAR modeli razvijeni na podacima iz baze ToxCast dobro pokrili kemijski prostor skupa Sava i predviđanja na tom skupu padala bi u domenu primjenjivosti.

5.5.2 Presjeci kemijskog prostora u latentnom prostoru (UMAP)

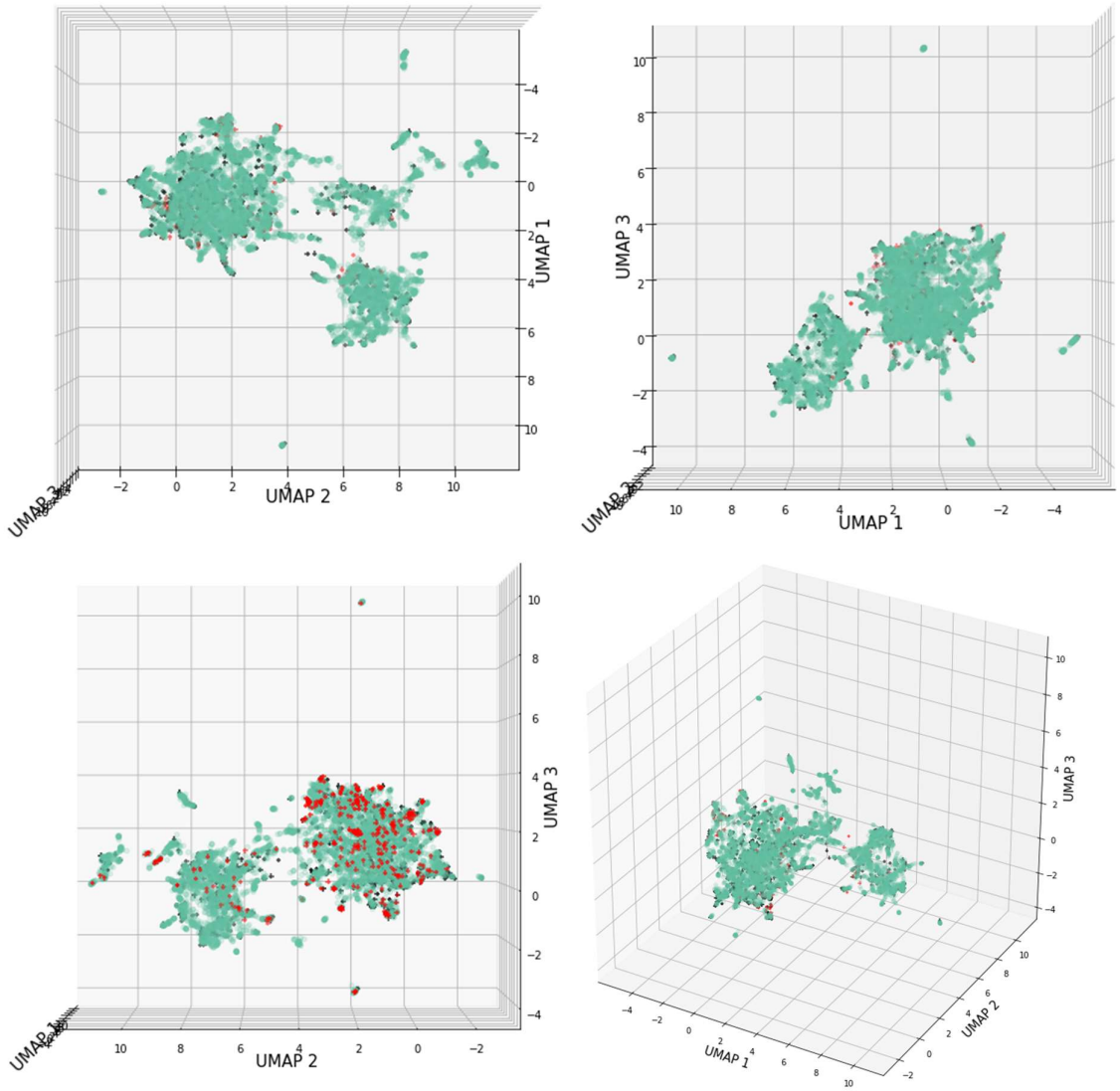
Tehnikom UMAP, koja je objašnjena u Poglavlju 2.9, nelinearno je transformiran kemijski prostor skupova Tox21 i ToxCast. Za tu je svrhu skup molekulskih otisaka Tox21 transformiran na temelju Jaccardove udaljenosti u 3D prostor. [185] Jaccardova je udaljenost mjera sličnosti koja se račun iz molekulskih otisaka kao binarnih podataka (s vrijednostima 0 ili 1 [153]). Isti je transformator potom primijenjen na molekulske otiske skupova ToxCast i Sava. Tako transformirani podaci prikazani su na slikama 27 - 29. U 3D prostoru zadržana je kemijska sličnost (najbliži susjedi) koja se putem struktura reflektira i u molekularnim otiscima. Sa slike 27 razvidno je da transformirani kemijski prostor skupova Tox21 (žuta boja) i ToxCast (crne oznake) pokazuju dobro preklapanje. Uočava se da nema kemijskih spojeva iz skupa ToxCast

koji izrazito odstupaju od spojeva iz skupa Tox21, što potvrđuje opravdanost korištenja prostora molekulskih otisaka skupa Tox21 za transformaciju skupa molekula iz skupa ToxCast.

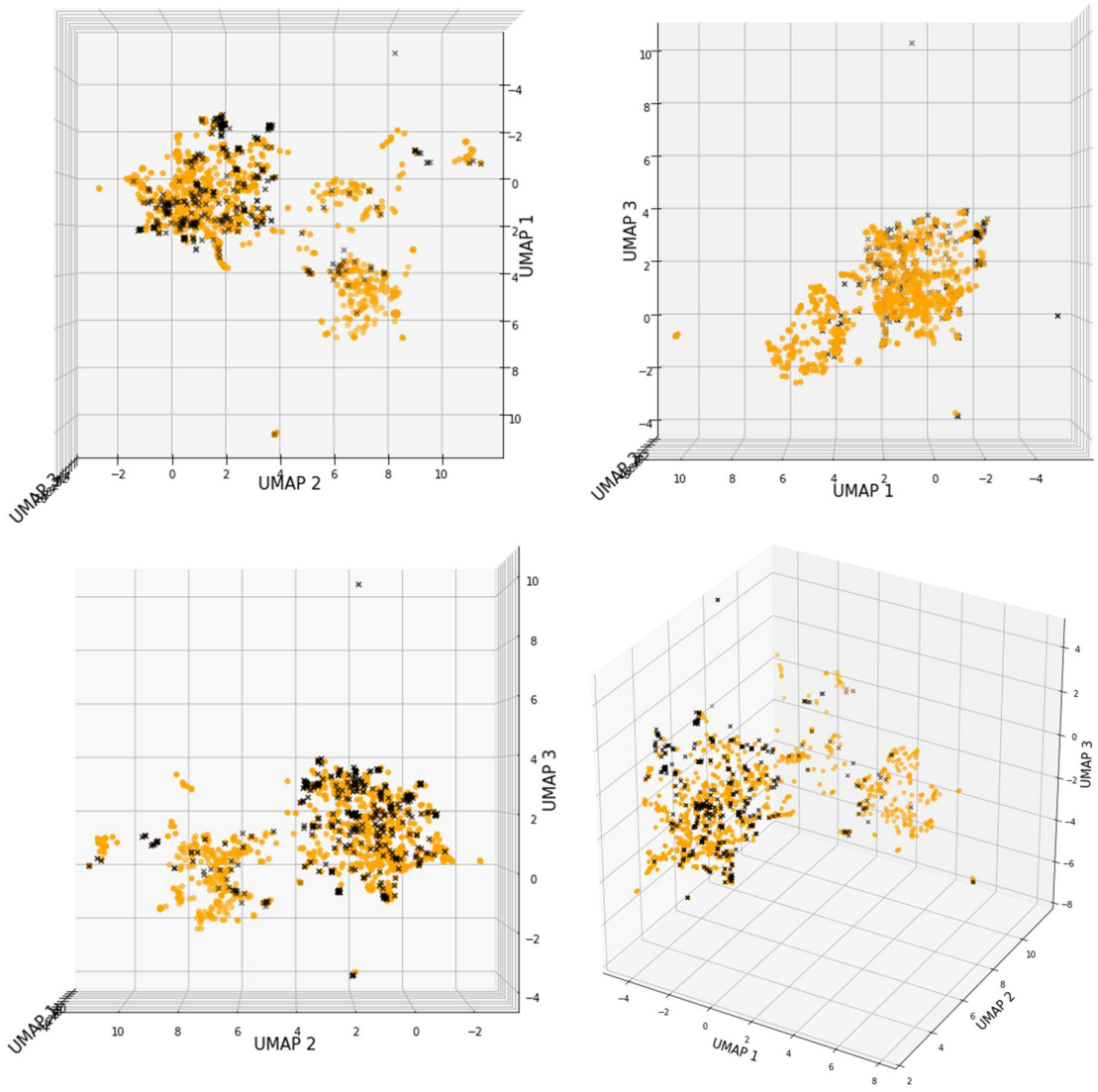


Slika 27. 3D prikaz presjeka kemijskog prostora skupova Tox21 (žuta boja) i ToxCast skupa (crna boja), koji su prethodno transformirani metodom UMAP. Uz 3D prikazani su i 2D presjeci prostora. UMAP1, UMAP2 i UMAP3 novi su vektori transformiranog kemijskog prostora.

Sa slike 28 na kojoj su prikazana sva tri skupa (Tox21, ToxCast, Sava) vidi se da volumen kemijskog prostora Tox21 (zelena boja) u potpunosti pokriva prostor druga dva spomenuta skupa (crna i crvena boja). Zaključno - transformirani su kemijski prostori (preko skupa Tox21) za skupove ToxCast i Sava prikazani na slici 29 iz koje se vidi da se ta dva skupa dobro preklapaju.



Slika 28. 3D prikaz presjeka kemijskog prostora skupa Tox21 (žuta boja) i skupova Sava i ToxCast (crna boja), koji su prethodno nelinearno transformirani metodom UMAP. Uz 3D prikazani su i 2D presjeci prostora. UMAP1, UMAP2 i UMAP3 novi su vektori transformiranog kemijskog prostora.



Slika 29. 3D i prikaz UMAP transformiranog kemijskog prostora presjeka skupova Tox21 (žuta boja), Sava skupa ToxCast skupa (crna boja). Uz 3D prikazani su i 2D presjeci prostora. UMAP1, UMAP2 i UMAP3 vektori su transformiranog kemijskog prostora.

5.5.3 Deskriptori izvedeni iz transformiranog prostora

Postupak dobivanja latentnih vektora za skupove ToxCast i Sava opisan je u Poglavlju 4.5. Nakon transformacije najvećeg skupa Tox21 (8144×3) postupkom UMAP, na isti su način transformirani i skupovi ToxCast (1018×3) i Sava (357×3). Dobivena su, dakle, dva skupa latentnih vektora (matrica ugnježđenja):

a) Preslika skupa ToxCast na latentni prostor skupa Tox21 (UMAP ToxCast)

b) Preslika skupa Sava na latentni prostor skupa Tox21 (UMAP Sava)

Latentni skupovi predstavljaju kemijska susjedstva iz skupova ToxCast i Sava na skupu Tox21, i u ovom će radu biti korišteni kao dodatni kemijski deskriptori.

5.6 Rezultati QSAR modela razvijenih u disertaciji

5.6.1 Računanje bioloških deskriptora na skupu molekula i svojstava baze Tox21

Skup Tox21 opisan u Poglavlju 3.3.2 korišten je za izradu QSAR modela. Skup Tox21 dobro je poznat te često korišten za ispitivanje metodoloških doprinosa u području modeliranja poput konformacijskog uzorkovanja i uravnotežavanja skupova. [172,186] Također, poznati su i dosezi QSAR modela razvijenih na skupu Tox21 i njihove sposobnosti (točnosti) predviđanja za različite biološke aktivnosti. [84,93,99,147,187] Analizom podataka u skupu Tox21 uočeni su nedostaci poput nedovoljno precizne detekcije, postojanje duplikata, kao i nedostatna predobrada struktura molekula iz tog skupa, a i pojedini literarni izvori svjedoče slično. [172] Uz biološku važnost, budući je skup neuravnotežen (znatno više spojeva je u klasi 0 nego u klasi 1), on će poslužiti i za usporedbu dosega metodologije QSAR modeliranja u ovom radu. Kompletni skup detaljno je pročišćen postupcima opisanima u Poglavlju 4.4.1 i nasumično je podijeljen na skup za učenje i skup za vanjsku provjeru (80 i 20 %). Za ukupno 12 pokazatelja toksičnosti (opisani u Poglavlju 3.3.2) iz skupa Tox21 (SR-HSE, NR-AR, SR-ARE, NR-Aromatase, NR-ER-LBD, NR-AhR, SR-MMP, NR-ER, NR-PPAR-gamma, SR-p53, SR-ATAD5, NR-AR-LBD) razvijeni su QSAR modeli postupcima opisanima u Poglavlju 4.5. Ukupno je razvijeno 60 QSAR modela, 36 metodom RF, te po 12 metodama LogReg i NN/MLP, tj. pet modela za svaku ciljnu varijablu. Kvaliteta svih QSAR modela razvijenih na spojevima iz skupa Tox21 iskazana je s više parametara (mjera) točnosti. Objedinjeni su rezultati dani u prilogu u elektroničkom dodatku E7. U ovom poglavlju dani su isključivo konačni rezultati QSAR modela razvijenih na spojevima iz skupa Tox21. Dodatni opis rezultata i doprinosa modela s više pojedinosti dan je u Poglavlju 5.7.

Najbolji modeli za pojedine ciljne varijable izabrani su prema kriteriju: Kappa/MCC Test iznad 0,20. Za svih 12 ciljnih varijabli barem jedan model zadovoljio je spomenute kriterije. Izbor konačnog modela, iz niza učenih modela prema toksičnom učinku, definiran je prema kriteriju SM (Scoring MCC) bodovanju prema jednadžbi 2.16. Popis konačnih QSAR modela razvijenih na skupu Tox21, koji su zadovoljili postavljene kriterije točnosti, dan je u tablici 17. Odabrani najbolji modeli spremljeni su u *.sav* format za daljnje korištenje i ekstrapolaciju na dodatnim skupovima molekula. Spremljeni su modeli korišteni za razvoj modela i predviđanju toksičnih učinaka na skupovima ToxCast i Sava.

Tablica 17. Kvaliteta 12 konačnih (izabranih) QSAR modela na skupu Tox21 (Fin21), iskazana parametrima (mjerama) točnosti na skupu za učenje (Train i CV) i na skupu za vanjsku provjeru (Test).[a)]

Toksični učinak	MCC Train	MCC CV	MCC Test	ΔQ2 Test	Točnost Test	Kappa Test	BA	Algoritam	Pred. varij.	Selekc. varij.	Broj varij.	SM
SR-HSE	0,67	0,27	0,28	0,03	0,92	0,28	0,66	LogReg	FP	Ne	5120	0,49
NR-AR	0,61	0,57	0,71	0,05	0,98	0,70	0,79	RF	FP	Ne	5120	0,73
SR-ARE	0,67	0,36	0,37	0,11	0,81	0,37	0,70	LogReg	FP	Ne	5120	0,55
NR-Aromatase	0,95	0,33	0,43	0,04	0,95	0,42	0,69	LogReg	FP	Ne	5120	0,55
NR-ER-LBD	0,64	0,51	0,44	0,03	0,96	0,43	0,67	RF	DS	Ne	203	0,60
NR-AhR	0,82	0,51	0,46	0,11	0,87	0,45	0,75	LogReg	FP	Ne	5120	0,61
SR-MMP	0,83	0,62	0,57	0,14	0,89	0,57	0,77	RF	DS	Ne	203	0,70
NR-ER	0,62	0,39	0,39	0,06	0,89	0,36	0,64	RF	DS	Da	89	0,57
NR-PPAR-gamma	0,83	0,27	0,32	0,02	0,96	0,32	0,67	LogReg	FP	Ne	5120	0,50
SR-p53	0,91	0,32	0,29	0,04	0,90	0,28	0,67	LogReg	FP	Ne	5120	0,50
SR-ATAD5	0,44	0,33	0,31	0,03	0,93	0,30	0,68	RF	DS	Ne	203	0,51
NR-AR-LBD	0,71	0,63	0,63	0,03	0,98	0,60	0,73	NN	FP	Ne	5120	0,74

[a)] **BA Test** - uravnotežena točnost na test skupu koja se računa prema jedn. (2.11); **Pred. varij.**– skup prediktorskih varijabli na kojem je model temeljen/razvijen; **Selekc. varij**. – selekcija (izbor) varijabli (provedena ili nije provedena); **Broj varij.** - broj varijabli uključenih u model; **SM** - bodovane kvalitete modela prema jedn. (2.16)

U kasnijim će poglavljima ovdje izračunati biološki deskriptori biti uvršteni u tzv. prošireni skup deskriptora (PDS) uz fizikalno-kemijske deskriptore (DS) (vidi Poglavlje 4.4.3). Svih 357 spojeva iz skupa Sava (osim Flusilazola) korišteno je u izračunu bioloških deskriptora (ekstrapolaciji) pomoću 12 konačnih Tox21 modela. Tako je dobiveno 12 novih predikorskih varijabli (bioloških deskriptora) za uključivanje u QSAR modele na skupu ToxCast. Skup ovih 12 konačnih modela ubuduće se skraćeno označava s Fin21 (Final Tox21 modeli).

5.6.2 QSAR modeliranje na spojevima skupa ToxCast

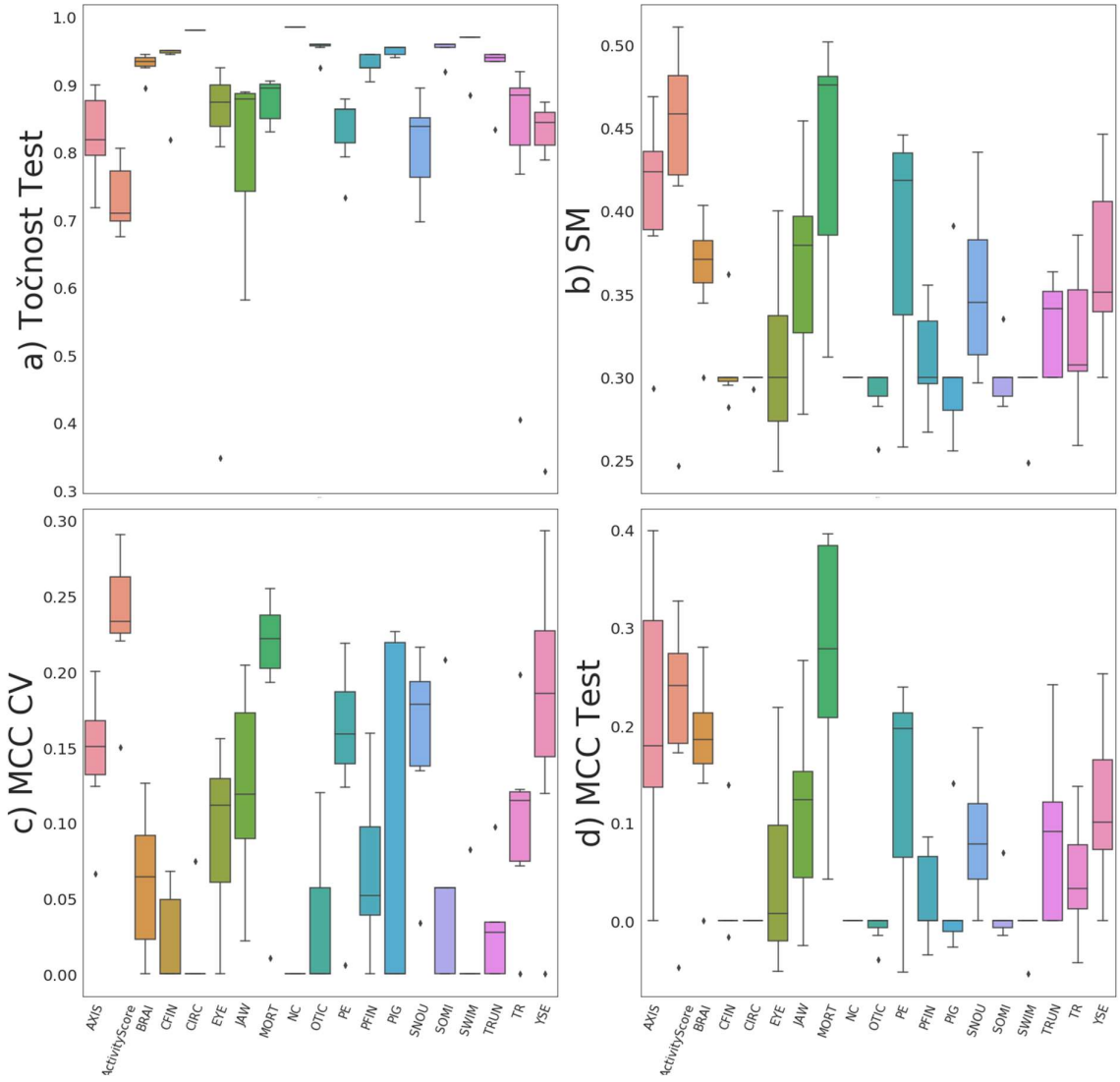
Razvijeno je ukupno 133 modela, sedam za svaku od 19 toksičnosti iz skupa ToxCast opisanih u Poglavlju 3.3.1. Vrijednosti parametara točnosti i druge pojedinosti svih modela dani su u elektroničkom dodatku E7. Modeli su optimirani i opisani postupkom opisanim u Poglavlju 4.5. U narednim su Poglavljima (5.6.3 – 5.6.10) opisani doprinosi pojedinih faktora tijekom simultanog modeliranja ovog skupa toksičnih učinaka.

5.6.3 ToxCast: Informativnost skupa toksičnih učinaka

Za potrebe ispitivanja mogućnosti ugađanja pojedinih toksičnih učinaka (njihove modelabilnosti), svojstva modela ispitana su na tri skupa prediktorskih varijabli:

- 5120 molekulskih otisaka (engl. *fingerprints* ili FP),
- 203 molekulska deskriptora (osnovni skup, DS) i
- proširenim molekulskim deskriptorima (PDS) (203 + 12 bioloških deskriptora + 3 deskriptora iz matrice ugnježđenja, ukupno 218 deskriptora).

Cilj testa bio je provjeriti općenitu mogućnost ugađanja modela za pojedine ciljne varijable, bez obzira na kemijsku reprezentaciju i algoritam. Za lakše razumijevanje prikazani su na slici 30. pravokutni dijagrami prediktivnih metrika svih modela podijeljeni prema toksičnim učincima. Cilj je dočarati doseg mogućnosti modeliranja toksičnosti i robusnost optimiranih/razvijenih modela, neovisno o hiperparametrima. Iako je izračunat velik broj mjera kvalitete, radi jednostavnosti ovdje su izabrani parametri MCC CV, MCC Test, Q_2 Test i SM, koje su prethodno definirani u Poglavlju 2.6. MCC CV odnosi se na Matthewsov korelacijski koeficijent tijekom križne validacije, MCC Test na Matthewsov korelacijski koeficijent izračunat na skupu za vanjsko vrednovanje (test skup), Q_2 Test na točnost na skupu za vrednovanje, a SM (Scoring MCC) kriterij je prethodno definiran u jednadžbi 2.16. Medijan MCC Test svih (133) modela je 0,01, što ukazuje na to da je većina modela s rezultatom u području nasumičnog vrednovanja (definiran prethodno kao MCC < 0,20), prema Q_2 Test koji je 0,90, što ukazuje na mogućnost krivog tumačenja kvalitete modela, kako je već utvrđeno u Poglavlju 2.6. Detaljna je usporedba metrika opisana u Poglavlju 5.6.8. Sudeći prema medijanima i prethodno definiranim kriterijima za prihvatljive modele [106], Kappa > 0,20 (tj. MCC > 0,20) može se utvrditi da su klasifikacijski rezultati mnogih modela na ToxCast skupu na razini nasumične točnosti (kvalitete). Za internu je provjeru kvalitete modela u postupku ugađanja/optimizacije na skupu za učenje korištena deseterostruka križna-validacija, koja je relativno strog postupak.



Slika 30. Usporedba mogućnosti ugađanja/predviđanja toksičnosti iz skupa ToxCast prema odabranim kriterijima; (a) Točnost Test, (b) SM, (c) MCC CV, (d) MCC Test za sve toksične učinka skupa ToxCast (x-os). Stupci predstavljaju 19 toksičnosti modeliranih u 19 prethodno opisanih kombinacija.

Prema kriteriju MCC Test > 0,2 , 22 od ukupno 133 modela imaju točnost klasifikacije iznad nasumične razine (tj. pokazuju prihvatljivu točnost). Iz toga zaključujem da je modelabilnost ovog skupa ograničena. Sudeći prema SM bodovanju (jednadžba 2.16), koje je postavljeno kao odlučujući kriterij u ovom radu, toksičnosti za koje su dobiveni najbolji QSAR modeli (prikazane kao medijan svih modela) su: MORT (SM = 0,48), ActivityScore (SM = 0,46), AXIS (SM = 0,42), PE (SM = 0,42), JAW (SM = 0,38). SM rezultati dani su za sve modele u tablici 18. Prema kriteriju metrike MCC Test koja se odnosi na prediktivnost modela na vanjskim

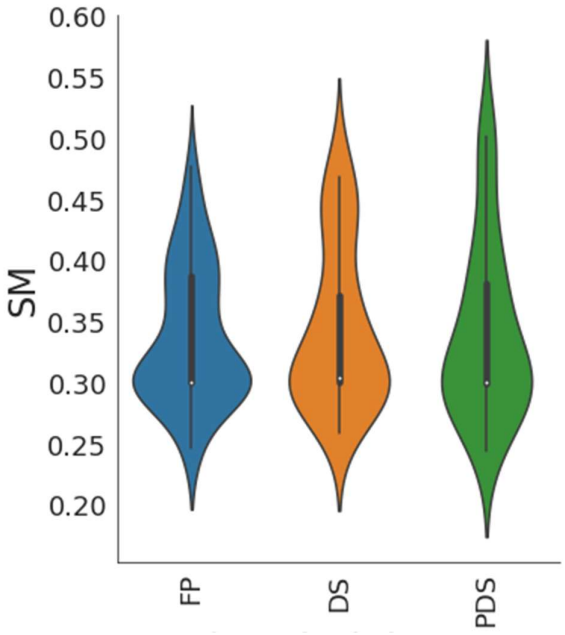
skupovima, 2 od 19 toksičnosti ima medijan iznad 0,20 (MORT i ActivityScore). Stoga, najinformativnije su toksičnosti MORT i ActivityScore.

Tablica 18. Usporedba informativnosti toksičnosti iz skupa ToxCast: mortalitet (MORT), spontano plivanje (SWIM), edem žumanjčane vrećice (YSE), deformacija notokorda (NC), deformacija tjelesne osi (AXIS), deformacija oka (EYE), deformacija njuške (SNOU), deformacija čeljusti (JAW), deformacija ušnog mjehurića (OTIC), perikardijalni edem (PE), deformacija mozga (BRAI), skraćeno tijelo (TRUN), deformacija somita (SOMI), deformacija prsnih peraja (PFIN), deformacija repne peraje (CFIN), pigmentacija (PIG), promjena u cirkulacijskom sustavu (CIRC) i reakcija na dodir (TR). Rezultati su prikazani kao medijan svih modela pojedinog kriterija kvalitete (MCC (Matthews korelacijski koeficijent) CV, MCC Test, SM (bodovanje prema MCC), ROBM (robusnost)). Tablica je poredana prema SM bodovanju.

Toksični učinak	MCC CV	MCC Test	SM	ROBM	ΔQ2 Test, %
MORT	0,22	0,28	0,48	0,82	8,62
ActivityScore	0,23	0,24	0,46	0,96	17,58
PE	0,16	0,20	0,42	0,95	5,97
BRAI	0,06	0,19	0,37	0,88	1,90
AXIS	0,15	0,18	0,42	0,93	6,96
JAW	0,12	0,12	0,38	0,95	2,69
YSE	0,19	0,10	0,35	0,95	4,51
TRUN	0,03	0,09	0,34	0,99	1,57
SNOU	0,18	0,08	0,34	0,97	2,63
TR	0,12	0,03	0,31	0,96	0,71
EYE	0,11	0,01	0,30	0,88	0,18
NC	0,00	0,00	0,30	1,00	0,00
CIRC	0,00	0,00	0,30	1,00	0,00
CFIN	0,00	0,00	0,30	1,00	0,00
PFIN	0,05	0,00	0,30	0,96	0,00
PIG	0,00	0,00	0,30	1,00	0,00
SOMI	0,06	0,00	0,30	0,94	0,00
SWIM	0,00	0,00	0,30	1,00	0,00
OTIC	0,00	0,00	0,30	1,00	0,00

5.6.4 ToxCast: Doprinos kemijske reprezentacije (prediktorskih skupova) i selekcije varijabli

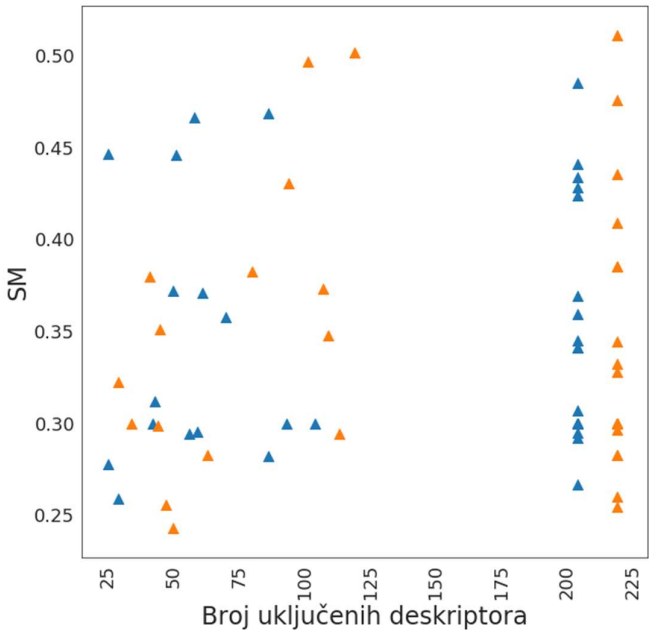
Od ukupno 133 modela, njih 38 trenirano je na skupu deskriptora (DS), 38 na proširenom skupu deskriptora (PDS), te 57 na molekulskim otiscima (FP). Kako bi se usporedio doprinos kvaliteti predviđanja odvojeno su prikazani modeli izgrađeni na skupu deskriptora, proširenih deskriptora i na skupu molekulskih otisaka (slika 31).



Slika 31. Usporedba skupova prediktorskih varijabli prema mjeri kvalitete SM (bodovanje prema MCC) danom jednadžbom 2.16.

Rezultati doprinosa pojedinih prediktivnih skupova pokazuju da su prema kriteriju odlučivanja SM najkvalitetniji oni modeli koji su izgrađeni na skupu proširenih deskriptora. Na slici 32 dan je dijagram raspršenja koji prikazuje odnos broja varijabli i kvalitete modela prema kriteriju SM (jednadžba 2.16). Na slici se vide modeli s punim brojem deskriptora (218 za PDS i 203 za DS) i zatim praznina po x-osi do nižih brojeva deskriptora, što se temelji na postavkama selekcije varijabli (Poglavlje 2.8). Iz slike je razvidno da su u izboru prediktorskih skupova DS i PDS dominirali PDS kao informativniji skup (narančasti trokutići na slici). Također se vidi da su za dobivanje većine najboljih modela bilo kojom od tri metode bili potrebni svi dostupni deskriptori (jer su postignuti kvalitetniji modeli prema SM s punim skupom bez selekcije varijabli). Za modele koji su koristili deskriptore i (DS, PDS) uz selekciju varijabli izračunat je prosječan broj izabranih varijabli. Veliki broj izabranih varijabli u konačnim modelima upućuje na složenost toksičnosti kao ciljne varijable u QSAR modeliranju. Za izračun srednje vrijednost

korištena su dva RF modela po toksičnom učinku. Rezultat je prikazan u tablici 19. gdje se vidi da su prema navedenom kriteriju manje složeni modeli za toksične učinke: JAW, YSE, PE, SNOU, TR (modeli s prosječno manje od 50 varijabli/deskriptora).



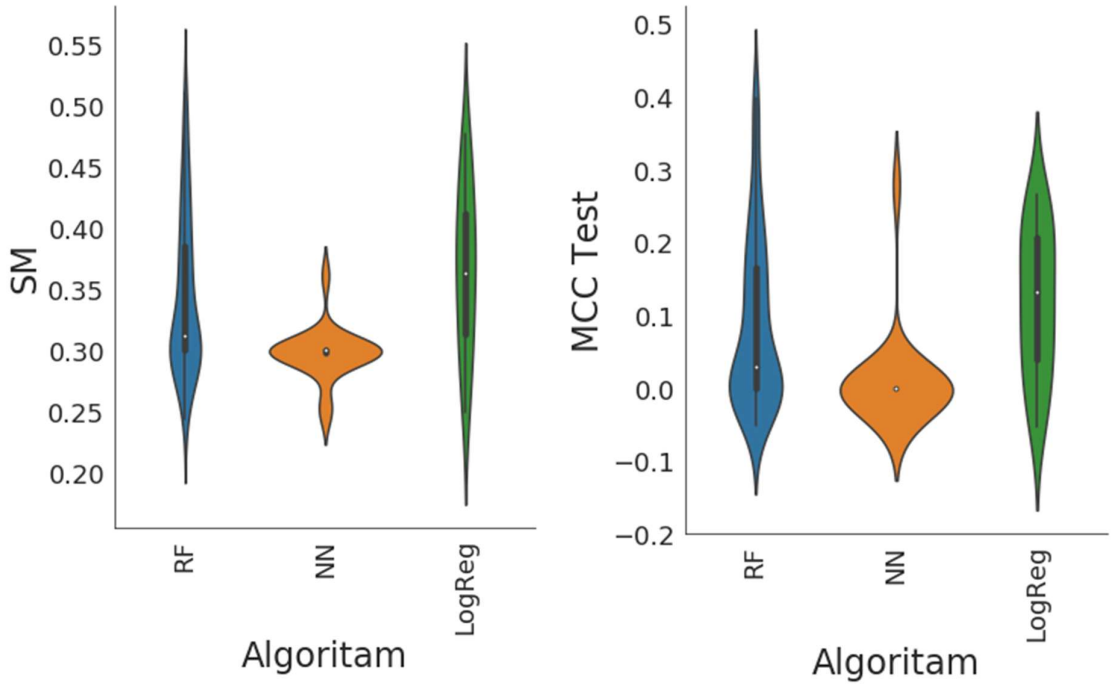
Slika 32. Prikaz ovisnosti indeksa SM o broju korištenih molekulskih deskriptora za modele optimirane na skupu deskriptora (DS) i proširenih deskriptora (PDS). Slika se odnosi na skup za vrednovanje (Test set).

Tablica 19. Prosjek broja korištenih deskriptora za QSAR modele toksičnosti na skupovima DS i PDS gdje je korištena selekcija varijabli.

Toksični učinak	Broj varijabli za dva modela	Srednja vrijednost
JAW	{25, 41}	33,0
YSE	{25, 45}	35,0
PE	{51, 29}	40,0
SNOU	{50, 44}	47,0
TR	{80, 29}	54,5
SOMI	{56, 63}	59,5
CFIN	{34, 86}	60,0
EYE	{50, 70}	60,0
PIG	{93, 47}	70,0
OTIC	{113, 42}	77,5
ActivityScore	{58, 101}	79,5
MORT	{43, 119}	81,0
PFIN	{59, 109}	84,0
BRAI	{107, 61}	84,0
AXIS	{94, 86}	90,0
SWIM	{104, 218}	161,5
NC	{218, 204}	211,5
CIRC	{218, 204}	211,5
TRUN	{218, 204}	211,5

5.6.5 ToxCast: Doprinos klasifikacijskog algoritma

Od ukupno 133 modela, 95 modela razvijeno je algoritmom Random Forest, 19 algoritmom neuronskih mreža i 19 logističkom regresijom. Na slici 33 prikazan je violinski dijagram rezultata modela prema SM i MCC testu u odnosu na algoritme na kojima se modeli temelje. Iz slike je vidljivo da su modeli temeljeni na logističkoj regresiji (LogReg) i algoritmu Random Forest (RF) bolji od onih temeljenih na metodi neuronskih mreža (NN) prema kriteriju/metrici SM. S obzirom na definiciju SM-a, razvidno je da modeli na osnovi NN preugađaju.



Slika 33. Usporedba utjecaja algoritma na ishod prediktivnih modela.

Iz rezultata je izračunato da su omjeri MCC CV/MCC Test kako slijedi RF – 1,11; LogReg – 1,25; NN – 2,00. Pojedini RF modeli pak postižu visoke vrijednosti MCC test (slika 33), dok se većina modela optimiranih s pomoću NN nalazi u području MCC test < 0,2, što pokazuje da nemaju sposobnost predviđanja, tj. predviđanje je potencijalno nasumično.

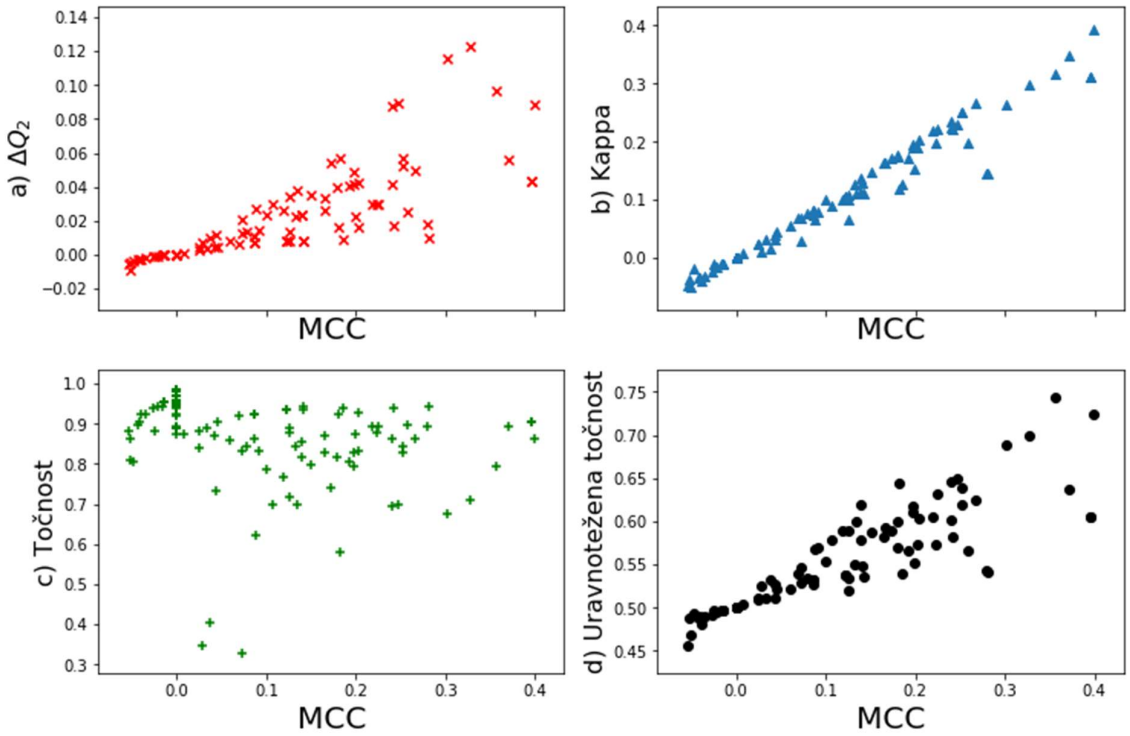
5.6.6 ToxCast: Važnost mjera kvalitete modela

U Poglavlju 2.6 objašnjene su mjere kvalitete modela, a detaljnije su obrađene u radu. [188] Iz rezultata 133 ToxCast modela izračunate su aritmetičke sredine za pojedine mjere kvalitete modela. Zatim je izračunata korelacijska matrica prema Pearsonu za spomenute srednje vrijednosti. U tablici 20 prikazani su rezultati korelacije vrijednosti različitih metrika na skupu za vrednovanje (*test set*). Iz tablice je vidljivo da mjere MCC, Kappa i ΔQ_2 međusobno pozitivno koreliraju ($r > 0,83$), od čega je između vrijednosti Kappa i MCC koeficijent korelacije r = 0,98.

Tablica 20. Korelacijska matrica primijenjenih metrika na skupu za vrednovanje (test set) između vrijednosti metrika dobivenih s pomoću 133 modela.

	ΔQ_2	**MCC**	**Kappa**	**Točnost**	**Uravnotežena točnost (BA)**
ΔQ_2	1,00	0,83	0,86	-0,49	0,95
MCC	0,83	1,00	0,98	-0,31	0,90
Kappa	0,86	0,98	1,00	-0,30	0,93
Točnost	-0,49	-0,31	-0,30	1,00	-0.47
Uravnotežena točnost (BA)	0,95	0,90	0,93	-0,47	1,00

Često korištena i preporučena mjera u popularnim radovima [107,108] u klasifikaciji neuravnoteženih skupova je uravnotežena točnost (engl. *balanced accuracy*, BA). BA (jednadžba 2.11) je aritmetička sredina osjetljivosti (Sn) i specifičnosti (Sp), a to znači da BA može imati relativno visoku vrijednost uz nisku vrijednost jednog od dva (Sn ili Sp) ulazna parametra za računanje BA. U slučajevima s neuravnoteženim skupovima Sp ima u pravilu vrlo visoke vrijednosti jer se odnosi na većinsku (negativnu/netoksičnu) klasu i visoka vrijednost BA može tako prikriti nedostatnu klasifikaciju manjinske klase. BA nešto slabije korelira s MCC i Kappa dok najbolju korelaciju ima s ΔQ_2 (tablica 20). Sve četiri metrike imaju negativnu korelaciju s točnošću.



Slika 34. Prikaz svih ovisnosti vrijednosti odabranih metrika o vrijednostima MCC na vanjskom (test) skupu modela razvijenih na skupu ToxCast (ukupno 133 modela). Na x-osi je MCC Test, a na y-osi redom: (a) ΔQ_2, (b) Kappa, (c) Točnost (Q_2), (d) Uravnotežena točnost (BA).

Na slici je vidljivo da postoje modeli iz ovog eksperimenta koji imaju nizak MCC i iznimno visoku točnost. To je posljedica upravo objašnjenog nedostatka slabe osjetljivosti parametra (metrike) točnosti za neuravnotežene skupove, i zato je važno koristiti višestruke i osjetljivije mjere kvalitete modela. Raspon MCC je (-1; 1) dok je BA u rasponu (0,5; 1,0). Na slici 34 vidi se da je raspon svih razvijenih modela prema BA do 0,75 što je oko 50 % definiranog raspona. Istovremeno su maksimalni Kappa i MCC približno 0,4 što pokazuje veću osjetljivost tih metrika kao i ΔQ_2 koji je za ove modele do ~0,13 od maksimalnih 0,50. Ove vrijednosti metrike ΔQ_2 pokazuju da su skoro svi razvijeni modeli, a svi najbolji modeli, iznad razine točnosti koja bi se dobila odgovarajućim nasumičnim modelom.

5.6.7 ToxCast: Kombinirani doprinos algoritama i prediktorskih varijabli. Najbolji modeli

U nastavku su ispitani doprinosi kombinacije algoritama i tri skupa prediktorskih varijabli putem agregacije po modelu i prediktorskom skupu (LogReg – FP, NN – FP, RF- PDS, RF - DS. RF -FP). Agregacijska vrijednost je aritmetička sredina mjere SM. Analiza je ograničena jer nisu svi algoritmi kombinirani sa svim skupovima zbog daljnje uporabe modela u serverskim

aplikacijama. Analiza je prikazana u tablici 21, iz koje se vidi da su najbolji modeli prema kriteriju bodovanja SM (jednadžba 2.16) oni koji su temeljeni na strukturnim prediktorskim varijablama koji su uključivali samo molekulske otiske (FP).

Tablica 21. Doprinos kombinacije vrste algoritma i skupa prediktorskih varijabli na 133 modela skupa ToxCast. Mjere su agregirane prema skupu prediktors se odnose na skup za vrednovanje.

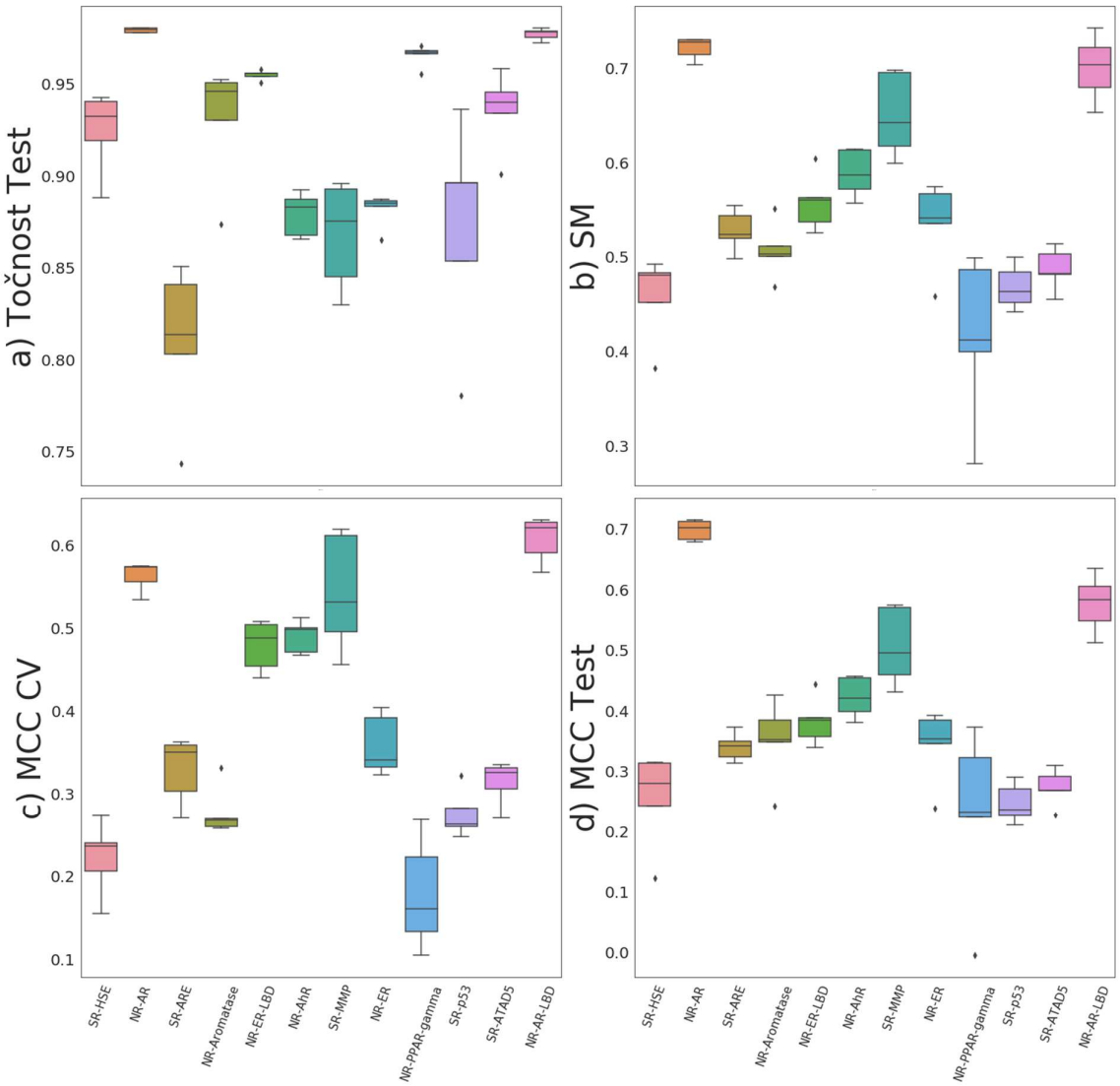
Klasifik. algoritam	**LogReg**	**NN**		**RF**	
Metrike/Pred varijable	***FP***	***FP***	***PDS***	***DS***	***FP***
MCC CV	***0,15**	0,02	0,12	0,08	0,08
MCC Train	***0,81**	0,12	0,44	0,44	0,37
MCC Test	***0,12**	0,01	0,08	0,09	0,09
SM	0,37	0,30	0,34	0,34	0,35

Iz praktičnih razloga, kako bi razvijeni modeli bili primjenjivi na vanjskom skupu molekula, na skupovima DS i PDS nisu razvijani modeli algoritmima LogReg i NN. Naime, oba ta algoritma zahtijevaju skaliranje numeričkih podataka, koje bi bilo jako teško objektivno provesti na novim skupovima molekula. RF je algoritam koji ne zahtijeva skaliranje jer dijeli varijablu tijekom učenja na nekoj vrijednosti bez obzira na skalu. Najbolji razvijeni modeli služit će daljnjoj uporabi na vanjskim skupovima, a u disertaciji to je skup Sava.

5.6.8 Tox21: Rezultati modeliranja i usporedba kvalitere modela skupova Tox21 i ToxCast

U ovom dijelu opisani su detalji modeliranja skupa Tox21 u svrhu dobivanja bioloških deskriptora (Poglavlje 5.6.1) a opis je koncipiran kao usporedno Poglavljima 5.6.3 – 5.6.8 (ToxCast) i nastavno na ta poglavlja. Naime, rabi se identična metodologija, ali se koristi skup Tox21 umjesto skupa ToxCast. Razvijeni su modeli na 12 toksičnosti baze Tox21 postupkom opisanim u Poglavlju 4.5. Ukupno je trenirano i validirano 60 modela, od čega je 36 RF, 12 LogReg i 12 NN modela. Ukupno je razvijeno po pet modela za svaku toksičnost iz baze Tox21. Rezultati svih izračunatih Tox21 modela sa svim metrikama nalaze se u prilogu u elektroničkom dodatku E7. Ukupno 24 RF modela razvijena su polazeći od RDKit deskriptora, 12 RF modela razvijeno je na molekulskim otiscima, dok je po 12 modela razvijeno na molekulskim otiscima uporabom LogReg i NN. Svi su modeli optimirani Bayesovom optimizacijom na 10x križnoj validaciji (Poglavlje 4.4). Pravokutni dijagram rezultata svih

modela prikazan je na slici 35. Iz slike je razvidno da su toksičnosti skupa Tox21 užih raspodjela u odnosu na skup ToxCast (slika 35). To se pripisuje kvaliteti eksperimentalnih podataka koji su kvalitetniji za modeliranje kod staničnih linija nego kod cijelih organizama, posebice ako postoji utjecaj stresora koji ne dolaze od primjenjenog kemijskog spoja nego okoline (velika kontrolna smrtnost). [189,190]



Slika 35. Usporedba svih 60 modela razvijenih na skupu Tox21. Korištena su tri algoritma (RF, NN, LogReg) s dva različita skupa prediktorskih varijabli (skup s različitim fizikalno-kemijskim deskriptorima (DS) i skup samo s molekulskim otiscima (FP)). Kvaliteta modela prikazana je pomoću metrika (parametara kvalitete): (a) Točnost Test, (b) SM, (c) MCC CV, (d) MCC Test za sve toksične učinke skupa Tox21 (x-os). Tamne točke oko pravokutnika su odstupajuće vrijednosti (engl. *outlier*).

Također je razvidno da postoje velike razlike u informativnosti točaka, tako NR-AR ima medijan MCC Test ~ 0,7, dok su pojedine toksičnosti u području MCC Test ~ 0,2 – 0,3što graniči s nasumičnim modelima po definiranim kriterijima. Modeli razvijeni na skupu fizikalno-kemijskih deskriptora izračunanih programom RDKit (DS) pokazuju približno iste rezultate na skupu Tox21 kao i modeli razvijeni na skupu koji sadrži samo strukturne otiske kao prediktorske varijable (FP) (tablica 22).

Tablica 22. Medijani svih razvijenih QSAR modela razvijenih na podacima baze Tox21 uz korištenje dva skupa prediktorskih varijabli (DS i FP).

Metrike/Prediktorske varijable	**DS**	**FP**
MCC CV	0,40	0,38
MCC Train	0,59	0,71
MCC Test	0,40	0,38
SM	0,56	0,55
ROBM	0,94	0,94

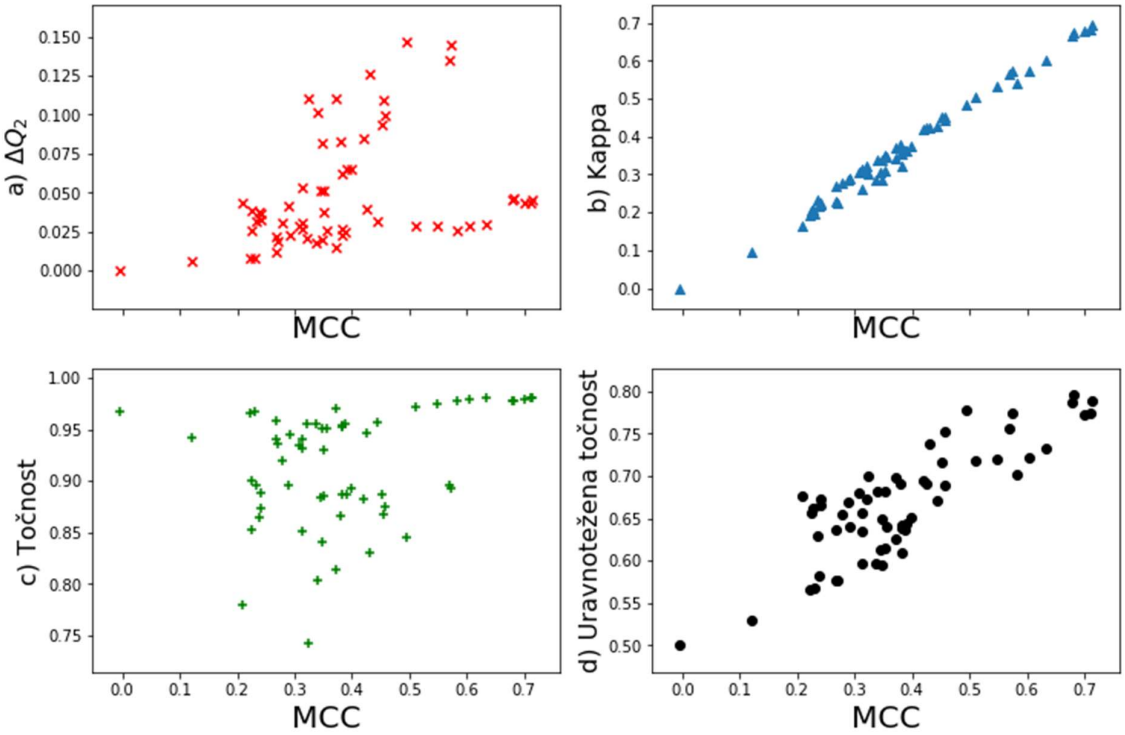
Kvalitete modela bitno su različite, što se vidi iz usporedbe medijana u tablici 23 za sve modele kroz oba skupa prediktora.

Tablica 23. Usporedba medijana svih modela razvijenih na skupovima molekula iz baza Tox21 (60 modela) i ToxCast (133 modela).

	Tox21	**ToxCast**
MCC CV	**0,35**	0,07
MCC Train	**0,64**	0,50
MCC Test	**0,35**	0,01
SM	**0,54**	0,30
ROBM	**0,95**	0,96

Prema svim ključnim metrikama (MCC CV, MCC Train, MCC Test, SM, ROBM) postignuti su bolji rezultati modeliranja sa skupom Tox21 nego je to bio slučaj s modelima razvijenim na skupu ToxCast. Dodatno, skup Tox21 sadrži 5-6 puta više molekula nego skup ToxCast. Iz toga se može izvući zaključak da su vrijednosti parametara kvalitete (metrika) modela razvijenih na skupu Tox21 (u pravilu) iznosom veće, a statistički gledano uvijek daleko signifikantnije, od vrijednosti odgovarajućih metrika modela razvijenim na skupu ToxCast. Izabrano je 12

najboljih konačnih modela (opisani u Poglavlju 5.6.3) koji su poslužili za računanje bioloških deskriptora, tj. predviđanje (ekstrapolaciju) na skupovima svih molekula i toksičnosti baze ToxCast i skupu spojeva izmjerenih u rijeci Savi. Od tih 12 modela šest je razvijeno uporabom metode logističke regresije (LogReg), pet s pomoću RF te jedan s pomoću NN. Jedan RF model temeljen je na molekulskim otiscima, dok su ostala četiri RF modela temeljeni na fizikalno-kemijskim deskriptorima. Preostali NN i LogReg modeli temeljeni su na molekulskim otiscima kao deskriptorima. što upućuje na to da je ta vrsta molekulskih deskriptora informativnija na skupu Tox21. Četiri od pet najboljih modela temeljenih na fizikalno-kemijskim deskriptorima koristili su sve dostupne deskriptore (njih 203). Usporedba međuodnosa mjera kvalitete na skupu Tox21 pokazuje slične obrasce kao i rezultati na skupu ToxCast. Na slici 36 vidljivo je da postoje pozitivne korelacije između vrijednosti MCC (na jednoj strani) i ΔQ_2, Kappa, Q_2 i uravnoteženoj točnosti (na drugoj strani).



Slika 36. Prikaz svih međuovisnosti vrijednosti odabranih metrika o vrijednostima MCC na vanjskom (test) skupu modela razvijenih na skupu Tox21 (ukupno 60 modela). Na x-osi je MCC Test, a na y-osi redom: (a) ΔQ_2, (b) Kappa, (c) Točnost (Q_2), (d) Uravnotežena točnost (BA).

Usporedba vrijednosti ΔQ_2 i MCC pokazuje da ne postoje modeli s visokim vrijednostima ΔQ_2 a da nemaju i relativno visoke vrijednosti MCC. Nadalje, usporedba točnosti i MCC pokazuje da na ovako velikom skupu postoji jako puno modela s visokim vrijednostima Q_2 i s jako niskim vrijednostima MCC. To potvrđuje nalaze iz poglavlja 5.6.7 da Q_2 ne može biti odlučujući kriterij odabira najboljeg (najboljih) klasifikacijskih modela razvijenim na neuravonoteženim skupovima podataka kakvi su baze Tox21 i ToxCast. Pritom, misli se da su ti skupovi neuravnoteženi samo u smislu velike razlike između broja molekula koji prema toksičnom učinku pripadaju skupini (klasi) aktivnih i broja molekula koje pripadaju skupini neaktivnih.

5.6.9 ToxCast: konačni QSAR modeli toksičnosti

Na temelju prethodno definiranih kriterija (MCC Test > 0,2 i ΔQ_2 Test > 0,001) izabrano je 7 modela od skupine modela razvijenih na bazi ToxCast koji su prešli oba praga. Zbog manjeg broja molekula na skupu ToxCast (u odnosu na skup Tox21) uveden je dodatni stroži kriterij za odlučivanje o kvaliteti modela, ΔQ_2. Taj dodatni kriterij osigurava da je predviđanje za barem 1 % iznad razine nasumičnog predviđanja (tj. pogađanja). Modeli koji su prešli definirane pragove odnose se na sljedeće toksične učinke: AXIS, ActivityScore, EYE, JAW, MORT, PE i YSE. Ukoliko je više modela prešlo definirani prag (MCC Test > 0,2 i ΔQ_2 Test > 0,001), za svaku od toksičnosti izabran je u konačnici najbolji model prema SM bodovanju. SM vrijednosti konačnih modela kreću se od 0,40 do 0,51, dok su vrijednosti MCC Test u rasponu od 0,22 do 0,40, a MCC CV između 0,11 i 0,29. Ovaj skup konačnih (najboljih) modela razvijenih na bazi spojeva i aktivnosti ToxCast nazvan je (zbog lakšeg razumijevanja)FinTC (od Final ToxCast), i nalazi se u tablici 24. Općenito niže vrijednosti za MCC CV u ovih sedam modela pokazuju da je to stroži kriterij od MCC Test. Razlog je taj što je MCC CV prosjek od 10 podjela križne validacije, što znači da pojedine podjela unutar tih 10 mogu narušiti mjeru kvalitete. Rezultati potvrđuju da najbolji modeli razvijeni na bazi ToxCast za ovih sedam toksičnosti imaju sve MCC Test vrijednosti iznad 0,2, što je prethodno postavljeno kao mjera dobrog slaganja.

Tablica 24. Sedam najboljih modela razvijenih na aktivnostima molekula iz baze ToxCast prema kriterijima MCC Test > 0,2 i ΔQ2 > 0,001 (1 %).[a)]

Toks. učinak	MCC Train	MCC CV	MCC Test	ΔQ_2 Test	Točnost Test	Kappa Test	BA Test	Algoritam	Pred. varij.	Selekc. varij.	Broj varij.	SM
AXIS	0,46	0,18	0,40	0,09	0,86	0,39	0,72	RF	DS	Da	86	0,47
A.S. [b)]	0,54	0,29	0,33	0,12	0,71	0,30	0,70	RF	PDS	Ne	219	0,51
EYE	0,91	0,11	0,22	0,03	0,89	0,22	0,61	LogReg	FP	Ne	5120	0,40
JAW	0,99	0,20	0,27	0,05	0,86	0,27	0,62	LogReg	FP	Ne	5120	0,45
MORT	0,68	0,26	0,37	0,06	0,90	0,35	0,64	RF	PDS	Da	119	0,50
PE	0,54	0,20	0,24	0,04	0,86	0,23	0,60	RF	DS	Da	51	0,45
YSE	0,50	0,19	0,25	0,05	0,84	0,25	0,62	RF	DS	Da	25	0,45

[a)] **Toks. učinak** - toksični učinak; **BA Test** - uravnotežena točnost na test skupu koja se računa prema jedn. (2.11); **Pred.varij.**– skupovi prediktorskih varijabli na kojem je model temeljen/razvijen; **Selekc. varij**. – selekcija (izbor) varijabli (provedena ili nije provedena); **Broj varij.** - broj varijabli uključenih u model; **SM** - bodovane kvalitete modela prema jedn. (2.16); [b)] **A. S.** – Activity Score (toksični učinak iz baze ToxCast opisan u radu)

Također. vrlo stroga metrika ΔQ_2 (koja opisuje stvarni doprinos modela iznad razine nasumične točnosti) u ovim najboljim modelima ima vrijednosti od 0,03 (3 %) pa na više. To je dodatna potvrda da je predviđanje s pomoću odabranih modela iznad točnosti nasumičnog modela.

5.6.10 ToxCast: Značajne varijable u QSAR modelima toksičnosti

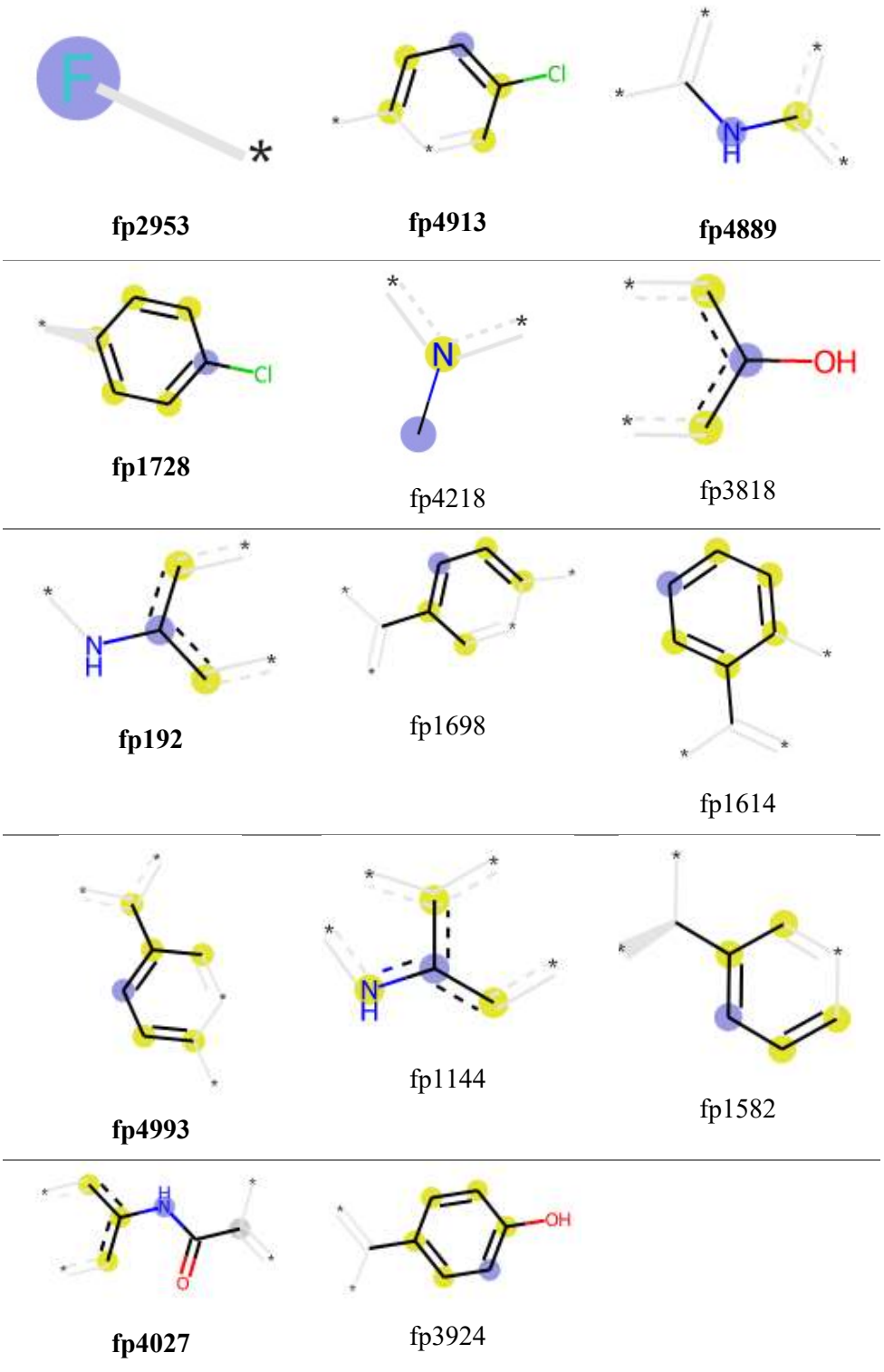
Doprinos pojedinih prediktorskih varijabli u FinTC modelima (konačni modeli na skupu ToxCast) analiziran je s pomoću permutacijske važnosti (PVV) opisane u Poglavlju 2.8. Ovdje su analizirani doprinosi 10 najvažnijih varijabli za sedam FinTC modela od kojih su dva razvijena na kombiniranom skupu prediktorskih varijabli PDS (AcitivityScore i MORT), tri na skupu DS (AXIS, PE, YSE) te dva na skupu prediktorskih varijabli FP (EYE, JAW). Tablica s PVV vrijednostima nalazi se u elektroničkom dodatku E8. U nastavku je dana kumulativna analize doprinisa prediktorskih varijabli (kemijske reprezentacije).

Za dva modela temeljena na molekulskim otiscima (FP) u tablici 25 prikazani su otisci koji su se pokazali bitnima, a pojedini su prikazani i grafički na slici 37. Vidi se da odabrani otisci označavaju prisutnost halogenih i drugih elemenata vezanih na aromatski prsten (AR), primjeri su (F-AR, Cl-AR. HN-AR, AR-N-R, AR-N-AR) kao i veliki broj položaja u aromatskih prstenovima. Poznato je od prije da se halogeni supstituenti, amini i pirimidini povezuju s

toksičnošću [191], stoga odabir ovih otisaka u modelu potkrepljuje dosadašnje razumijevanje toksikologije na molekulskoj razini. S tumačenjem molekuskih otisaka u modelima je potreban oprez jer algoritam kodiranja otisaka nije nužno uvijek jednoznačan i može doći do tzv. kolidirajućih bitova, što je poznato iz literature. [192,193]

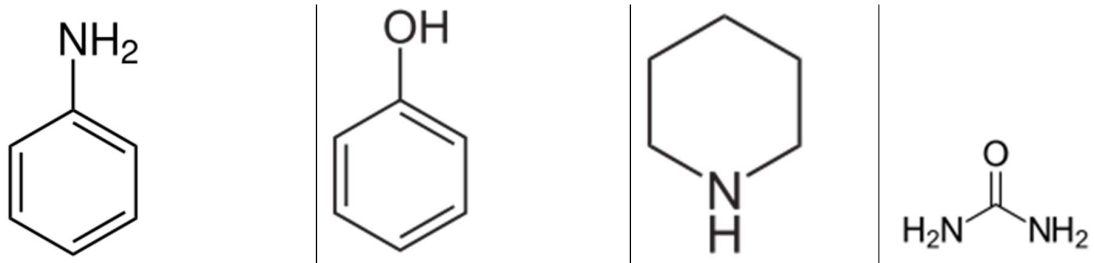
Tablica 25. Molekulski otisci koji su pokazali visoku važnost u klasifikaciji toksičnosti. Prva četiri otiska prikazana su na slici 37.

Molekulski otisak	**N**	**EYE**	**JAW**
fp2953*	134	x	
fp4913*	70	x	
fp4889*	61	x	
fp1728*	60		x
fp4218	59		x
fp3818	54		x
fp192	43	x	
fp1698	39		x
fp1614	38	x	
fp4993	35	x	
fp326	29		x
fp1144	16		x
fp1582	14		x
fp2984	14		x
fp4027	12	x	
fp3924	10		x
fp4219	7	x	
fp1411	5	x	
fp2126	3		x
fp3953	3		



Slika 37. Izabrani molekulski otisci prema učestalosti pojavljivanja iz tablice 25. fp326, fp2984, fp4219, fp2126, fp3953 ovdje nisu prikazati jer snažno podliježu kolidirajućim bitovima i nisu jednoznačni.

Među 10 najvažnijih varijabli unutar najboljih modela (razvijenih na skupovima prediktora DS i PDS) našle su se i pojedine funkcionalne grupe koje ne uzimaju u obzir atomarno susjedstvo (kao molekulski strukturni otisci FP). Takvi deskriptori nazivaju se još i fragmentni deskriptori. Četiri takva fragmentna deskriptora (anilinska grupa, fenolna, piperidinska, urea) prikazana su u slici 38, a pokazali su utjecaj u modelima toksičnosti za ishode AXIS, MORT, PE i YSE. Navedeni fragmentni deskriptori pokazuju slična strukuralna obilježja kao i otisci korišteni u modelima za EYE i JAW - a to su heteroatomi vezani za prstenove (slika 38 a,b), kao -NH- , -NH_2 fragmenti (slika 38 c,d). Iz usporedbe modela izgrađenim na otiscima (slika 37) i onih na fragmentnim deskriptorima (slika 38) razvidno je da su u modele uključena slična struktura obilježja.



a) Anilin b) Fenol c) Piperidin d) Urea

Slika 38. Fragmentni deskriptori koji su pokazali važnost (PVV) u četiri modela na skupovima prediktora DS i PDS.

Skup prediktorskih varijabli PDS obuhvaća pojedine biološke deskriptore koji su produkt ekstrapolacije Fin21 modela (Poglavlje 5.6.1). Modeli izgrađeni na skupu prediktora PDS pokazali su najbolje rezultate za toksične ishode ActivityScore i MORT iz baze ToxCast. Za MORT korišteni su biodeskriptori NR-Aromatase, NR-ER-LBD, SR-MMP te SR-p53, i pritom je biodeskriptor SR-MMP bio jedan od važnijih deskriptora (varijabli) u modelu. Za ActivityScore visoku PVV pokazali su biodeskriptori NR-AR, SR-ARE te također SR-MMP. Načelno su se toksični ishodi ActivityScore i MORT pokazali najpogodnijima za izgradnju modela. Zanimljivo je da su za te obje toksičnosti važni deskriptori upravo biološki deskriptori, posebice SR-MMP koji se asociran s negativnim učinkom na mitohondrijsku membranu. Važnost procesa na membrani za staničnu toksičnost dobro je poznata činjenica u literaturi [194]. Preostali deskriptori iz skupova DS i PDS koji su pokazali visoke PVV vrijednosti dani su u tablici 26.

Tablica 26. Najvažniji fizikalno-kemijski deskriptori uključeni u najbolje konačne QSAR modele razvijene na skupu molekula i svojstava iz baze ToxCast.

Tip deskriptora	Korišteni pojedini deskriptori
Topološki deskriptori	Chi1n, Chi3n
EState VSA	EState_VSA1, EState_VSA2, EState_VSA4, EState_VSA6, EState_VSA7
MOE Charge VSA	PEOE_VSA10, PEOE_VSA13, PEOE_VSA5
MOE logP VSA	SlogP_VSA11, SlogP_VSA12, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, SlogP_VSA7, SlogP_VSA8
VSA Estate	VSA_EState10, VSA_EState2, VSA_EState3, VSA_EState4, VSA_EState5, VSA_EState7, VSA_EState8, VSA_EState9
Gustoća molekulskih otisaka	FpDensityMorgan3
EState indeks	MinAbsEStateIndex
Polarna površina molekule	TPSA
MOE MR VSA	SMR_VSA5
Sličnost lijekovima	qed

Deskriptori dobiveni iz matrica ugnježđenja našli su se u najvažnijim deskriptorima u modelima. Ovdje je u razmatranju obuhvaćeno samo 10 deskriptora po modelu, preostali deskriptori također imaju doprinose u modelima, ali su u ovoj analizi zanemareni zbog malih doprinosa.

5.6.11 Usporedba QSAR modela na skupovima ToxCast i Tox21 s modelima iz literature

U velikom broju radova objavljeni su rezultati modeliranja toksičnih učinaka (njih 12) iz skupa Tox21. [84,93,200–206,99,172,187,195–199] Usporedba s mnogim radovima je otežana jer pojedini imaju manje stroge procjene kvalitete modela kao što je to slučaj u [93] gdje autori koriste procjenu parametera tijekom križne validacije i nemaju odvojen vanjski skup za provjeru modela (test set) kao što je slučaj u ovom radu. Većina autora koristi podjelu skupa Tox21 koja je dana na natjecanju u predviđanju toksičnih ishoda „Tox21 challenge 2014“ [199].U sklopu ovog rada podaci su pažljivo pročišćeni i uviđen je niz nedostataka u izvornim podacima koji navode na

zaključak da su kvalitete dosad objavljenih modela potencijalno pristrane zbog prisutnosti strukturnih duplikata između podskupova. Dodatnu razliku čini činjenica je izvorni skup podijeljen na u podskupove za učenje i vrednovanje na omjere 5500 - 11 764 (skupovi za učenje) i 296 (skup za validaciju), spram 647 spojeva u skupu za vanjsko vrednovanje (test skup) što predstavlja 5-10 % ukupnog broja spojeva u prva dva skupa.

U ovom radu je veličina skupa za vanjsko vrednovanje čini stalni udio od 20 % nasumično odvojenih spojeva što daje manji prostor za učenje modela, ali i jednu širu distribuciju pri ispitivanju kvalitete predviđanja. Dominante mjere kvalitete modela u spomenutim radovima su površina ispod ROC (engl. *receiver operating characteristic*) krivulje (engl. *area under the curve*, AUC) ili uravnotežena točnost (engl. *Balanced accuracy*, BA) [84]. Tek mali broj radova koristi kritične mjere za rad s uravnoteženim skupovima kao što su MCC ili Kappa. [99,172,196] S obzirom na razlike u podjeli skupova ovdje dana usporedba prema AUC i BA nije striktna nego je orjentacijska, s obzirom da se vanjski skupovi razlikuju od jednog do drugog modela. Pritom, važna prednost modela razvijenih u disertaciji (i najboljih modela prikazanih u tablici ispod) u tome je što su test skupovi na kojem su izračunane vrijednosti metrika (parametara) najmanje dvostruko veći od svih ostali navedenih modela iz literature.

Važno je napomenuti da test skup koji predstavlja 20 % ukupnih podataka nije isti kod svih 12 toksičnih učinaka iz baze Tox21 koji su modelirani u disertaciji stoga što je nakon opisanih postupaka pročišćavanja skupova podataka za svaki toksični učinak preostao različit ukupan broj molekula. Test skupovi sadržavali su od 1200 do 1500 spojeva. Rezultati usporedbe na skupu Tox21 predstavljeni su u tablicama 27 i 28. Uočava se da su za svih 12 toksičnih učinaka vrijednosti metrike BA modela razvijenih u disertaciji (posljednji stupac) usporedivi ostalim modelima razvijenim tijekom sudjelovanja na natjecanju 2014 godine. Treba napomenuti važnu pojednost a to je da su ovi rezultati prema BA dobiveni u disertaciji osjetno više statističke značajnosti stoga što su skupovi na kojima su računani u disertaciji više od dva puta veći nego kod preostalih metoda navedenih u tablici 27.

Nadalje, modeli iz literature optimirani su s ciljem postizanja visoke vrijednosti parametra AUC koji je bio prvi kao i BA koji je bio drugi kriterij vrednovanja kvalitete modela u sklopu natjecanja u predviđanju toksičnosti održanom 2014 godine (https://ncats.nih.gov/news/releases/2015/tox21-challenge-2014-winners, preuzeto 01. travnja 2021. godine). Svi timovi koji su sudjelovali u natjecanju u završnom predviđanju na test skupu imali su pravo poslati tri predviđanja. To znači da njihovo predviđanje nije bilo posve čisto (realistično) jer su kroz prve dvije predikcije mogli korigirati njihove modele (npr. optimiranjem praga odluke između toksičnih i netoksičnih spojeva u završnom koraku

predviđanja) s ciljem dobivanja viših vrijednosti metrika AUC i BA. S druge strane, modeli 12 toksičnih učinaka skupova iz baze Tox21 razvijeni u disertaciji optimirani su primarno s obzirom na metriku MCC. Stoga, tako izabrani modeli neće nužno biti i optimalni prema materici BA. Međutim, vrijednosti metrika BA svih 12 Tox21 modela iz disertacije usporedive su s vrijednostima BA drugih modela iz literature prema tablici 27.

Tablica 27. Rezultati modeliranja iz literature prema uravnoteženoj točnosti (BA) na skupu za vrednovanje. U ovom radu skup za vrednovanje višestruko je veći (~ 20% ukupnih spojeva) nego kod ostalih metoda gdje je bio ~ 647 spojeva (ili manje). „*“ označava najbolji model.

Uravnotežena točnost	**Barta**[205]	**Drwal**[201]	**Banerjee**[197]	**Idakwo**[172]	**Abdelaziz**[196]	**Natjecanje**[196]	**Zhang FP**[99]	**Zhang MD**[99]	**Ovaj rad**
NR-AR	0,61	0,60	*0,86	0,64	0,75	0,74	0,80	0,79	**0,79**
NR-Ahr	0,56	0,82	*0,91	0,82	0,86	0,85	0,83	0,85	**0,75**
NR-AR-LBD	0,49	0,66	0,83	0,61	0,60	0,65	0,84	*0,85	**0,73**
NR-Aromatase	0,56	0,75	*0,89	0,73	0,76	0,74	0,77	0,79	**0,69**
NR-ER	0,66	0,61	*0,91	0,792	0,74	0,75	0,72	0,73	**0,64**
NR-ER-LBD	0,59	0,73	*0,89	0,70	0,76	0,72	0,80	0,79	**0,67**
NR-PPAR-gamma	0,55	0,58	*0,85	0,75	0,76	0,79	0,76	0,77	**0,67**
SR-ARE	0,52	0,67	*0,87	0,85	0,71	0,73	0,76	0,80	**0,70**
SR-ATAD5	0,61	0,69	*0,84	0,71	0,73	0,74	0,80	0,80	**0,68**
SR-HSE	0,56	0,81	*0,86	0,67	0,77	0,80	0,71	0,74	**0,66**
SR-MMP	0,69	0,75	*0,91	0,853	0,90	0,90	0,85	0,88	**0,77**
SR-p53	0,58	0,76	*0,89	0,78	0,76	0,77	0,78	0,81	**0,67**

Usporedbe s modelima iz literature prema metrici (parametru) MCC u tablici 28 pokazuje visoku kvalitetu 12 Tox21 modela iz disertacije u usporedbi s onim modelima iz literature [99,172,199] za koje je iz objavljenih i dostupnih podataka bilo moguće izračunati pripadajuće vrijednosti MCC.

Tablica 28. Rezultati modeliranja iz literature prema MCC na skupu za vrednovanje. U ovom radu je skup za vrednovanje višestruko veći (~20 %).

MCC	Abdelaziz[196]	Idakwo[172]	Zhang FP[99]	Zhang MD[99]	Ovaj rad
NR-AR	0,25	0,29	0,50	0,48	***0,71**
NR-Ahr	0,51	0,53	0,52	*0,58	**0,46**
NR-AR-LBD	0,09	0,16	0,60	0,62	***0,63**
NR-Aromatase	0,26	*0,47	0,28	0,33	**0,43**
NR-ER	0,34	*0,56	0,37	0,38	**0,39**
NR-ER-LBD	0,22	0,34	*0,45	0,41	**0,44**
NR-PPAR-gamma	0,24	*0,38	0,32	0,27	**0,32**
SR-ARE	0,36	*0,62	0,46	0,50	**0,37**
SR-ATAD5	0,27	0,34	0,37	*0,39	**0,31**
SR-HSE	0,23	0,26	0,31	*0,33	**0,28**
SR-MMP	0,59	0,55	0,63	*0,69	**0,57**
SR-p53	0,31	0,39	0,42	*0,46	**0,29**

Velik broj radova koristi kombinacije prediktorskih skupova (strukturnih otisaka, molekulskih deskriptora, maccs otisaka). [99,187,196,197,200–202,204,205] Većina navedenih autora koristi više od 1000 prediktorskih varijabli za učenje modela. Pojedini autori koriste konvolucijske neuronske mreže (engl. *Convolutional neural networks*) koje uče na dvodimenzionalnim slikama kemijskih spojeva. [93,198] Uz neuronske mreže [84,196] koriste se i algoritam nasumičnih šuma [197,202,204] i druge ansambl metode [99,200,205] te metoda potpornih vektora. [197,204] Pojedini autori koriste sve navedene metode i mnoge druge. [187] Potrebno je naglasiti da je pobjednik natjecanja "Tox21 challenge" [199], grupa Mayr sa suradnicima. pobijedila neuronskim mrežama koristeći strukturne otiske i učenje na višestrukim ciljnim varijablama (engl. *multilabel classification*) umjesto individualnih modela pojedinačno po toksičnom učinku. [84]

U usporedbi s modelima iz literature gdje se nekad koristi velik broj deskriptora i softverskih alata za računanje prediktorskih varijabli te se modelira velik broj modela kao što je to 1000 modela u radu [196], 5000 modela u [187], u ovom radu je trenirano svega 5 modela po toksičnom učinku iz baze Tox21. Također su u ovom radu korišteni samo softverski alat otvorenog koda, biblioteka RDKit [173] sa skupom od ~200 fizikalno-kemijskih deskriptora i molekulskim

otiscima (5120), dok je razvidno da pojedini autori koriste sve raspoložive softvere za računanje prediktorskih varijabli (primjerica MOE, maccs, Toxprint, Dragon, Mold2, Mordred). [187,201] Tako je u redu Abdelaziz i sur.[196] uporabljeno 10 programa za računanje strukturnih deskriptora (prediktorskih varijabli) te je uz varijacije algoritma generirano za svaki toksični učinak preko 1000 modela. Takvim pristupom dobiveni rezultati nisu značajno bolji, a iza takvoga postupka krije se mogućnost dobivanja preoptimističnih rezultata uzrokovanih isrpobavanjem i uporabom velikog broja mogućih kombinacija (u optimizaciji i razvoju) modela u konačnom predviđanju. Za usporedbu je također ključna veličina skupa za vrednovanje koji je u ovom radu veći u odnosu na radove koji su koristili podjelu iz natjecanja. S obzirom na sve dano i prema rezultatima iz tablica 27 i 28 razvidno je da su modeli iz ovog rada sličinih kvaliteta kao i modeli iz literature, posebice prema MCC gdje su modeli ovog rada 2/12 učinaka pokazali najbolji rezultat (NR-AR, NR-AR-LBD u tablici 28). Iz rezultata uravnotežene točnosti u tablici 27 razvidno je da postoji dominacija modela iz rada Banerjee i sur. u 11/12 toksičnih učinaka [197]. Iz rada nije razvidna kako su postignute tako visoke vrijednosti parametara kvalitete jer se prediktorski skupovi i klasifikacijski algoritmi ne razlikuju značajno od preostalih modela iz literature.

Ono što je vrijedno spomena jest činjenica da su modeli u toj studiji objavljeni 2018., tj. dvije godine nakon drugih modela koji su optimirani tijekom prediktivnog natjecanja 2014. (a objavljeni su 2016. godine). To upućuje na mogućnost skrivenog pretjeranog optimiranja (isprobavanjem raznih opcija i metoda modeliranja) kvalitete modela prema vanjskom skupu spojeva koji je bio umnogome sličan onom s prediktivnog natjecanja. U radu Banerjee i sur. konačni modeli predstavljaju ansamble velikog broja RF i SVM modela. [197] Nadalje, vanjski skup spojeva u Banerjee i sur. [197] na kojem je provedena usporedba s modelima iz literature i iz disertacije (tablica 27) bio je tri (2,8 – 3,2) puta manji nego odgovarajući vanjski skupovi molekula korišteni u disertaciji za provjeru kvalitete 12 modela toksičnosti koji se odnose na skup Tox21. Nadalje, za svaku od 12 toksičnih učinaka razvijeno je u disertaciji po pet modela, što je daleko manje nego kod drugih metoda iz tablice 27, što ukazuje na daleko manju razinu slučajne korelacije kao posljedica isprobavanja velikog broja mogućih kombinacija i optimizacijskih parametara u metodama koje se koriste za razvoj modela, ili isprobavanja različitih skupova deskriptora odnosno podjela skupova molekula u postupku modeliranja.

Potrebno je napomenuti da na kvalitetu modela osim podjele skupova (što je u ovom radu učinjeno značajno strože nego u odgovarajućim analizama iz literature) utječe i način penalizacije. Modeli u disertaciji penalizirani su putem MCC (Poglavlje 2.7). Iz svih navedenih literaturnih izvora nije jasno jesu li (i na koji način) modeli penalizirani.

Dobri rezultati usporedbe 12 Tox21 modela razvijenih u disertaciji s modelima razvijenih od više timova iz literature pokazuju da su ispravno odabrani skupovi deskriptora (prediktoskih varijabli) i da je odabrana ispravna metodologija (algoritmi) u razvoju, optimizaciji i validaciji QSAR modela toksičnosti kemijskih spojeva.

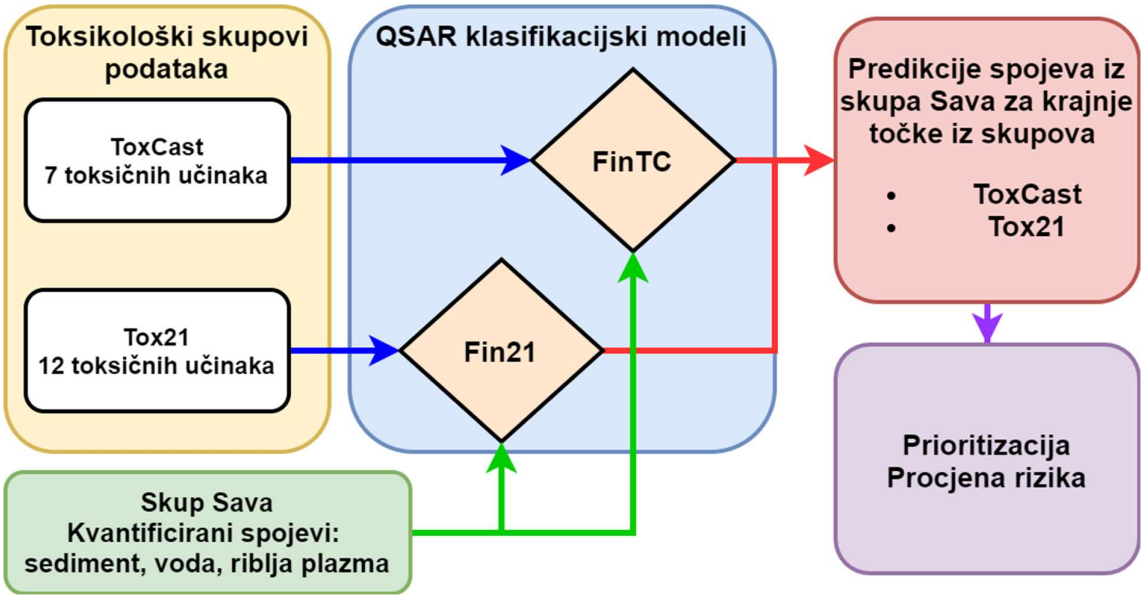
Za razliku od skupa Tox21, skup ToxCast koji se odnosi na 19 toksičnih učinaka proveden na zebricama [22] nije bio predmet velikog broja radova u području QSAR-a. U literaturi je nađen jedan rad [207] koji je koristio sirove podatke iz eksperimentalnih mjerenja na zebricama[22] koji su dani kao minimalna koncentracija učinka (engl. *lowest effect concentration*) i stoga nisu za usporediti s modelima iz ovog rada. Za potrebe ovog rada koriütena su provjereni podaci kao binarni izlaz. Međutim razvidno je da Garg i sur.[207] nisu napravili potpuno QSAR istraživanje s objavom rezultata validacije modela nego je rad samo deskriptivne (ne i prediktivne) prirode i opisuje odnose s nekoliko deskriptora poput log*P*.

Nekoliko radova predstavlja QSAR modele na podskupovima ToxCast baze podataka koji su također modelirani na eksperimentalim podacima sa zebricama ali nisu iz skupova podataka korištenih u ovom radim, nego na skupu podataka *NHEERL_ZF_144hpf_TERATOSCORE.* Navedeni skup, skraćeno *NHEERL_ZF*, koji je izvorno opisan u radu [208], koristi embrije zebrica da bi se ocjenio toksični učinak 309 spojeva iz okoliša (tzv ToxCast Faza 1), koji su pretežito pesticidi. Dodatna razlika ogleda se i u tome što je toksični učinak u navedenom skupu proveden u dužem vremenskom periodu testiranja (144 sata nakon oplodnje) i predstavlja kumulativno bodovanje svih opservacija. Za razliku od toga skupa, toksični učinci koji su modelirani u disertacijibitno su osjetljiviji jer se odnose na pojedine razvoje poremećaje i mortalite embrija zebrice, a sva mjerenja provedena su uz značajno niže koncentracije kemijskih spojeva. Opisana su dva modela razvijena na *NHEERL_ZF* testu u literaturi [209,210]. Međutim, obje su studije uključivale samo jednu krajnju točku toksičnosti zebrica (tzv. *teratoscore*). U usporedbi s QSAR modelima toksičnosti spojeva na embrije zebrice razviu ovom radu koji su razvijeni na skupu od preko 1000 spojeva, ovi modeli iz literature [208-210] razvijeni su i validirani na skupovima podataka s manje od 300 konačnih spojeva. Pokazali su razumnu kvalitetu predviđanja za samo-definirane granične vrijednosti toksičnosti, s vrijednostima parametra MCC na vanjskim (test) skupovima spojeva od 0,89 (n = 58) [209] i 0,77 (n = 61) [210].

5.7 Prioritizacija spojeva na temelju najboljih QSAR modela toksičnosti

Dobiveno je 7 konačnih ToxCast (FinTC) modela (Poglavlje 5.6.10) i 12 konačnih Tox21 (FinT21) modela (Poglavlje 5.6.1) koji su zadovoljili potrebne kriterije kvalitete. Osim što su Fin21 modeli korišteni za toksikološku evaluaciju kemijskih spojeva, oni su iskorišteni i za

računanje biodeskriptora. Na temelju ekstrapolacije modela FinTC i Fin21 na skup Sava, tj. na temelju predviđanja toksičnosti spojeva iz skupa Sava modelima FinTC i Fin21, izvršena je prioritizacija smjesa kemijskih spojeva iz skupa Sava. Prioritizacija je provedena pojedinačno za svaki spoj zasebno, ali i za postaje koje su smjese kemijskih spojeva. Shema postupka prikazana je na slici 39.



Slika 39. Shematski prikaz metodologije ovog rada. Najbolji odabrani modeli koriste se za ekstrapolaciju (predviđanje toksičnosti) i procjenu rizika na skupovima spojeva detektiranih u rijeci Savi.

Cilj istraživanja bio je dobiti modele za procjenu toksičnosti koji pokrivaju širi kemijski prostor i koji su razvijeni na točnijim i osjetljivijim mjerenjima toksičnosti nego je to slučaj s postojećim popularnim alatima za procjenu rizika koji su opisani u Poglavlju 3.4. Nadalje, cilj je bio i izgraditi modele na skupovima spojeva koji su izvan domene tradicionalnih (sigurnih) onečišćivala, i tako preseliti procjenu rizika u kemijski prostor KORZ stvarno detektiranih u rijeci Savi. U svrhu pripreme podataka spojevi su podvrgnuti postupcima pripreme SMILES strukture opisanima u Poglavlju 4.4. Zatim su ti spojevi podvrgnuti provjeri zastupljenosti u kemijskom prostoru u Poglavlju 5.5 te je provedena ekstrapolacija (predviđanje) najboljim odabranim modelima (FinTC, Fin21).

5.7.1 Predviđanje (ekstrapolacija) toksičnosti na skupu Sava najboljim QSAR modelima

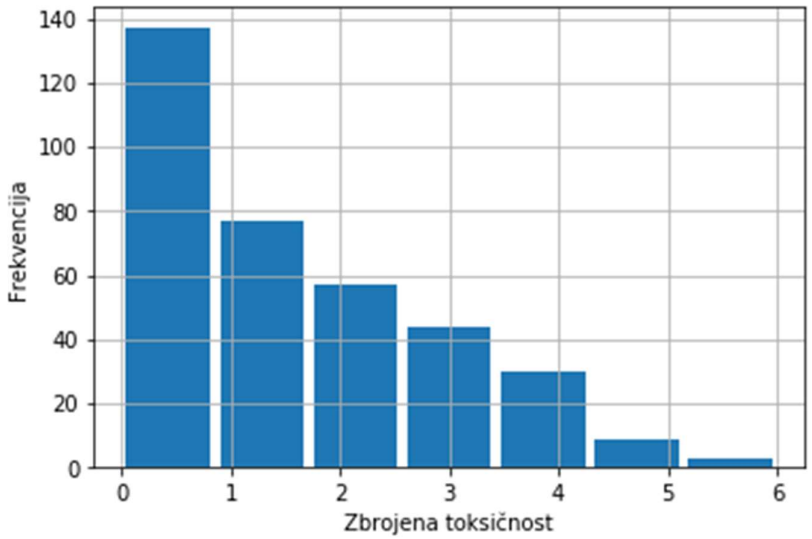
U ovom su poglavlju prikazani rezultati predviđanja ishoda s 19 QSAR modela na spojevima iz skupa Sava. Rezultati predviđanja za pojedinačne spojeve skupa Sava priloženi su u elektroničkom dodatku E9. U tablici 29 prikazani su kumulativni rezultati predviđanja toksičnosti (ekotoksikoloških učinaka) svih spojeva iz skupa Sava za sedam toksičnosti dobivenih modelima razvijenim na podacima iz baze ToxCast (modeli FinTC). Značenje kratica za sedam toksičnosti dano je u Poglavlju 3.3.1. Iako je problem primarno klasifikacijske prirode, rezultati modeliranja mogu se prikazati i kao vjerojatnost pripadnosti toksičnoj klasi (klasa 1), što je priloženo u elektroničkom dodatku E9. Prikaz toksičnosti po spojevima dan je na slici 40 u obliku histograma.

Postoji velika diskrepancija između pojedinih toksičnih učinaka. Toksičnosti s najmanje toksičnih spojeva prema modelima su MORT i EYE s manje od 5 % aktivnih na skupu Sava, dok je prema predviđanjima tri toksičnosti (ActivityScore, AXIS i PE) aktivno 25 % spojeva. Iz slike se vidi da je većina spojeva prema modelima netoksična (38 % ili 137 spojeva), tj. molekule su netoksične prema predviđanjima svih sedam FinTC modela.

Tablica 29. Kumulativni rezultati predviđanja toksičnosti spojeva skupa Sava dobivenih sa sedam modela razvijenih na podacima skupa ToxCast

Toksični učinak	Neaktivno	Aktivno
ActivityScore	206	151
AXIS	232	125
PE	246	111
YSE	297	60
JAW	317	40
EYE	342	15
MORT	353	4

Nadalje, 21 % ili 77 spojeva predviđeno je kao toksično prema samo jednom od sedam modela. Dobiveni rezultati upućuju na ispravnost dobivenih sedam modela toksičnosti temeljnih na podacima iz baze ToxCast.



Slika 40. Kumulativni rezultati toksičnosti prikazani kao frekvencijski dijagram za 357 spojeva iz skupa Sava. Iz slike je vidljivo da je većina spojeva netoksična ili minimalno toksična prema sedam FinTC modela.

Tek je mali broj spojeva "iznimno toksičan" prema sedam promatranih modela te bi pri malim koncentracijama uzrokovao smrt embrija zebrice, ili ozbiljnije deformacije. Kvantificirani spojevi u ovom radu su velikim dijelom FADM, koji su u ljudskoj uporabi. Mnogi spojevi iz skupa Sava prošli su potrebna ispitivanja nadležnih regulatornih tijela u Republici Hrvatskoj i EU u smislu prihvatljivosti njihove primjene kao pesticida, herbicida, lijeka, itd. Stoga, u malim koncentracijama oni bi trebali biti neškodljivi (ili malo škodljivi) po žive organizme u okolišu. Međutim, oni kumulativno, iako u niskim koncentracijama, mogu prouzročiti mjerljiv toksični učinak na organizme u tom onečišćenom okolišu. Potrebno je naglasiti kako je raspon koncentracija za ispitivanje toksičnosti na skupu ToxCast nizak (0,0064 – 64 μM), dok su mnogi poznati ekotoksikološki modeli razvijeni i optimirani na ekotoksikološkim učincima izmjerenim pri osjetno višim koncentracijama (mM). [153,211] Takvi učinci daju određenu informaciju o toksičnosti, ali pri koncentracijama koje ne odgovaraju svakodnevnim situacijama u okolišu u kojima uočavamo i pokušavamo izmjeriti i druge učinke kemijskih spojeva u okolišu pri nižim koncentracijama koji, pritom, nisu letalni po organizam koji se promatra (embrij zebrice u slučaju baze ToxCast). Ti su subletalni učinci povezani s jasno mjerljivim oštećenjima embrija koji predstavljaju mjerljivi učinak koji proizvodi izmjerena niska koncentracija kemikalije u kojoj ona stvarno može biti prisutna u okolišu.

Stoga je ukupno ~59 % ili 214 spojeva prema sedam najboljih ToxCast modela minimalno toksično ili netoksično (zbrojena toksičnost $N \leq 1$, slika 40). U nastavku je navedeno nekoliko

istaknutih spojeva za koje je predviđeno da su toksični prema pet ili više ($N \geq 5$) toksičih učinaka modela FinTC (od njih ukupno sedam).

Prema ukupnom broju predviđenih toksičnosti modelima FinTC (od ukupno sedam mogućih) prednjače ovi spojevi: miklobutanil (pesticid), propikonazol (pesticid), ciprokonazol (pesticid) po šest, azakonazol (pesticid), dimetomor (pesticid), haloperidol (farmaceutik), difenhidramin (farmaceutik), fenamidon (pesticid), fenbukonazol (pesticid), trifloksistrobin (pesticid), unikonazol-p (pesticid), imazalil (pesticid) po pet. Iz popisa je razvidno da su 10 od 12 spojeva pesticidi.

Kumulativni rezultati predviđanja toksičnosti spojeva skupa Sava s 12 modela razvijenih na podacima iz baze Tox21 prikazani su u tablici 30. Značenje kratica za 12 toksičnosti dano je u Poglavlju 3.3.2. Prvih pet toksičnosti s manje od 5 % aktivnih spojeva na skupu Sava prema predviđanjima redom su NR-ER-LBD, NR-ER, NR-AR, NR-AR-LBD, SR-ATAD5, SR-HSE te NR-PPAR-gamma, dok je 25 % spojeva iz skupa Sava predviđeno kao toksično prema toksičnom ishodu SR-ARE.

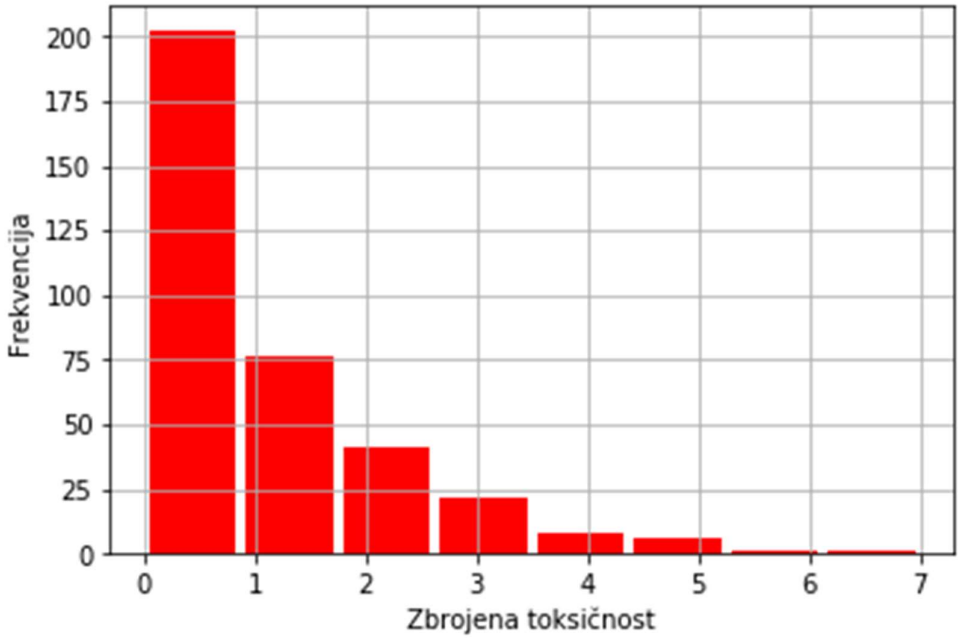
Tablica 30. Kumulativni rezultati predviđanja toksičnosti spojeva skupa Sava modelima razvijenim na skupu Tox21 (Fin21 modeli) za 12 toksičnih učinaka.

Toksični učinak	Neaktivno	Aktivno
SR-ARE	263	94
NR-AhR	293	64
NR-Aromatase	321	36
SR-MMP	323	34
SR-p53	330	27
NR-PPAR-gamma	342	15
SR-HSE	346	11
SR-ATAD5	350	7
NR-AR-LBD	352	5
NR-AR	353	4
NR-ER	355	2
NR-ER-LBD	357	0

Prikaz predviđenih toksičnosti zbrojeno po svim spojevima za svaki toksični ishod dan je na slici 41 u obliku histograma. Iz slike se vidi da je većina spojeva (57 % ili 202 spoja) prema modelima netoksična (zbrojena toksičnost $N = 0$, slika 41). Nadalje, samo jedan toksični učinak

od 12 modeliranih (N = 1) ima 21 % ili 76 spojeva. Stoga je ukupno ~ 78 % spojeva prema Tox21 modelima netoksično ili minimalno toksično (N ≤ 1). Ovo je, kao i u slučaju ranije opisanih predviđanja sa sedam modela razvijenih na bazi ToxCast, dodatna potvrda ispravnosti 12 QSAR modela toksičnosti razvijenih u ovim istraživanjima na skupu kemijskih spojeva i njihovih 12 izmjerenih toksičnosti iz baze Tox21. Naime, kao i u slučaju sedam modela toksičnosti razvijenih na skupu podataka baze ToxCast, spojevi iz skupa Sava odobreni su za uporabu nakon provjere i odobrenja regulatornih tijela, i u malim koncentracijama u okolišu ne bi trebali biti toksični za stanične linije iz baze Tox21.

Ovdje je navedeno nekoliko istaknutih spojeva iz skupa Sava koji su predviđeni s pomoću 12 Tox21 (Fin21) modela kao toksični rangirani prema zbrojenoj toksičnosti: digitoksin 7 (farmaceutik), boskalid 6 (pesticid), kloroksuronimazalil(i), triadimenol (pesticid), digoksin (farmaceutik), pentaklorofenol (pesticid) i triklozan po pet (antiseptik), 7-aminoklonazepam (farmaceutik), indoksakarb (pesticid), forklorfenuron (pesticid), avermektin b1a (pesticid), avermektin b1b (pesticid), desmedifam (pesticid), kanabinol (droga) i kvinoklamin (pesticid) po četiri. Iz popisa razvidno je da se pretežito radi o pesticidima.



Slika 41. Kumulativni rezultati toksičnosti prikazani kao frekvencijski dijagram za 357 spojeva iz skupa Sava. Iz slike je vidljivo da je većina spojeva netoksična ili minimalno toksična prema 12 Fin21 modela.

Uz prioritizaciju je pojedinačno po skupovima ToxCast i Tox21 napravljena i kumulativna prioritizacija na temelju zajedničkih rezultata kako slijedi:

Uspoređene su dvije metode prioritizacije spojeva prema modeliranoj toksičnosti:

(1) prema zbroju pripadnosti pozitivnoj klasi nekog toksičnog učinka (1 ili 0, prebrojavanjem) koja je označena kraticom ZB (od Zbroj Binarnih) te

(2) prema prosječnoj vjerojatnosti (PV) pripadnosti pozitivnoj klasi uprosječeno za svih 19 predviđanja toksičnosti. što daje broj na skali 0 – 1 (npr. kad je vjerojatnost iznad 0.5 spoj se pripisuje pozitivnoj klasi).

Radi lakšeg razumijevanja, u tablici 31 dan je primjer zbrajanja spomenutih rezultata modeliranja na tri primjera.

Tablica 31. Primjer računanja PV i ZB. Za tri kemijska spoja predviđene su vrijednosti za dva toksična učinka u obliku vjerojatnosti (na skali 0 - 1.0), pri čemu kad je vjerojatnost viša od 0.50 spoj se klasificira kao aktivan (binarno 1). PV računa prosječne vjerojatnosti a ZB zbraja binarne vrijednosti predviđanja.

	Toks. učinak 1 (vjeroj.)	**Toks. učinak (binarno), ZB**	**Toks. učinak 2 (vjeroj.)**	**Toks. učinak 2 (binarno)**	**Prosj. vjeroj. (PV)**	**Zbroj binarnih (ZB)**	**Rang**
Kemijski spoj 1	0,51	2	0,60	1	0,555	3	1
Kemijski spoj 2	0,49	1	0,37	0	0,430	1	2
Kemijski spoj 3	0,20	0	0,72	1	0,460	1	3

Prema shemi iz primjera u tablici 31 za potrebe prioritizacije KORZ čija je prisutnost utvrđena (kvantificirana) u rijeci Savi (skup Sava) spojevi su najprije pojedinačno rangirani na temelju rezultata predviđanja pojedinačnih toksičnih učinaka prema oba skupa modela (FinTC – 7 konačnih modela i Fin21 – 12 konačnih modela, ukupno 19 modela). Zatim su pojedinačni rangovi usrednjeni (rezultati su u tablici 32).

Tablica 32. Kumulativni rezultati predviđanja toksičnosti modelima FinTC i Fin21 spojeva skupa Sava za 20 najaktivnijih spojeva. U tablici su prikazane brojnosti (N) spojeva koji su prema predviđanjima klasificirani kao pozitivni, njihovi rangovi prema predviđanjima modelima ToxCast (FinTC) i Tox21 (Fin21), te prosječni rang svakog spoja prema toksičnosti.

KORZ[a]	Kategorija	ToxCast (PV)	ToxCast (ZB), N	Tox21 (PV)	Tox21 (ZB), N	Prosječni rang [b]
imazalil	pesticid	0,57	5	0,47	5	3,5
triadimenol	pesticid	0,56	4	0,38	5	8
propikonazol	pesticid	0,66	6	0,31	3	9
azakonazol	pesticid	0,56	5	0,28	3	10,5
unikonazol-p	pesticid	0,57	5	0,30	3	10,5
prazepam	farmaceutik	0,45	4	0,29	3	15
fenamidon	pesticid	0,59	5	0,30	2	21,5
kloroksuron	pesticid	0,40	3	0,33	5	23
loratadin	farmaceutik	0,38	4	0,24	2	26
terbukonazol	pesticid	0,42	4	0,23	2	26
citalopram	farmaceutik	0,41	4	0,20	2	26
desmedifam	pesticid	0,45	3	0,26	4	26
7-aminoklonazepam	farmaceutik	0,42	3	0,38	4	26
izoksaben	pesticid	0,48	4	0,23	2	26
tritikonazol	pesticid	0,50	4	0,25	2	26
klorazepat	farmaceutik	0,39	3	0,30	3	30
flutriafol	pesticid	0,52	3	0,30	3	30
pikolinafen	pesticid	0,50	3	0,27	3	30
ciprokonazol	pesticid	0,63	6	0,21	1	40,5
miklobutanil	pesticid	0,65	6	0,29	1	40,5

[a] KORZ - kemijska onečišćivala od rastućeg značaja za okoli**š;** [b] Manji broj (rang) znači da je spoj toksičniji; [c] Kod spoja imazalil broj 5 u stupcu „ToxCast (ZB), N“ znači da je od ukupno 7 FinTC modela njih 5 predvidjelo da je spoj toksičan. Nadalje, za isti spoj broj 5 u stupcu „Tox21 (ZB), N“ znači da je od ukupno 12 Fin21 modela njih 5 predvidjelo da je ovaj spoj toksičan.

Iz tablice 32 je razvidno da su 15 od 20 spojeva pesticidi, dok su preostalih pet farmaceutici. Prema saznanjima autora u literaturi nema primjera korištenja ToxCast modela na zebricama u svrhe prioritizacije i procjene ekotoksikološkog rizika.

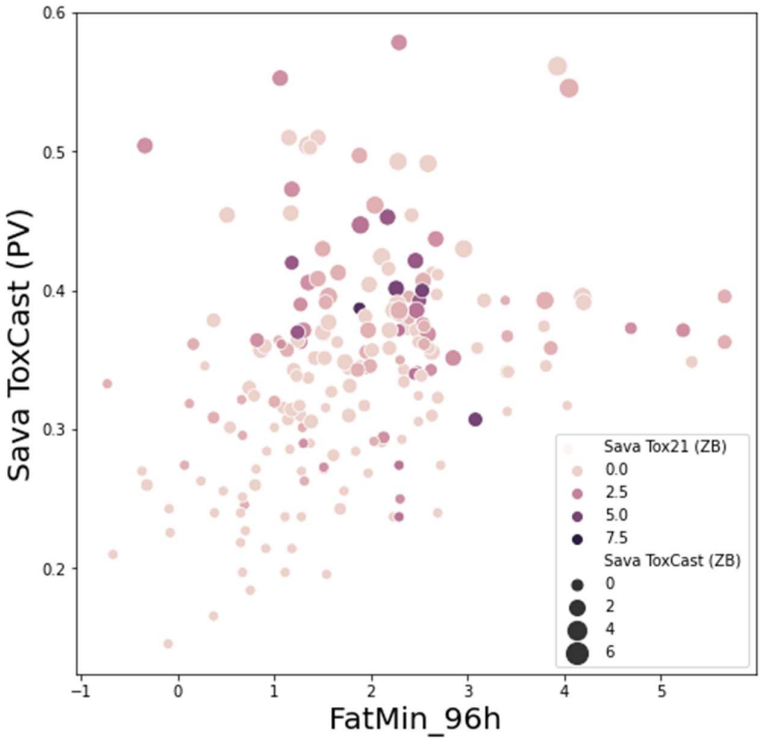
5.7.2 Usporedba predložene prioritizacije s prioritizacijom temeljnom na modelima VEGA-QSAR

Prve metode prioritizacije prikazane u ovom radu temelje se na modelu iz alata VEGA-QSAR koji je predstavljen u Poglavlju 3.4. Napravljena je usporedba predviđene toksičnosti spojeva (rangiranje) za organizam *Fathead Minnow 96h* (FatMin_96h) s pomoću modela toksičnosti razvijenih i opisanih u literaturi. [148] Budući da ne postoji referentna metoda procjene (engl. *ground truth*), metode su uspoređene međusobno nakon predviđanja na molekulama iz skupa Sava. Usporedba je izvršena putem Spearmanovog koeficijenta korelacija (rang-korelacija) između vrijednosti rangova molekule iz skupu Sava dobivenih prioritizacijskim postupcima osmišljenim u disertaciji. Uspoređene su dvije metode za kumulativni prikaz rezultata modeliranja kako je opisano ranije: (1) prema zbroju pripadnosti toksičnoj klasi ZB te (2) prema prosječnoj vrijednosti (PV) vjerojatnosti pripadnosti toksičnoj klasi na temelju predviđanja s pomoću sedam modela toksičnosti temeljnih na bazi ToxCast ili s 12 modela toksičnosti temeljenih Tox21. Vjerojatnosti su brojevi na skali 0 – 1, npr. kad je vjerojatnost iznad 0,5 spoj se pripisuje toksičnoj klasi (binarno 1). Rezultati su prikazani u tablici 33.

Tablica 33. Spearmanova korelacija između prioritizacija (rangiranja) spojeva skupa Sava modelom toksičnosti FatMin_96h temeljenom na VEGA-QSAR, sedam modela FinTC i 12 modela FinT21 putem prebrojavanja toksičnosti (N) za koje je spoj predviđen kao toksičan (pozitivan, 1) i s pomoću prosječne vjerojatnosti da je spoj toksičan promatrano preko modela svih sedam FinTC odnosno 12 FinT21 toksičnih učinaka.

	ToxCast (PV)	**Tox21 (PV)**	**ToxCast (ZB)**	**Tox21 (ZB)**	**FatMin 96h**
Sava ToxCast (PV)	1	0,48	0,81	0,35	***0,39**
Sava Tox21 (PV)	0,48	1	0,36	0,78	0,14
Sava ToxCast (ZB)	0,81	0,36	1	0,33	***0,28**
Sava Tox21 (ZB)	0,35	0,78	0,33	1	0,07
FatMin_96h	0,39	0,14	0,28	0,07	1

Na slici 42 prikazan je radi lakšeg razumijevanja dijgram raspršenja odnosa predviđene kumulativne toksičnosti putem ToxCast modela i one predviđene na FatMin_96h LC50.



Slika 42. Slika sadrži tri razine informacija vezano za prioritizaciju spojeva iz skupa Sava: 1) Dijagram raspršenja predviđanja toksičnih učinaka prema modelima Sava ToxCast (PV) na x- osi i FatMin_96h (predviđene LC50 vrijednosti) na y-osi; 2) veličine oznaka (kružića) predstavljaju brojnost predviđenih toksičnih učinaka (binarno) spojeva prema rangiranju Sava ToxCast (ZB) (objašnjeno u tekstu, od 7 modela FinTC); 3) gradijent boja oznaka (kružića) označava brojnost toksičnih učinaka prema Sava Tox21 (ZB) vrijednostima (od 12 modela Fin21).

Rezultat usporedbe prioritizacije temeljene na novim QSAR modelima (FinTC i Fin21) pokazuje najvišu rang-korelaciju (po Spearmanu) s FatMin_96h za usrednjenu vjerojatnost kroz sve ToxCast toksične učinke, Sava TC (PV), od 0,39. Iduća je najviša rang-korelacija (0,28) između ranga prema VEGA-QSAR modelu FatMin_96h i ranga prema sedam modela ToxCast (ZB), tj. zbroju pozitivnih klasifikacija. Znatno slabije slaganje ranga temeljenog na VEGA-QSAR modelu FatMin_96h je sa rangovima dobivenim na 12 modela razvijenih na bazi Tox21. Rezultati predviđanja rangova molekula ovih prioritizacijskih postupaka pokazuju međusobne korelacije u rasponu 0,07 – 0,14.

5.7.3 Usporedba rezultata predviđanja FinTC modela na skupu Sava s rezultatima testova embriotoksičnosti (ZET)

Ovdje je dana usporedba ekotoksikološkog rizika ispitanih uzoraka sedimenta (Poglavlje 3.1.1) izračunata uporabom QSAR modela sa stvarnim rizikom uzoraka dobivenim primjenom testova embriotoksičnosti na zebricama (ZET) (Poglavlje 5.2.1). Ishodi ZET mapirani su s predviđenim toksičnim učincima iz FinTC modela. Vrijednosti toksičnosti molekula iz reduciranog skupa Sava (N = 313) (kvantificirani u sedimentu i vodi) dobivene su predviđanjima (aktivno = 1, neaktivno = 0) pomoću sedam FinTC modela. Dobivena matrica *Potencijal ToxCast*, skraćeno P_{TC} ima oblik 313×7. Vrijednosti matrice P_{TC} su 0 ili 1. Kako bi se prioritizirale lokacije uzorkovanja sedimenta, matrica P_{TC} pomnožena je s matricom koncentracija spojeva izmjerenih u sedimentu (C_{Sed}) za svaku lokaciju (313×1) prema Jednadžbi 5.1, što daje za pojedinu lokaciju matricu *TP* koja predstavlja toksični potencijal te lokacije. Dimenzije matrice *TP* su (313×7) za svaku lokaciju (Jesenice = J, Rugvica = R, Galdovo = G, Lukavec = L, Poglavlje 3.1.1). S obzirom da su uzorkovanja provedena na četiri lokacije (J, R, G, L) dobivene su četiri *TP* matrice dimenzija (313×7).

$$TP_{ij} = \ (P_{TC} \odot \mathrm{C_{Sed}})_{\mathrm{ij}} = (P_{TC})_{\mathrm{ij}}(\mathrm{C_{Sed}})_{\mathrm{ij}} \tag{5.1}$$

Metoda *TP* inspirirana je mjerom *TU*, opisanom u Poglavlju 4.2.1, gdje se izmjerene koncentracije spojeva dijele s vrijednosti LC50. Budući da za predviđanja dobivena klasifikacijskim modelima razvijenim u disertaciji (FinTC i Fin21) nisu dostupne informacije u obliku LC50, uzeti su binarni izlazi modela (0 ili 1). Nedostatak ovakve metode koja ne računa s kontinuiranim nego s binarnima vrijednostima stroga je granica između spojeva koji se ocjenuju kao toksični i onih koji to nisu. Stoga, ovim rezultatima nedostaje doprinos granično toksičnih spojeva. Rezultati množenja prema jednadžbi 5.1 prikazani su u tablici 34 za pojedinu lokacije (J, R, G, L) i za pojedini toksični učinak (AXIS, EYE, JAW, MORT, PE, YSE, ActivityScore). Potom su ti podaci mapirani (povezani) s rezultatima ZET (toksične učinke mjeren na embrijima zebrice) kako slijedi. Poremećaji optjecajnog sustava (POS) (edemi, nakupljanje krvi) iz ZET mapirani su s toksičnim učincima PE i YSE iz skupa ToxCast (POS → PE, YSE). Oštećenja tkiva (OT) iz ZET mapirana su s toksičnim učinkom AXIS iz skupa ToxCast (OT → AXIS). Razvojne deformacije tkiva (RDT) (oko, mozak) iz ZET mapirane su s toksičnim učinkom EYE iz skupa ToxCast (RDT → EYE).

Mortalitet iz ZET (MRT) nakon 24 h i 48 h (vidi Poglavlje 5.2.1) mapiran je s MORT iz skupa ToxCast (MRT24h → MORT, MRT48h → MORT). ZET Razvojne abnormalnosti (RA)

mapiran je s toksičnim učinkom ActivityScore kao i sa Zbrojem aktivnosti (ZA) AXIS, EYE, JAW, MORT, PE i YSE iz skupa ToxCast (RA → ActivityScore, RA → ZA).

Tablica 34. Vrijednosti toksičnog potencijala *TP* prema mjestima uzorkovanja u rijeci Savi i razvrstane prema analiziranim toksičnim učincima izračunatim s pomoću modela FinTC.[a)]

	Act.Sc.	**MORT**	**POS**	**OT**	**RDT**	**ZA**
RUGVICA (R)	5487,9	28,1	7081,0	4677,3	363,5	12461,9
JESENICE (J)	2525,7	10,6	3276	2253,8	66,4	5745,9
LUKAVEC (L)	917,9	2,1	1738,9	1089,9	103,2	3104
GALDOVO (G)	510,4	15,8	487,6	162,3	13,1	707,8

[a)] **Act.Sc. – (ActivityScore) i MORT** – toksični učinci iz skupa ToxCast; **POS** – poremećaji optjecajnog sustava; **OT** – oštećenja tkiva; **RDT** – razvojne deformacije tkiva; **ZA** – zbrojene aktivnosti

ZET su pokazali sljedeće poretke mjesta uzorkovanja prema kumulativnoj toksičnosti predviđenoj prema pojedinim značajkama (pokazateljima) toksičnosti (tablica 13, Poglavlje 5.2.1):

POS: G > L > R > J,

OT: R > J > L > G,

RDT: R > J > L > G,

RA: R > J > L > G,

MRT_{24h}: R > J > L > G,

MRT_{48h}: R > G > J > L.

Poretci mjesta uzorkovanja iz tablice 34 prema toksičnom potencijalu (*TP*) smjese spojeva pronađenim u uzetim uzorcima i prema biotoksičnostima predviđenim s pomoću sedam modela razvijenih na podacima baze ToxCast (FinTC modeli):

$TP_{ActivityScore}$: R > J > L > G,

TP_{MORT}: R > G > J > L,

TP_{POS}: R > J > L > G,

TP_{OT}: R > J > L > G,

TP_{RDT}: R > L > J > G,

TP_{ZA}: R > J > L > G.

Iz usporedbe rezultata ZET vezanih za histopatološke značajke toksičnosti (POS, OT, RDT) i mapiranih rezultata modela razvidno je da se:

a) poretci POS i TP_{POS} ne podudaraju (u rangiranju);

b) OT i TP_{OT} u potpunosti podudaraju u poretku;

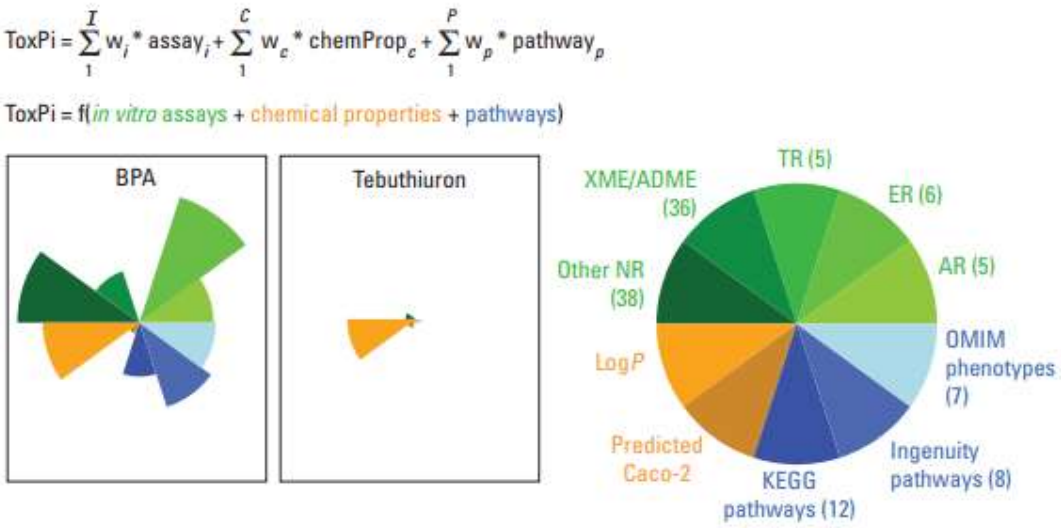
c) RDT i TP_{RDT} djelomično podudaraju gdje R (Rugvica) ima najviši rang, G (Galdovo) najniži, dok su J (Jesenice) i L (Lukavec) zamijenjeni.

Usporedba mortaliteta iz ZET i rezultata modela za mortalitet (prema učinku iz baze ToxCast – MORT) pokazuje potpunu podudarnost za ZET na MRT_{48h}, dok za MRT_{24h} odgovara samo za Rugvicu koja je rangom na prvom mjestu i u modelu i u ZET. Razvojna abnormalnost (RA) iz ZET pokazuje potpunu podudarnost s modelom ZA (zbrojene bioaktivnosti) kao i za ActivityScore. Rezultati metode *TP* pokazuju kumulativno dobru podudarnost s rezultatima ZET koji se odnose na toksične učinke na embrije zebrica. Ograničenje ovog pristupa prioritizacije toksičnosti je u broju toksičnih učinaka koji su uspoređeni. Naime za ZET se koristi više histopatoloških pokazatelja [33] dok je za usporedbu dostupan relativno mali broj toksičnih učinaka iz baze ToxCasta . Tako je za FinTC modele mapirana po jedna toksičnost za OT (→ ToxCast AXIS) i RDT (→ ToxCast EYE) dok su dva toksična učinka (→ ToxCast PE i ToxCast YSE) mapirana s POS iz ZET. Mortalitet je vrlo dobro definiran toksični učinak koji se smatra najpouzdanijim testom jer ne podliježe subjektivnoj analizi. Prednost ove metode veća je osjetljivost procjene ekotoksikološkog rizika u odnosu na metode poput onih iz alata VEGA-QSAR, jer je koncentracijski raspon testiranja toksičnosti razvijene modele temeljene na bazi ToxCast vrlo nizak (0,0064–64 μM) [22], dok su testovi za LC50 poput FatMin_96h [153,211] u rasponu koncentracija mM.

5.7.4 Usporedba prioritizacijskih metoda iz ovog rada s metodama iz literature

U mnogim radovima koriste se prioritizacijske metode za procjenu rizika po okoliš usljed prisutnosti pojedinih spojeva i njihovih mješavina u slatkim vodama. Velik broj autora koristi metodu TU (*toxic unit*) korištenu i u ovoj disertaciji [161,163,167,212–215] za procjenu rizika i prioritizaciju kemijskih spojeva. Schäfer i sur. koristili su tako TU za prioritizaciju 97 pesticida na 24 postaje u rijekama u Australiji s predviđenom toksičnošću na planktonskom račiću *Daphnia magna* i algi *Selenastrum capricornutum*, dok je u ovom radu procjena toksičnosti bila na ribi *Pimephales promelas.* Zbog nedostatnih eksperimentalnih podataka za toksičnost kemijskih spojeva također su korišteni i QSAR modeli za procjenu toksičnih učinaka. U radu Feng i sur. [213] s pomoću TU (iako nazvana *toxicity-weighted concentration*) prioritizirano je 59 spojeva kvantificiranih u riječnim estuarijima.

Mjereni spojevi su u koncentracijskom rasponu ng/L i spadaju u pesticide, farmaceutike i njihove metabolite poput spojeva priorirtiziranih u ovom radu. Mnogi autori koriste za prioritizaciju i procjenu rizika i druge podatke o toksičnostima spojeva iz baze ToxCast (ili CompTox) [216] a ne samo podatke koji se odnose na toksičnost na embrije zebrice). Primjer takvog pristupa metoda je ToxPi [217] koja profilira kemijske spojeva prema podacima dostupnim u bazi (za fazu 1, ukupno 309 spojeva) i zatim utežava i normalizira njihove doprinose te se u konačnici dobiju kružni grafovi prikazani na slici 43.



Slika 43. Prikaz računanja i prioritizacije metodom ToxPi. Pojedini čimbenici koji utječu na rizik vezan za određeni kemijski spoj uteže se i normaliziraju, zatim grafički predstave u kružnim grafovima (desno na slici) [217].

Metoda ToxPi prvenstveno je usmjerena na toksičnost koja uzrokuje endokrine poremećaje kao i na i pojedina kemijska svojstva poput log*P*. Iako je ToxPi koristan u smislu procjene načina djelovanja, ovisan je o postojećim podacima u bazi ToxCast i za procjenu velikog broja spojeva (primjerice onih za koje u bazi nema podataka o njihovoj toksičnosti) bili bi potrebni QSAR modeli kako bi se nadopunili podaci koji nedostaju.

Sličan (noviji) pristup nazvan CardioToxPi, uz veći broj toksičnih učinaka iz baze ToxCast (stanične linije) i s fokusom na kardiotoksičnost, procjenu toksičnih učinaka spojeva temelji na toksičnim učincima ispitanim na staničnim linijama (*in vitro*). [218] S obzirom na cjelokupnu strategiju metoda ToxPi i CardioToxPi po svojoj sveobuhvatnosti podsjećaju na PBTr, metodu korištenu u ovom radu za prioritizaciju koja koristi tri čimbenika: perzistentnost, toksičnost i mogućnost biokoncentracije pojedinog spoja.

Sve popularniji pristup u prioritizaciji je izračunom omjera izloženosti (engl. *exposure*) i određene biološke aktivnosti izmjerene u *in vitro* testovima toksičnosti (engl. *exposure-activity ratio*, EAR). [219,220] EAR vrijednosti stoga podsjećaju na TU koji je korišten u ovom radu. U EAR pristupu koriste se svi dostupni toksični učinci kvantificirani u bazi ToxCast (kao % aktivnosti poput AC50 (50% aktivnosti), AC10 ili arbitrarnih vrijednosti između). Metoda je u odnosu na TU, potencijalno pouzdanija jer procjenjuje toksičnost na velikom broju učinaka u bazi, međutim velika većina toksičnih učinaka temelji se na staničnim linijama pretežno sisavaca, dok je u procjeni utjecaja na okoliš važno utvrditi djelovanje na živući organizam. Koristi se također za procjenu toksičnosti većeg broja spojeva kvantificiranih u plazmi [66,221] i vodi [222].

Još jedna korištena metoda je CI (engl. *Concern Index*) [223] koja podsjeća na TU i EAR jer koristi omjer mjerenje koncentracije spoja i AC50 kroz sve dostupne toksične učinke iz baze ToxCast. Razlika je u tome što se za toksični učinak koristi vrijednost 95 percentila svih AC50 iz baze. Uz metode koje se oslanjaju na podatke iz javno dostupnih baza podataka, biološki testovi (engl. *bioassays*) su i dalje jedina stvarna mjera ekotoksikološkog stanja. Biološke metode pokrivaju embriotoksikološke testove na zebricama (ZET) [160,224] kao i biotestove na beskralježnjacima poput planktonskih račića vrste *Daphnia magna* ili bakterijskih organizama kao što je *Vibrio fischeri* i dr. [161,225]. Uz testove na cijelim organizmima često se koriste i ljudske i animalne stanice (citotoksičnost, endokrini poremećaji) [50,217].

Iz literature je razvidno da su korištene metode međusobno vrlo slične. Tako CI, TU i EAR koriste omjere toksičnih učinaka (aktivnost) i izloženosti kemijskim spojevima (koncentracija u okolišu). Razlike su pretežito u kvaliteti podataka toksičnih učinaka. Iako su EAR i CI potencijalno pouzdanije u smislu većeg broja obuhvaćenih učinaka, one ne pokrivaju dosad nekvantificirane spojeve u bazi. Ključ prioritizacija u ovom radu je upravo računanje QSAR modela kako bi se pokrili i dosad nepoznati (novi) spojevi koji su upravo u fokusu ovog rada. Na sličnom principu poput EAR, TU i CI temelji se i ovdje osmišljena metoda toksičnog potencijala *TP* (Poglavlje 5.7.3) gdje se izloženost množi s kvalificiranim (binarnim) toksičnim učinkom iz QSAR modela. Iako je u validaciju te metode uključen relativno mali broj ZET podataka na smjesama, *TP* pokazuje dobro slaganje s utvrđenim histopatološkim promjenama. Nadalje, QSAR modelima može se procjeniti toksični učinak širokog spektra spojeva. Jako je malo prioritizacijskih analiza posvećeno učincima mješavina, i jedine pouzdane procjene zasad su učinci binarnih mješavina [225] te biološki testovi na kompleksnim mješavinama. Pojedini autori sugeriraju da aditivni pristup pokazuje dobro slaganje s biološkim testovima [226], ali

potrebna su dodatna istraživanja toksičnih učinaka mješavina spojeva i ispitivanju slaganja računalnih procjena toksičnosti s rezultatima sveobuhvatnih bioloških analiza.

§ 6. ZAKLJUČAK

U radu je korišten cjelovit pristup procjene ekotoksikološkog rizika vodenih staništa za ekosustav rijeke Save. Utvrđeno je da je rijeka Sava (sediment, voda i riblja plazma), kontaminirana s 358 bioaktivnih onečišćivala antropogenog porijekla. KORZ (kemijska onečišćivala od rastućeg značaja za okoliš) se za razliku od tradicionalnih kemijskih onečišćivala ne prate redovitim kemijskim analizama stanja vode. Kemijski spojevi kvantificirani u sedimentu pretežno su blago hidrofobni i male molekulske mase. S obzirom na izmjerenu koncentraciju i toksičnost predviđenu modelima razvijenim u disertaciji, neki od detektiranih spojeva označeni su toksičnima (prema pojedinim toksičnim učincima) te predstavljaju veliki rizik za vodene organizme. Čak 90 FADM (farmaceutici, droge te metaboliti farmaceutika i droga) nađeno je i kvantificirano u ribljoj plazmi. Koristeći Model riblje plazme pomoću humane terapeutske koncentracije u plazmi i omjera učinka između riba i čovjeka, u ovom je radu izračunat potencijalni učinak na ribe. Aktivni spojevi mogu pokazati iste toksične učinke u ribama kao i kod ljudi, zbog evolucijskog očuvanja ciljnih proteina. Omjer učinka, koji je definiran kao omjer minimalne koncentracije u ljudskoj plazmi na najmanjoj terapijskoj dozi lijeka/spoja s izmjerenim koncentracijama u ribljoj plazmi, pokazuje visok stupanj aktivnosti za pojedine FADM.

Budući da se postojeće metode za procjenu rizika temelje na QSAR modelima razvijenima na malom broju spojeva, u ovom su radu razvijeni novi QSAR modeli toksičnosti temeljeni na nekoliko tisuća spojeva šire kemijske raznolikosti preuzeti iz javnih toksikoloških baza ToxCast (1018 spojeva i 19 toksičnih učinaka na *D. rerio*) i Tox21 (8144 spojeva i 12 izmjerenih staničnih toksičnih učinaka). Zadani skupovi sadrže binarne podatke o toksičnom učinku (0 ili 1). Utvrđeno je da su skupovi neuravnoteženi. Pristupilo se stoga posebnom postupku modeliranja i izabrane su odgovarajuće metrike za što objektivniju procjenu kvalitete modela poput Cohenove Kappe, uravnotežene točnosti i Matthewsovog korelacijskog koeficijenta. Uveden je i novi parametar ΔQ_2 koji daje informaciju o doprinosu modela ukupnoj toksičnosti binarne klasifikacije kad se od ukupne točnosti oduzme nasumična točnost. Nasumična točnost često može biti visoka kad se model razvija na skupovima s neuravnoteženim brojevima elemenata (molekula) u jednoj i u drugoj klasi toksičnog učinka. Razvijeni su klasifikacijski modeli na temelju neuronskih mreža, algoritma nasumičnih šuma i logističke regresije uz optimizaciju širokog parametarskog prostora kako bi se postigla što točnija klasifikacija.

Ukupno su u ovom radu razvijena 193 modela, 133 na bazi skupa ToxCast i 60 na bazi skupa Tox21. Uz širok hiperparametarski prostor algoritama ispitani su i doprinosi raznih skupova deskriptora poput strukturnih otisaka, fizikalno-kemijskih deskriptora (koji su izračunani programima drugih autora) te bioloških deskriptora (osmišljenih i izračunanih po prvi puta u ovoj disertaciji). Modeli su optimirani deseterostrukom križnom validacijom unutar skupova za učenje (optimizaciju), koji su činili 75 % (za svaki toksični učinak) kemijskih spojeva, dok je kvaliteta modela testirana na skupovima za učenje i za vanjsko vrednovanje (20 % spojeva, podataka). Konačni izbor modela proveden je s nakanom primjene na kemijskim spojevima iz vanjskog skupa Sava - za buduću primjenu u prioritizaciji i procjeni ekotoksikološkog rizika za vodeni okoliš. Sedam QSAR modela razvijenih za toksičnosti spojeva (mjerene na embrijima zebrice) iz baze ToxCast i 12 QSAR modela za toksičnosti spojeva (mjerene na staničnim linijama) iz baze Tox21 ispunili su kriterije prihvatljivih modela prema metodi bodovanja uvedenoj u disertaciji u analogiji sa sličnim postupcima iz literature. Ispostavilo se da su svi modeli razvijeni na skupu molekula i toksičnosti iz baze Tox21 osjetno lakši za modeliranje nego toksičnosti na skupu molekula iz baze ToxCast. Svi parametri kvalitete (točnosti) za Tox21 modele viših su vrijednosti, iako su dobiveni na 5-6 puta većim skupovima molekula. Kriteriji bodovanja definirani u disertaciji zahtijevaju od modela prihvatljivu točnost na skupu za učenje i u predviđanju na vanjskom skupu. Uvedeni kriteriji bodovanja kvalitete modela postroženi su na taj način da je pridodana veća težina parametru koji iskazuje točnost modela na vanjskom (test) skupu spojeva. Nadalje, ukupno odabranih 19 najboljih modela pokazuje točnost predviđanja na vanjskom skupu spojeva koja je iznad nasumične, što je u disertaciji uvedeno po prvi puta kao kriterij izbora modela. Svih 19 odabranih modela temeljeno je na molekularnom strukturnim otiscima kao deskriptorima, što omogućuje njihovu jednostavnu primjenu na velikom skupu novih spojeva, jer nije potrebno provoditi postupke normiranja deskriptora. Iako je mehanistička interpretacija ovakvih kompleksnih modela teška, analizom permutacijske važnosti prediktorskih varijabli (deskriptora) utvrđeno je da su u modelu najvažniji očekivani fragmenti molekula poput vezanih heteroatoma i halogena na molekulama, fragmenti halogena vezanih na aromatski prsten, piridinski prsten te NH i NH_2 skupine.

Rezultati modeliranja pokazuju dobru generalizaciju na vanjskim skupovima za pojedini toksični učinak. Na temelju 19 odabranih najboljih QSAR modela toksičnosti razrađena je originalna metodologija procjene rizika i prioritizacija molekula na temelju njihovog toksičnog učinka. Razvijeni postupci prioritizacije primijenjeni su na procjenu toksičnog potencijala smjesa više od 350 kemijskih spojeva (KORZ) detektiranih u sedimentu, vodi i krvnoj plazmi riba na četiri postaje u rijeci Savi (Rugvica, Jesenice, Lukavec i Galdovo). Rezultati izračuna

toksičnosti na zebricama preko razvijenih QSAR modela pokazali su dobro slaganje s postojećim prioritizacijskim metodama, poput postupka temeljenog na toksičnosti (LC50) kemijskih spojeva na ribi *Pimephales promelas*. Također metoda računanja toksičnog potencijala pokazuje dobro slaganje s rezultatima provedenih testova embriotoksičnosti. Prednost ovog postupka prioritizacije toksičnog potencijala molekula u odnosu na postupke iz literature ogleda se u korištenju većeg broja toksičnih učinaka koji su mjereni pri značajno nižim koncentracijama spojeva, i na modelima razvijenim na većim skupovima molekula. Uz međusobno dobro slaganje, konačni modeli i prioritizacijski postupci pokrivaju višestruko veći kemijski prostor, što će omogućiti kvalitetnije procjene rizika za KORZ, kao i za nova, dosad nepoznata kemijska onečišćivala. Modeli razvijeni u ovom radu spremni su za uporabu i korištenje za potrebe predviđanje toksičnosti na novim skupovima kemijskih spojeva.

§ 7. POPIS OZNAKA, KRATICA I DODATAKA

BA	engl. *Balanced Accuracy*, uravnotežena točnost (Poglavlje 2.6)
CLP	Uredba o razvrstavanju, označivanju i pakiranju
DTXSID	Identifikator kemijskih spojeva u toksikološkoj bazi ToxCast
DS	Oznaka za skup fizikalno-kemijskih deskriptora računatih putem softvera RDKit (Poglavlje 4.4.3)
ER	engl. *effect ratio*, omjer učinka pri terapeutskim koncentracijama kod riba i čovjeka
FADM	Skupni naziv za farmaceutike, droge te metabolite farmaceutika i droga
Fin21	Konačni skup od 12 QSAR modela učenih na staničnim linijama iz skupa Tox21 (Poglavlje 5.6.1)
FinTC	Konačni skup od 7 QSAR modela učenih na skupu ToxCast. (Poglavlje 5.6.9)
Fathead Minnow	riblja vrsta, velikoglavi klen, . *Pimephales promelas*
FP	engl. *fingerprints*, molekulski otisci izračunati programom RDKit (Poglavlje 4.4.3)
FatMin_96h	Skup LC50 podataka na ribi vrste *Pimephales promelas*, proveden kroz 96 sati
generalizacija	mogućnost nekog modela da ispravno predvidi ciljnu varijablu
hpf	engl. *hours post fertilization*, sati nakon začetka (embrija)
HTKP	humana terapeutska koncentracija nekog kemijskog spoja u plazmi
HTS	engl. *high-throughput screening*, visokoprotočno testiranje
kemijski prostor	višedimenzionalni prostor varijabli koji opisuje skup kemijskih spojeva
KO	kemijska onečišćivala

KORZ	kemijska onečišćivala od rastućeg značaja za okoliš
logP	logaritam koeficijenta raspodjele, logK_{ow}
LogReg	logistička regresija (Poglavlje 2.5.1)
MCC	engl. *Matthews correlation coefficient*, korelacijski koeficijent po Matthewsu (Poglavlje 2.6)
MOA	*Mode of Action*, način djelovanja toksičnih spojeva
MRP	model riblje plazme
NN	engl. *neural networks*, neuronske mreže (Poglavlje 2.5)
ODV	Okvirna direktiva o vodama, ODV (Vode, 2001))
PBT	skupni pojam za procjenu perzistentnih, bioakumulativnih i toksičnih spojeva
prioritizacija	metodologija procjene rizika spojeva, prioritizacija opasnosti spojeva u nekom biotopu
PDS	prošireni skup deskriptora (Poglavlje 4.4.3)
Prediktorske varijable	bilo koji skup varijabli (X) koji se koristi kao ulaz u model za predviđanje nekog svojstva (y)
PVV	permutacijska važnost varijabli, *post-hoc* metoda za ispitivanje doprinosa prediktorskih varijabli u modelu (Poglavlje 2.8.1)
QSAR	engl. *Quantitative structure-activity relationships* (Poglavlje 2.3)
REACH	Europsko zakonodavstvo: registracija, procjena, autorizacija i ograničavanje kemikalija
RF	engl. *random forest*, algoritam nasumičnih šuma (Poglavlje 2.5.2)
SMILES	zapis molekulske strukture ASCII znakovima
TKO	tradicionalna kemijska onečišćivala
Tox21	baza podataka toksikoloških testova na staničnim linijama (Poglavlje 3.3.2)
ToxCast	baza podataka toksikoloških testova, u ovom radu se referira na skupove toksičnost na zebricama (*D. Rerio*) (Poglavlje 3.3.2)

Dodatak D1

2D deskriptori

Autocorr2D, BalabanJ, BertzCT, Chi0, Chi1, Chi0n - Chi4n, Chi0v - Chi4v, EState_VSA1 - EState_VSA11, ExactMolWt, FpDensityMorgan1, FpDensityMorgan2, FpDensityMorgan3, FractionCSP3, HallKierAlpha, HeavyAtomCount, HeavyAtomMolWt, Ipc, Kappa1 - Kappa3, LabuteASA, MaxAbsEStateIndex, MaxAbsPartialCharge, MaxEStateIndex, MaxPartialCharge, MinAbsEStateIndex, MinAbsPartialCharge, MinEStateIndex, MinPartialCharge, MolLogP, MolMR, MolWt, NHOHCount, NOCount, NumAliphaticCarbocycles, NumAliphaticHeterocycles, NumAliphaticRings, NumAromaticCarbocycles, NumAromaticHeterocycles, NumAromaticRings, NumHAcceptors, NumHDonors, NumHeteroatoms, NumRadicalElectrons, NumRotatableBonds, NumSaturatedCarbocycles, NumSaturatedHeterocycles, NumSaturatedRings, NumValenceElectrons, PEOE_VSA1 - PEOE_VSA14, RingCount, SMR_VSA1 - SMR_VSA10, SlogP_VSA1 - SlogP_VSA12, TPSA, VSA_EState1 - VSA_EState10, qed

Deskriptori bazirani na prebrojavanju fragmenata

---

fr_Al_COO, fr_Al_OH, fr_Al_OH_noTert, fr_ArN, fr_Ar_COO, fr_Ar_N, fr_Ar_NH, fr_Ar_OH, fr_COO, fr_COO2, fr_C_O, fr_C_O_noCOO, fr_C_S, fr_HOCCN, fr_Imine, fr_NH0, fr_NH1, fr_NH2, fr_N_O, fr_Ndealkylation1, fr_Ndealkylation2, fr_Nhpyrrole, fr_SH, fr_aldehyde, fr_alkyl_carbamate, fr_alkyl_halide, fr_allylic_oxid, fr_amide, fr_amidine, fr_aniline, fr_aryl_methyl, fr_azide, fr_azo, fr_barbitur, fr_benzene, fr_benzodiazepine, fr_bicyclic, fr_diazo, fr_dihydropyridine, fr_epoxide, fr_ester, fr_ether, fr_furan, fr_guanido, fr_halogen, fr_hdrzine, fr_hdrzone, fr_imidazole, fr_imide, fr_isocyan, fr_isothiocyan, fr_ketone, fr_ketone_Topliss, fr_lactam, fr_lactone, fr_methoxy, fr_morpholine, fr_nitrile, fr_nitro, fr_nitro_arom, fr_nitro_arom_nonortho, fr_nitroso, fr_oxazole, fr_oxime, fr_para_hydroxylation, fr_phenol, fr_phenol_noOrthoHbond, fr_phos_acid, fr_phos_ester, fr_piperdine, fr_piperzine, fr_priamide, fr_prisulfonamd, fr_pyridine, fr_quatN, fr_sulfide, fr_sulfonamd, fr_sulfone, fr_term_acetylene, fr_tetrazole, fr_thiazole, fr_thiocyan, fr_thiophene, fr_unbrch_alkane, fr_urea

3D deskriptori

Asphericity, Autocorr3D, Eccentricity, GETAWAY, InertialShapeFactor, MORSE, NPR1, NPR2, PMI1, PMI2, PMI3, RDF, RadiusOfGyration, SpherocityIndex, WHIM

Popis elektroničkih dodataka

Elektronički dodaci ovog rada mogu se preuzeti iz javnog repozitorija [227] na poveznici https://doi.org/10.5281/zenodo.4678187 i na CD-u priloženom ovom radu.

Oznaka elektroničkog dodatka	Opis
E1	Pročišćeni podaci skupa ToxCast za modeliranje (toksični učinci)
E2	Tablica s kvantificiranim kemijskih spojeva iz rijeke Save (plazma, sediment, voda)
E3	Skup SMILES zapisa iz skupa ToxCast (.smiles)
E4	Datoteka za standardizaciju struktura (.xml)
E5	Deskriptori i molekulski otisci za skup ToxCast
E6	Izračunate TU za skup Sava (sediment i voda)
E7	Rezultati svih ToxCast i Tox21 modela
E8	Tablica premutacijskih važnosti za FinTC modele
E9	Predviđanja Fin21 i FinTC modela na skupu Sava kao binarne vrijednosti i vjerojatnosti

§ 8. LITERATURNI IZVORI

1. R. Loos, B. M. Gawlik, G. Locoro, E. Rimaviciute, S. Contini, i G. Bidoglio, *Environ. Pollut.*, **2009**, *157*, 561.
2. E. Commission, REACH - Chemicals - Environment - European Commission, https://ec.europa.eu/environment/chemicals/reach/reach_en.htm. , preuzeto 29. travnja 2021. god.
3. E. Commission, Understanding CLP - ECHA, https://echa.europa.eu/hr/regulations/clp/understanding-clp. , preuzeto 29. travnja 2021. god.
4. A. Lombardo, A. Roncaglioni, E. Benfenati, M. Nendza, H. Segner, S. Jeram, E. Pauné, i G. Schüürmann, *Environ. Res.*, **2014**, *135*, 156.
5. OECD, Types of Chemicals Assessed in the OECD Cooperative Chemicals Assessment Programme - OECD, https://www.oecd.org/chemicalsafety/risk-assessment/typesofchemicalsassessedintheoecdcooperativechemicalsassessmentprogramme.htm. , preuzeto 29. travnja 2021. god.
6. E. Commission, OECD Test Guidelines for the Chemicals - OECD, https://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm. , preuzeto 29. travnja 2021. god.
7. EAA, Chemicals in the European Environment: Low Doses, High Stakes? — European Environment Agency, https://www.eea.europa.eu/publications/NYM2. , preuzeto 29. travnja 2021. god.
8. E. Malaj, P. C. Von Der Ohe, M. Grote, R. Kühne, C. P. Mondy, P. Usseglio-Polatera, W. Brack, i R. B. Schäfer, *Proc. Natl. Acad. Sci. U. S. A.*, **2014**, *111*, 9549.
9. E. Commission, DIREKTIVA 2000/60/EC EUROPSKOG PARLAMENTA I VIJEĆA KOJOM SE USPOSTAVLJA OKVIR ZA DJELOVANJE ZAJEDNICE NA PODRUČJU POLITIKE VODA, Zagreb, Hrvatska, **2001**.
10. R. Altenburger, S. Ait-Aissa, P. Antczak, T. Backhaus, D. Barceló, T. B. Seiler, F. Brion, W. Busch, K. Chipman, M. L. de Alda, G. de Aragão Umbuzeiro, B. I. Escher, F. Falciani, M. Faust, A. Focks, K. Hilscherova, J. Hollender, H. Hollert, F. Jäger, A. Jahnke, A. Kortenkamp, M. Krauss, G. F. Lemkine, J. Munthe, S. Neumann, E. L. Schymanski, M. Scrimshaw, H. Segner, J. Slobodnik, F. Smedes, S. Kughathas, I.

Teodorovic, A. J. Tindall, K. E. Tollefsen, K. H. Walz, T. D. Williams, P. J. Van den Brink, J. van Gils, B. Vrana, X. Zhang, i W. Brack, *Sci. Total Environ.*, **2015**, *512–513*, 540.

11. V. A. Buonsante, H. Muilerman, T. Santos, C. Robinson, i A. C. Tweedale, *Environ. Res.*, **2014**, *135*, 139.
12. Q. Li, L. Chen, L. Liu, i L. Wu, *Environ. Sci. Pollut. Res.*, **2016**, *23*, 4908.
13. S. D. Richardson i T. A. Ternes, *Anal. Chem.*, **2011**, *83*, 4616.
14. E. Commission, EUR-Lex - 32013L0039 - EN - EUR-Lex, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32013L0039. , preuzeto 29. travnja 2021. god.
15. V. Christen, S. Hickmann, B. Rechenberg, and K. Fent, Aquat. Toxicol., 2010, 96, 167.
16. W. Brack, V. Dulio, M. Ågerstrand, I. Allan, R. Altenburger, M. Brinkmann, D. Bunke, R. M. Burgess, I. Cousins, B. I. Escher, F. J. Hernández, L. M. Hewitt, K. Hilscherová, J. Hollender, H. Hollert, R. Kase, B. Klauer, C. Lindim, D. L. Herráez, C. Miège, J. Munthe, S. O'Toole, L. Posthuma, H. Rüdel, R. B. Schäfer, M. Sengl, F. Smedes, D. van de Meent, P. J. van den Brink, J. van Gils, A. P. van Wezel, A. D. Vethaak, E. Vermeirssen, P. C. von der Ohe, i B. Vrana, *Sci. Total Environ.*, **2017**, *576*, 720.
17. R. Truhaut, *Ecotoxicol. Environ. Saf.*, **1977**, *1*, 151.
18. R. Relyea i J. Hoverman, *Ecol. Lett.*, **2006**, *9*, 1157.
19. A. B. Raies i V. B. Bajic, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **2016**, *6*, 147.
20. A. M. Richard, R. Huang, S. Waidyanatha, P. Shinn, B. J. Collins, I. Thillainadarajah, C. M. Grulke, A. J. Williams, R. R. Lougee, R. S. Judson, K. A. Houck, M. Shobair, C. Yang, J. F. Rathman, A. Yasgar, S. C. Fitzpatrick, A. Simeonov, R. S. Thomas, K. M. Crofton, R. S. Paules, J. R. Bucher, C. P. Austin, R. J. Kavlock, i R. R. Tice, *Chem. Res. Toxicol.*, **2020**.
21. P. H. Rowe, "*In Silico Toxicology: Principles and Applications*", ur. Mark Cronin i Judith Madden, **2010**, 252.
22. L. Truong, D. M. Reif, L. S. Mary, M. C. Geier, H. D. Truong, i R. L. Tanguay, *Toxicol. Sci.*, **2014**, *137*, 212.
23. P. H. Rowe, *Issues Toxicol.*, **2010**, *1,* 1.
24. H. van de Waterbeemd i E. Gifford, *Nat. Rev. Drug Discov.*, **2003**, *2*, 192.
25. J. Bailey, M. Thew, i M. Balls, *ATLA Altern. to Lab. Anim.*, **2014**, *42*, 181.
26. R. S. Settivari, N. Ball, L. Murphy, R. Rasoulpour, D. R. Boverhof, i E. W. Carney, *J.*

Am. Assoc. Lab. Anim. Sci., **2015**, *54*, 214.

27. T. Hartung i S. Hoffmann, *ALTEX*, **2009**, *26*, 155.
28. N. Klüver, C. Vogs, R. Altenburger, B. I. Escher, i S. Scholz, *Chemosphere*, **2016**, *164*, 164.
29. M. Krauss, H. Singer, and J. Hollender, *Anal. Bioanal. Chem.*, **2010**, *397*, 943.
30. S. Comtois-Marotte, T. Chappuis, S. Vo Duy, N. Gilbert, A. Lajeunesse, S. Taktek, M. Desrosiers, É. Veilleux, i S. Sauvé, *Chemosphere*, **2017**, *166*, 400.
31. A. J. Ebele, M. Abou-Elwafa Abdallah, i S. Harrad, *Emerg. Contam.*, **2017**, *3*, 1.
32. B. F. da Silva, A. Jelic, R. López-Serna, A. A. Mozeto, M. Petrovic, i D. Barceló, *Chemosphere*, **2011**, *85*, 1331.
33. S. Babić, J. Barišić, D. Stipaničev, S. Repec, M. Lovrić, O. Malev, D. Martinović-Weigelt, R. Čož-Rakovac, i G. Klobučar, *Sci. Total Environ.*, **2018**, *643*, 435.
34. M. Šrut, L. Traven, A. Štambuk, S. Kralj, R. Žaja, V. Mićović, i G. I. V. V Klobučar, *Toxicol. Vitr.*, **2011**, *25*, 308.
35. M. Sigurnjak Bureš, M. Cvetnić, M. Miloloža, D. Kučić Grgić, M. Markić, H. Kušić, T. Bolanča, M. Rogošić, i Š. Ukić, *Environ. Chem. Lett.*, **2021**.
36. W. et al. Brack, R. Altenburger, G. Schüürmann, M. Krauss, D. López Herráez, J. van Gils, J. Slobodnik, J. Munthe, B. M. Gawlik, A. van Wezel, M. Schriks, J. Hollender, K. E. Tollefsen, O. Mekenyan, S. Dimitrov, D. Bunke, I. Cousins, L. Posthuma, P. J. van den Brink, M. López de Alda, D. Barceló, M. Faust, A. Kortenkamp, M. Scrimshaw, S. Ignatova, G. Engelen, G. Massmann, G. Lemkine, I. Teodorovic, K. H. Walz, V. Dulio, M. T. O. Jonker, F. Jäger, K. Chipman, F. Falciani, I. Liska, D. Rooke, X. Zhang, H. Hollert, B. Vrana, K. Hilscherova, K. Kramer, S. Neumann, R. Hammerbacher, T. Backhaus, J. Mack, H. Segner, B. Escher, i G. de Aragão Umbuzeiro, *Sci. Total Environ.*, **2015**, *503–504*, 22.
37. H. Hollert, S. Keiter, N. König, M. Rudolf, M. Ulrich, T. Braunbeck, N. König, M. Rudolf, M. Ulrich, i T. Braunbeck, *J. Soils Sediments*, **2004**, *4*, 94.
38. S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, i S. H. Bryant, *Nucleic Acids Res.*, **2016**, *44*, D1202.
39. K. J. Groh, R. N. Carvalho, J. K. Chipman, N. D. Denslow, M. Halder, C. A. Murphy, D. Roelofs, A. Rolaki, K. Schirmer, i K. H. Watanabe, *Chemosphere*, **2015**, *120*, 764.
40. C. Prasse, D. Stalter, U. Schulte-Oehlmann, J. Oehlmann, i T. A. Ternes, *Water Res.*, **2015**, *87*, 237.

41. "*The Sava River*", **2015**, prvo izdanje , ur. Radmila Milačič, Janez Scancar, i Momir Paunović, Springer-Verlag Berlin Heidelberg, Berlin, Germany.
42. I. Senta, I. Krizman, M. Ahel, i S. Terzic, *J. Chromatogr. A*, **2015**, *1425*, 204.
43. A. Bielen, A. Šimatović, J. Kosić-Vukšić, I. Senta, M. Ahel, S. Babić, T. Jurina, J. J. González Plaza, M. Milaković, i N. Udiković-Kolić, *Water Res.*, **2017**, *126*, 79.
44. S. Terzić i M. Ahel, *Arh. Hig. Rada Toksikol.*, **2006**, *57*, 297.
45. S. Terzić, I. Senta, M. Ahel, M. Gros, M. Petrović, D. Barcelo, J. Müller, T. Knepper, I. Martí, F. Ventura, P. Jovančić, i D. Jabučar, *Sci. Total Environ.*, **2008**, *399*, 66.
46. R. Ćuk, D. Tomas, i I. Vučković, Kakvoća rijeke Save u 2012. godini, **2014**.
47. G. I. V. Klobučar, A. Štambuk, M. Pavlica, M. Sertić Perić, B. Kutuzović Hackenberger, i K. Hylland, *Ecotoxicology*, **2010**, *19*, 77.
48. O. Malev, R. S. Klobučar, E. Fabbretti, i P. Trebše, *Pestic. Biochem. Physiol.*, **2012**, *104*, 178.
49. T. Smital i M. Ahel, "*Handbook of Environmental Chemistry*", ur. Andrey G. Barceló, Damià, Kostianoy, **2015**, Vol. 31, Springer-Verlag Berlin Heidelberg, Berlin, Germany, 177.
50. T. Smital, S. Terzic, R. Zaja, I. Senta, B. Pivcevic, M. Popovic, I. Mikac, K. E. Tollefsen, K. V. Thomas, i M. Ahel, *Ecotoxicol. Environ. Saf.*, **2011**, *74*, 844.
51. J. Vidmar, T. Zuliani, P. Novak, A. Drinčić, J. Ščančar, i R. Milačič, *J. Soils Sediments*, **2017**, *17*, 1917.
52. E. Heath, J. Ščančar, T. Zuliani, i R. Milačič, *Environ. Monit. Assess.*, **2010**, *163*, 277.
53. T. Smital, S. Terzić, J. Lončar, I. Senta, R. Žaja, M. Popović, I. Mikac, K. E. Tollefsen, K. V. Thomas, i M. Ahel, *Environ. Sci. Pollut. Res.*, **2013**, *20*, 1384.
54. T. Källqvist, R. Milačič, T. Smital, K. V. Thomas, S. Vranes, i K. E. Tollefsen, *Water Res.*, **2008**, *42*, 2146.
55. A. Marinović Ruždjak i D. Ruždjak, *Environ. Monit. Assess.*, **2015**, *187*, 1.
56. S. Li, D. L. Villeneuve, J. P. Berninger, B. R. Blackwell, J. E. Cavallin, M. N. Hughes, K. M. Jensen, M. D. Kahl, K. E. Stevens, L. M. Thomas, M. A. Weberg, G. T. Ankley, S. Li, J. P. Berninger, Z. Jorgenson, i A. L. Schroeder, *Sci. Total Environ.*, **2017**, *579*, 825.
57. P. Kay, S. R. Hughes, J. R. Ault, A. E. Ashcroft, i L. E. Brown, *Environ. Pollut.*, **2017**, *220*, 1447.
58. J. Fick, R. H. Lindberg, M. Tysklind, i D. G. J. Larsson, *Regul. Toxicol. Pharmacol.*, **2010**, *58*, 516.

59. I. Senta, I. Krizman-Matasic, S. Terzic, i M. Ahel, *J. Chromatogr. A*, **2017**, *1509*, 60.
60. I. Krizman-Matasic, P. Kostanjevecki, M. Ahel, i S. Terzic, *J. Chromatogr. A*, **2018**, *1533*, 102.
61. R. Pal, M. Megharaj, K. P. Kirkbride, i R. Naidu, *Sci. Total Environ.*, **2013**, *463–464*, 1079.
62. M. K. Yadav, M. D. Short, R. Aryal, C. Gerber, B. van den Akker, i C. P. Saint, *Water Res.*, **2017**, *124*, 713.
63. S. Terzic, I. Senta, i M. Ahel, *Environ. Pollut.*, **2010**, *158*, 2686.
64. T. H. Hutchinson, J. C. Madden, V. Naidoo, i C. H. Walker, *Philos. Trans. R. Soc. B Biol. Sci.*, **2014**, *369*.
65. D. B. Huggett, J. C. Cook, J. F. Ericson, i R. T. Williams, *Hum. Ecol. Risk Assess.*, **2003**, *9*, 1789.
66. O. Malev, M. Lovrić, D. Stipaničev, S. Repec, D. Martinović-Weigelt, D. Zanella, T. Ivanković, V. S. Đuretec, J. Barišić, M. Li, i G. Klobučar, *Environ. Pollut.*, **2020**, 115162.
67. G. Patlewicz, N. Ball, E. D. Booth, E. Hulzebos, E. Zvinavashe, i C. Hennes, *Regul. Toxicol. Pharmacol.*, **2013**, *67*, 1.
68. M. Rand-Weaver, L. Margiotta-Casaluci, A. Patel, G. H. Panter, S. F. Owen, i J. P. Sumpter, *Environ. Sci. Technol.*, **2013**, *47*, 11384.
69. A. David, A. Lange, C. R. Tyler, i E. M. Hill, *Sci. Total Environ.*, **2018**, *621*, 782.
70. T. G. Bean, B. A. Rattner, R. S. Lazarus, D. D. Day, S. R. Burket, B. W. Brooks, S. P. Haddad, i W. W. Bowerman, *Environ. Pollut.*, **2018**, *232*, 533.
71. D. Muir, D. Simmons, X. Wang, T. Peart, M. Villella, J. Miller, i J. Sherry, *Sci. Rep.*, **2017**, *7*, 1.
72. S. Poirier Larabie, M. Houde, i C. Gagnon, *J. Chromatogr. A*, **2017**, *1522*, 48.
73. R. Tanoue, K. Nomiyama, H. Nakamura, J. W. Kim, T. Isobe, R. Shinohara, T. Kunisue, i S. Tanabe, *Environ. Sci. Technol.*, **2015**, *49*, 11649.
74. C. Hansch, P. P. Maloney, T. Fujita, i R. M. Muir, *Nature*, **1962**, *194*, 178.
75. C. Hansch i T. Fujita, *J. Am. Chem. Soc.*, **1964**, *86*, 1616.
76. K. Tanabe, B. Lučić, D. Amić, T. Kurita, M. Kaihara, N. Onodera, i T. Suzuki, *Mol. Divers.*, **2010**, *14*, 789.
77. M. Lovrić, *Kem. Ind.*, **2018**, *67*, 409.
78. C. Hansch i F. Helmer, *J. Polym. Sci. Part A-1 Polym. Chem.*, **1968**, *6*, 3295.
79. L. David, A. Thakkar, R. Mercado, i O. Engkvist, *J. Cheminform.*, **2020**, *12*, 1.

80. OECD, OECD GUIDELINE FOR THE TESTING OF CHEMICALS, https://www.oecd-ilibrary.org/environment/test-no-227-terrestrial-plant-test-vegetative-vigour-test_9789264067295-en. , preuzeto 29. travnja 2021. god.
81. D. Weininger, *J. Chem. Inf. Comput. Sci.*, **1988**, *28*, 31.
82. R. Todeschini i V. Consonni, "*Handbook of molecular descriptors. WileyVCH, Weinheim*", **2000**, ur. R. Mannhold, H. Kubinyi, i H. Timmerman, Wiley, New York.
83. D. Rogers i M. Hahn, *J. Chem. Inf. Model.*, **2010**, *50*, 742.
84. A. Mayr, G. Klambauer, T. Unterthiner, i S. Hochreiter, *Front. Environ. Sci.*, **2016**, *3*.
85. T. Hastie, R. Tibshirani, i J. Friedman, "*The Elements of Statistical Learning*", **2009**, Springer New York, New York, NY.
86. R. L. M. Robinson, A. Palczewska, J. Palczewski, i N. Kidley, *J. Chem. Inf. Model*, **2017**, *57*, 1773.
87. F. Murtagh, *Neurocomputing*, **1991**, *2*, 183.
88. L. Breiman, *Mach. Learn.*, **2001**, *45*, 5.
89. C. Developers, Random Forest Classifier, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. , preuzeto 29. travnja 2021. god.
90. M. Lovrić, K. Pavlović, P. Žuvela, A. Spataru, B. Lučić, R. Kern, i M. W. Wong, *J. Chemom.*, **2021**, e3349.
91. M. Lovrić, R. Meister, T. Steck, L. Fadljević, J. Gerdenitsch, S. Schuster, L. Schiefermüller, S. Lindstaedt, i R. Kern, *Adv. Model. Simul. Eng. Sci.*, **2020**, *7*, 46
92. T. Bolanča, Š. Cerjan-Stefanović, M. Luša, Š. Ukić, i M. Rogošić, *Sep. Sci. Technol.*, **2010**, *45*, 236.
93. M. Fernandez, F. Ban, G. Woo, M. Hsing, T. Yamazaki, E. Leblanc, P. S. Rennie, W. J. Welch, i A. Cherkasov, *J. Chem. Inf. Model.*, **2018**, *58*, 1533.
94. P. Probst, B. Bischl, i A.-L. Boulesteix, *J. Mach. Learn. Res.*, **2019**, 1.
95. J. Snoek, H. Larochelle, i R. P. Adams, Practical Bayesian Optimization of Machine Learning Algorithms, Zbornik radova Neural Information Processing Systems (NIPS) 2012, Lake Tahoe, Nevada, USA, **2012**, 2951–2959, ur. F. Pereira, C.J.C. Burges, L. Bottou i K.Q. Weinberger
96. J. Bergstra, D. Yamins, i D. D. Cox, Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures, PMLR, **2013**.
97. P. M. Lerman, *Appl. Stat.*, **1980**, *29*, 77.

98. F. Nogueira, Bayesian Optimization: Open source constrained global optimization tool for Python, https://github.com/fmfn/BayesianOptimization. , preuzeto 01. rujna 2020. god.
99. J. Zhang, D. Mucs, U. Norinder, i F. Svensson, *J. Chem. Inf. Model.*, **2019**, *59*, 4150.
100. A. Tharwat, *Appl. Comput. Informatics*, **2018**, *17*, 168.
101. A. Tharwat, Y. S. Moemen, i A. E. Hassanien, *J. Biomed. Inform.*, **2017**, *68*, 132.
102. H. He i E. A. Garcia, *IEEE Trans. Knowl. Data Eng.*, **2009**, *21*, 1263.
103. D. Chicco i C. Rovelli, *PLoS One*, **2019**, *14*.
104. S. Boughorbel, F. Jarray, i M. El-Anbari, *PLoS One*, **2017**, *12*.
105. G. Weiss, *ACM SIGKDD Explor. Newsl.*, **2004**, *6*, 7.
106. P. Czodrowski, *J. Comput. Aided. Mol. Des.*, **2014**, *28*, 1049.
107. K. Mansouri, N. Kleinstreuer, A. M. Abdelaziz, D. Alberga, V. M. Alves, P. L. Andersson, C. H. Andrade, F. Bai, I. Balabin, D. Ballabio, E. Benfenati, B. Bhhatarai, S. Boyer, J. Chen, V. Consonni, S. Farag, D. Fourches, A. T. García-Sosa, P. Gramatica, F. Grisoni, C. M. Grulke, H. Hong, D. Horvath, X. Hu, R. Huang, N. Jeliazkova, J. Li, X. Li, H. Liu, S. Manganelli, G. F. Mangiatordi, U. Maran, G. Marcou, T. Martin, E. Muratov, D.-T. Nguyen, O. Nicolotti, N. G. Nikolov, U. Norinder, E. Papa, M. Petitjean, G. Piir, P. Pogodin, V. Poroikov, X. Qiao, A. M. Richard, A. Roncaglioni, P. Ruiz, C. Rupakheti, S. Sakkiah, A. Sangion, K.-W. Schramm, C. Selvaraj, I. Shah, S. Sild, L. Sun, O. Taboureau, Y. Tang, I. V. Tetko, R. Todeschini, W. Tong, D. Trisciuzzi, A. Tropsha, G. Van Den Driessche, A. Varnek, Z. Wang, E. B. Wedebye, A. J. Williams, H. Xie, A. V. Zakharov, Z. Zheng, i R. S. Judson, *Environ. Health Perspect.*, **2020**, *128*, 027002.
108. K. Mansouri, A. M. Abdelaziz, A. Rybacka, A. Roncaglioni, A. Tropsha, A. Varnek, A. Zakharov, A. Worth, A. M. Richard, C. M. Grulke, D. Trisciuzzi, D. Fourches, D. Horvath, E. Benfenati, E. Muratov, E. B. Wedebye, F. Grisoni, G. F. Mangiatordi, G. M. Incisivo, H. Hong, H. W. Ng, I. V. Tetko, I. Balabin, J. Kancherla, J. Shen, J. Burton, M. Nicklaus, M. Cassotti, N. G. Nikolov, O. Nicolotti, P. L. Andersson, Q. Zang, R. Politi, R. D. Beger, R. Todeschini, R. Huang, S. Farag, S. A. Rosenberg, S. Slavov, X. Hu, i R. S. Judson, *Environ. Health Perspect.*, **2016**, *124*, 1023.
109. R. J. Urbanowicz i J. H. Moore, *Evol. Intell.*, **2015**, *8*, 89.
110. B. Krawczyk, M. Woźniak, i G. Schaefer, *Appl. Soft Comput. J.*, **2014**, *14*, 554.
111. N. V. Chawla, N. Japkowicz, i A. Kotcz, *ACM SIGKDD Explor. Newsl.*, **2004**, *6*, 1.
112. E. N. Muratov, A. G. Artemenko, E. V. Varlamova, P. G. Polischuk, V. P. Lozitsky, A.

S. Fedchuk, R. L. Lozitska, T. L. Gridina, L. S. Koroleva, V. N. Silnikov, A. S. Galabov, V. A. Makarov, O. B. Riabova, P. Wutzler, M. Schmidtke, i V. E. Kuzmin, *Future Med. Chem.*, **2010**, *2*, 1205.

113. V. E. Kuz'min, A. G. Artemenko, E. N. Muratov, I. L. Volineckaya, V. A. Makarov, O. B. Riabova, P. Wutzler, i M. Schmidtke, *J. Med. Chem.*, **2007**, *50*, 4205.
114. OECD, OECD PRINCIPLES FOR THE VALIDATION, FOR REGULATORY PURPOSES, OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP MODELS, https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf. , preuzeto 29. travnja 2021. god.
115. P. Polishchuk, *J. Chem. Inf. Model.*, **2017**, *57*, 2618.
116. I. Sushko, S. Novotarskyi, R. Körner, A. K. Pandey, A. Cherkasov, J. Li, P. Gramatica, K. Hansen, T. Schroeter, K. R. Müller, L. Xi, H. Liu, X. Yao, T. Öberg, F. Hormozdiari, P. Dao, C. Sahinalp, R. Todeschini, P. Polishchuk, A. Artemenko, V. Kuz'Min, T. M. Martin, D. M. Young, D. Fourches, E. Muratov, A. Tropsha, I. Baskin, D. Horvath, G. Marcou, C. Muller, A. Varnek, V. V. Prokopenko, i I. V. Tetko, *J. Chem. Inf. Model.*, **2010**, *50*, 2094.
117. V. E. Kuz'min, P. G. Polishchuk, A. G. Artemenko, i S. A. Andronati, *Mol. Inform.*, **2011**, *30*, 593.
118. A. Tropsha, P. Gramatica, i V. K. Gombar, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application i Interpretation of QSPR Models, Wiley-VCH Verlag, **2003**.
119. T. Fujita i D. A. Winkler, *J. Chem. Inf. Model.*, **2016**, *56*, 269.
120. U. Johansson, C. Sönströd, U. Norinder, i H. Boström, *Future Med. Chem.*, **2011**, *3*, 647.
121. R. Guha, *J. Comput. Aided. Mol. Des.*, **2008**, *22*, 857.
122. A. Altmann, L. Toloşi, O. Sander, i T. Lengauer, *Bioinformatics*, **2010**, *26*, 1340.
123. S. Khedkar, V. Subramanian, G. Shinde, i P. Gandhi, *SSRN Electron. J.*, **2019**.
124. M. Du, N. Liu, i X. Hu, *Commun. ACM*, **2020**, *63*, 68.
125. K. K. Nicodemus, J. D. Malley, C. Strobl, i A. Ziegler, *BMC Bioinformatics*, **2010**, *11*.
126. C. Strobl, A. L. Boulesteix, T. Kneib, T. Augustin, i A. Zeileis, *BMC Bioinformatics*, **2008**, *9*, 1.
127. I. Šimić, M. Lovrić, R. Godec, M. Kröll, i I. Bešlić, *Environ. Pollut.*, **2020**, *263*, 114587.
128. L. McInnes, J. Healy, i J. Melville, *J. Open Source Softw.*, **2018**, *3*, 861.

129. D. Probst i J. L. Reymond, *J. Cheminform.*, **2020**, *12*, 12.
130. J. M. Ali, D. L. D'Souza, K. Schwarz, L. G. Allmon, R. P. Singh, D. D. Snow, S. L. Bartelt-Hunt, i A. S. Kolok, *Sci. Total Environ.*, **2018**, *618*, 1371.
131. J. L. Tadeo, C. Sánchez-Brunete, B. Albero, A. I. García-Valcárcel, i R. A. Pérez, *Cent. Eur. J. Chem.*, **2012**, *10*, 480.
132. D. Stipaničev, Z. Dragun, S. Repec, K. Rebok, i M. Jordanova, *Ecotoxicol. Environ. Saf.*, **2017**, *135*, 48.
133. S. Dekić, G. Klobučar, T. Ivanković, D. Zanella, M. Vucić, J. P. Bourdineaud, i J. Hrenović, *Int. J. Environ. Health Res.*, **2018**, *28*, 315.
134. Python, Python, www.python.org. , preuzeto 01. travnja 2020. god.
135. S. Seabold i J. Perktold, Data Structures for Statistical Computing in Python, Zbornik radova 9th Python in Science Conference, ur. Stéfan van der Walt i Jarrod Millman, **2010**, 92
136. F. Pedregosa, V. Michel, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, G. Varoquaux, A. Gramfort, B. Thirion, V. Dubourg, A. Passos, M. Brucher, É. Duchesnay, B. Thirion, i P. Prettenhofer, *J. Mach. Learn. Res.*, **2011**, *12*, 2825.
137. S. Van Der Walt, S. C. Colbert, i G. Varoquaux, *Comput. Sci. Eng.*, **2011**, *13*, 22.
138. W. Mckinney, Statsmodels: Econometric and Statistical Modeling with PythonZbornik radova 9th Python in Science Conference, ur. Stéfan van der Walt i Jarrod Millman, **2010**, 51–56.
139. M. Waskom, O. Botvinnik, D. O'Kane, P. Hobson, S. Lukauskas, D. C. Gemperline, T. Augspurger, Y. Halchenko, J. B. Cole, J. Warmenhoven, J. de Ruiter, C. Pye, S. Hoyer, J. Vanderplas, S. Villalba, G. Kunter, E. Quintero, P. Bachant, M. Martin, K. Meyer, A. Miles, Y. Ram, T. Yarkoni, M. L. Williams, C. Evans, C. Fitzgerald, Brian, C. Fonnesbeck, A. Lee, i A. Qalieh, Mwaskom/Seaborn: V0.8.1 (September 2017).
140. J. D. Hunter, *Comput. Sci. Eng.*, **2007**, *9*, 99.
141. EPA, ECOTOX | Home, https://cfpub.epa.gov/ecotox/. , preuzeto 01. travnja 2020. god.
142. PAN Pesticide Database, http://www.pesticideinfo.org/ , preuzeto 01. travnja 2020. god.
143. R. Thomas, *ALTEX*, **2018**, *35*, 163.
144. A. M. Richard, R. S. Judson, K. A. Houck, C. M. Grulke, P. Volarath, I. Thillainadarajah, C. Yang, J. Rathman, M. T. Martin, J. F. Wambaugh, T. B. Knudsen,

J. Kancherla, K. Mansouri, G. Patlewicz, A. J. Williams, S. B. Little, K. M. Crofton, i R. S. Thomas, *Chem. Res. Toxicol.*, **2016**, *29*, 1225.

145. A. Morger, M. Mathea, J. H. Achenbach, A. Wolf, R. Buesen, K. J. Schleifer, R. Landsiedel, i A. Volkamer, *J. Cheminform.*, **2020**, *12*, 1.
146. D. M. Reif, L. Truong, D. Mandrell, S. Marvel, G. Zhang, i R. L. Tanguay, *Arch. Toxicol.*, **2016**, *90*, 1459.
147. R. Huang, *Challenges Adv. Comput. Chem. Phys.*, **2019**, *30*, 279.
148. E. Benfenati, A. Manganaro, i G. Gini, *CEUR Workshop Proc.*, **2013**, *1107*, 21.
149. E. Benfenati, C. Ileana Cappelli, M. Ifigenia Petoumenou, F. Pizzo, A. Lomardo, F. Albanese, A. Roncaglioni, A. Manganaro, i F. Lemke, PROMETHEUS – Prioritization Of chemicals: a METHodology Embracing PBT parameters into a Unified Strategy, **2016**., https://www.researchgate.net/publication/321484822_PROMETHEUS_-_PRioritization_Of_chemicals_a_METHodology_Embracing_PBT_parameters_into_a_Unified_Strategy, preuzeto 01. travnja 2020. god.
150. F. Grisoni, V. Consonni, S. Villa, M. Vighi, i R. Todeschini, *Chemosphere*, **2015**, *127*, 171.
151. A. Cassano, A. Manganaro, T. Martin, D. Young, N. Piclin, M. Pintore, D. Bigoni, i E. Benfenati, *Chem. Cent. J.*, **2010**, *4*, S4.
152. N. Amaury, E. Benfenati, E. Boriani, J. R. Chrétien, J. Cotterill, F. Lemke, N. Piclin, M. Pintore, C. Porcelli, N. Price, A. Roncaglioni, A. Toropov, M. Casalegno, A. Chana, and Q. Chaudhry, "Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes", ur. Emilio Benfenati, prvo izdanje, Elsevier, 2007, New York
153. M. Cassotti, D. Ballabio, R. Todeschini, i V. Consonni, *SAR QSAR Environ. Res.*, **2015**, *26*, 217.
154. J. Perktold, S. Seabold, i J. Perktold, Zbornik radova 9th Python in Science Conf, ur. Stéfan van der Walt i Jarrod Millman, **2010**, 92–96.
155. G. Lemaitre, F. Nogueira, D. Oliveira, i C. Aridas, imbalanced-learn API — imbalanced-learn 0.5.0 documentation, https://imbalanced-learn.readthedocs.io/en/stable/api.html. , preuzeto 01. travnja 2020. god.
156. eli5 · PyPI, https://pypi.org/project/eli5/. , preuzeto 01. travnja 2020. god.
157. ChemAxon - Software Solutions i Services for Chemistry & Biology, https://chemaxon.com/products/jchem-for-office. , preuzeto 01. travnja 2020. god.
158. M. Lovrić, J. M. Molero, i R. Kern, *Mol. Inform.*, **2019**, *38*.

159. OECD, Test No. 236: Fish Embryo Acute Toxicity (FET) Test, OECD, **2013**.

160. S. Babić, J. Barišić, H. Višić, R. Sauerborn Klobučar, N. Topić Popović, I. Strunjak-Perović, R. Čož-Rakovac, i G. Klobučar, *Water Res.*, **2017**, *115*, 9.

161. N. de Castro-Català, M. Kuzmanovic, N. Roig, J. Sierra, A. Ginebreda, D. Barceló, S. Pérez, M. Petrovic, Y. Picó, M. Schuhmacher, i I. Muñoz, *Sci. Total Environ.*, **2016**, *540*, 297.

162. "*Emerging Contaminants in River Ecosystems*", **2016**, prvo izdanje , ur. Mira Petrovic, Sergi Sabater, Arturo Elosegi, i Damià Barceló, Springer International Publishing.

163. A. Ginebreda, M. Kuzmanovic, H. Guasch, M. L. de Alda, J. C. López-Doval, I. Muñoz, M. Ricart, A. M. Romaní, S. Sabater, i D. Barceló, *Sci. Total Environ.*, **2014**, *468–469*, 715.

164. T. Backhaus i M. Faust, *Environ. Sci. Technol.*, **2012**, *46*, 2564.

165. D. M. Di Toro, C. S. Zarba, D. J. Hansen, W. J. Berry, R. C. Swartz, C. E. Cowan, S. P. Pavlou, H. E. Allen, N. A. Thomas, i P. R. Paquin, *Environ. Toxicol. Chem.*, **1991**, *10*, 1541.

166. R. Seth, D. Mackay, i J. Muncke, *Environ. Sci. Technol.*, **1999**, *33*, 2390.

167. J. P. Meador, A. Yeh, i E. P. Gallagher, *Environ. Pollut.*, **2017**, *230*, 1018.

168. M. Schulz, S. Iwersen-Bergmann, H. Andresen, i A. Schmoldt, *Crit. Care*, **2012**, *16*, R136.

169. M. Schulz i A. Schmoldt, Therapeutic and toxic blood concentrations of more than 800 drugs and other xenobiotics, **2003**.

170. D. Fourches, E. Muratov, and A. Tropsha, J. Chem. Inf. Model., 2010, 50, 1189.

171. *Tutorials in Chemoinformatics*, **2017**, ur. Alexandre Varnek, John Wiley & Sons, Ltd, Chichester, UK.

172. G. Idakwo, S. Thangapandian, J. Luttrell, Y. Li, N. Wang, Z. Zhou, H. Hong, B. Yang, C. Zhang, i P. Gong, *J. Cheminform.*, **2020**, *12*, 1.

173. G. Landrum, RDKit: Open-Source Cheminformatics Software, http://rdkit.org/. , preuzeto 01. travnja 2020. god.

174. S. A. Wildman i G. M. Crippen, *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 868.

175. K. Ito, K. R. Weinberger, G. S. Robinson, P. E. Sheffield, R. Lall, R. Mathes, Z. Ross, P. L. Kinney, i T. D. Matte, *Environ. Heal. A Glob. Access Sci. Source*, **2015**, *14*, 1.

176. E. Darcq i B. L. Kieffer, *Nat. Rev. Neurosci.*, **2018**, *19*, 499.

177. A. M. Drewes, R. D. Jensen, L. M. Nielsen, J. Droney, L. L. Christrup, L. Arendt-Nielsen, J. Riley, i A. Dahan, *Br. J. Clin. Pharmacol.*, **2013**, *75*, 60.

178. G. D. Bossé i R. T. Peterson, *Behav. Brain Res.*, **2017**, *335*, 158.
179. D. L. Cunha, F. G. de Araujo, i M. Marques, Psychoactive drugs: occurrence in aquatic environment, analytical methods, and ecotoxicity—a review, Springer Verlag, **2017**.
180. G. Mikus i J. Weiss, *Curr. Pharmacogenomics*, **2005**, *3*, 43.
181. European Monitoring Centre for Drugs and Drug Addiction, "*Europe Drug report 2017*", **2017**.
182. J. L. Scavone, R. C. Sterling, i E. J. Van Bockstaele, *Neuroscience*, **2013**, *248*, 637.
183. K. T. Ahmed, M. R. Amin, P. Shah, i D. W. Ali, *Sci. Rep.*, **2018**, *8*, 1.
184. R. Länge i D. Dietrich, *Toxicol. Lett.*, **2002**, *131*, 97.
185. S.-S. Choi, S.-H. Cha, i C. C. Tappert, A Survey of Binary Similarity and Distance Measures., https://www.researchgate.net/publication/266496381_A_Survey_of_Binary_Similarity_and_Distance_Measures, preuzeto 01. travnja 2020. god.
186. J. Hemmerich, E. Asilar, i G. F. Ecker, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, **2019**, *11731 LNCS*, 788.
187. L. Wu, R. Huang, I. V Tetko, Z. Xia, J. Xu, i W. Tong, *Chem. Res. Toxicol.*, **2020**.
188. B. Lučić, J. Batista, V. Bojović, M. Lovrić, A. Sović Kržić, D. Bešlo, D. Nadramija, i D. Vikić-Topić, *Croat. Chem. Acta*, **2019**, *92*.
189. M. Lovrić, O. Malev, G. Klobučar, R. Kern, J. J. Liu, i B. Lučić, *Molecules*, **2021**, *26*, 1617.
190. S. Scholz, N. Klüver, i R. Kühne, Analysis of the relevance and adequateness of using Fish Embryo Acute Toxicity (FET) Test Guidance (OECD 236) to fulfil the information requirements and addressing concerns under REACH, **2016**., www.umweltbundesamt.de%2Fen%2Ffish-embryo-acute-toxicity-fet-test-workshop-report&usg=AOvVaw3D-iysd8LU7zQTcQc1XgfH , preuzeto 15. ožujka 2019. god.
191. V. M. Alves, E. N. Muratov, S. J. Capuzzi, R. Politi, Y. Low, R. C. Braga, A. V. Zakharov, A. Sedykh, E. Mokshyna, S. Farag, C. H. Andrade, V. E. Kuz'Min, D. Fourches, and A. Tropsha, Green Chem., 2016, 18, 4348.
192. M. Gütlein i S. Kramer, *J. Cheminform.*, **2016**, *8*, 1.
193. G. Landrum, RDKit: Colliding Bits III, http://rdkit.blogspot.com/2016/02/colliding-bits-iii.html. , preuzeto 01. travnja 2020. god.
194. I. Freshney, "*Cell Culture Methods for In Vitro Toxicology*", **2001**, Springer Netherlands, 9.
195. K. Kurosaki, R. Wu, i Y. Uesawa, *Int. J. Mol. Sci.*, **2020**, *21*, 1.

196. A. Abdelaziz, H. Spahn-Langguth, K. W. Schramm, i I. V. Tetko, *Front. Environ. Sci.*, **2016**, *4*, 1.
197. P. Banerjee, A. O. Eckert, A. K. Schrey, i R. Preissner, *Nucleic Acids Res.*, **2018**, *46*, W257.
198. Yuan, Wei, Guan, Jiang, Wang, Zhang, i Li, *Molecules*, **2019**, *24*, 3383.
199. R. Huang, M. Xia, D.-T. Nguyen, T. Zhao, S. Sakamuru, J. Zhao, S. A. Shahane, A. Rossoshek, i A. Simeonov, *Front. Environ. Sci.*, **2016**, *3*, 85.
200. F. Stefaniak, *Front. Environ. Sci.*, **2015**, *3*, 77.
201. M. N. Drwal, V. B. Siramshetty, P. Banerjee, A. Goede, R. Preissner, i M. Dunkel, *Front. Environ. Sci.*, **2015**, *3*, 54.
202. Y. Uesawa, *Front. Environ. Sci.*, **2016**, *4*, 9.
203. K. Ribay, M. T. Kim, W. Wang, D. Pinolini, i H. Zhu, *Front. Environ. Sci.*, **2016**, *4*, 12.
204. A. Koutsoukas, J. St. Amand, M. Mishra, i J. Huan, *Front. Environ. Sci.*, **2016**, *4*, 11.
205. G. Barta, *Front. Environ. Sci.*, **2016**, *4*, 1.
206. S. J. Capuzzi, R. Politi, O. Isayev, S. Farag, i A. Tropsha, *Front. Environ. Sci.*, **2016**, *4*, 3.
207. R. Garg, G. M. Ko, i C. J. Smith, *Toxicol. Res. Appl.*, **2017**, *1*, 239784731770737.
208. S. Padilla, D. Corum, B. Padnos, D. L. Hunter, A. Beam, K. A. Houck, N. Sipes, N. Kleinstreuer, T. Knudsen, D. J. Dix, i D. M. Reif, *Reprod. Toxicol.*, **2012**, *33*, 174.
209. M. Ghorbanzadeh, J. Zhang, i P. L. Andersson, *J. Chemom.*, **2016**, *30*, 298.
210. G. J. Lavado, D. Gadaleta, C. Toma, A. Golbamaki, A. A. Toropov, A. P. Toropova, M. Marzo, D. Baderna, J. Arning, i E. Benfenati, *Ecotoxicol. Environ. Saf.*, **2020**, *202*, 110936.
211. Y. In, S. K. Lee, P. J. Kim, i K. T. No, *Bull. Korean Chem. Soc.*, **2012**, *33*, 613.
212. R. B. Schäfer, V. Pettigrove, G. Rose, G. Allinson, A. Wightwick, P. C. Von Der Ohe, J. Shimeta, R. Kühne, i B. J. Kefford, *Environ. Sci. Technol.*, **2011**, *45*, 1665.
213. X. Feng, D. Li, W. Liang, T. Ruan, i G. Jiang, *Environ. Sci. Technol.*, **2021**, *758*.
214. V. Osorio, A. Larrañaga, J. Aceña, S. Pérez, i D. Barceló, *Sci. Total Environ.*, **2016**, *540*, 267.
215. M. A. Wetzel, D. S. Wahrendorf, i P. C. von der Ohe, *Sci. Total Environ.*, **2013**, *449*, 199.
216. A. J. Williams, C. M. Grulke, J. Edwards, A. D. McEachran, K. Mansouri, N. C. Baker, G. Patlewicz, I. Shah, J. F. Wambaugh, R. S. Judson, i A. M. Richard, *J. Cheminform.*,

2017, *9*, 1.

217. D. M. Reif, M. T. Martin, S. W. Tan, K. A. Houck, R. S. Judson, A. M. Richard, T. B. Knudsen, D. J. Dix, i R. J. Kavlock, *Environ. Health Perspect.*, **2010**, *118*, 1714.
218. N. Kleinstreuer, S. Krishna, i B. Berridge, *Chem. Res. Toxicol.*, **2021**.
219. R. A. Becker, K. P. Friedman, T. W. Simon, M. S. Marty, G. Patlewicz, i J. C. Rowlands, *Regul. Toxicol. Pharmacol.*, **2015**, *71*, 398.
220. B. R. Blackwell, G. T. Ankley, S. R. Corsi, L. A. Decicco, K. A. Houck, R. S. Judson, S. Li, M. T. Martin, E. Murphy, A. L. Schroeder, E. R. Smith, J. Swintek, i D. L. Villeneuve, *Environ. Sci. Technol.*, **2017**, *51*, 8713.
221. S. M. Elliott, W. T. Route, L. A. DeCicco, D. D. VanderMeulen, S. R. Corsi, i B. R. Blackwell, *Environ. Pollut.*, **2019**, *244*, 861.
222. L. D. Rose, D. M. Akob, S. R. Tuberty, S. R. Corsi, L. A. DeCicco, J. D. Colby, i D. J. Martin, *Sci. Total Environ.*, **2019**, *677*, 362.
223. J. Barbosa, K. De Schamphelaere, C. Janssen, i J. Asselman, *Sci. Total Environ.*, **2021**, *758*.
224. T. Braunbeck, B. Kais, E. Lammer, J. Otte, K. Schneider, D. Stengel, i R. Strecker, *Environ. Sci. Pollut. Res.*, **2014**, *22*, 16247.
225. M. Sigurnjak Bureš, Š. Ukić, M. Cvetnić, V. Prevarić, M. Markić, M. Rogošić, H. Kušić, i T. Bolanča, *Environ. Pollut.*, **2021**, *275*, 115885.
226. R. Altenburger, M. Scholze, W. Busch, B. I. Escher, G. Jakobs, M. Krauss, J. Krüger, P. A. Neale, S. Ait-Aissa, A. C. Almeida, T.-B. B. Seiler, F. Brion, K. Hilscherová, H. Hollert, J. Novák, R. Schlichting, H. Serra, Y. Shao, A. Tindall, K. E. Tolefsen, G. Umbuzeiro, T. D. Williams, A. Kortenkamp, K. E. Tollefsen, G. Umbuzeiro, T. D. Williams, i A. Kortenkamp, *Environ. Int.*, **2018**, *114*, 95.
227. Mario Lovrić. (2021). Elektronički dodaci [doktorska disertacija] [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4905282

§ 9. ŽIVOTOPIS

Mario Lovrić rođen je 23. svibnja 1986. godine u Gradačcu, u Bosni i Hercegovini. Maturirao je po programu opće gimnazije u Srednjoj školi Sesvete. Diplomirao je primjenjenu kemiju na Fakultetu kemijskog inženjerstva i tehnologije, Sveučilišta u Zagrebu 2012. godine. Od 2012. do 2015. boravio je u inozemstvu na raznim poslovima u području analitičke kemije. Od ožujka 2015. do lipnja 2017. bio je zaposlen kao istraživač analitičar u PLIVA HRVATSKA d.o.o. u Zagrebu. Tijekom rada u PLIVA HRVATSKA d.o.o. (listopad 2015.) upisao je poslijediplomski studij analitičke kemije na Kemijskom odsjeku Prirodoslovno-matematičkog fakulteta, Sveučilišta u Zagrebu kao izvanredni student. Od srpnja 2017. zaposlen je u Know-Centru, Graz, Austrija kao podatkovni analitičar. Tijekom poslijediplomskog studija boravio je na tri inozemna istraživačka boravka: Sveučilište u Strasbourgu u Francuskoj u laboratoriju , profesora Varneka, na Sveučilištu u Reimsu u Francuskoj u grupi profesora Nuzillarda te u Bolnici Gentofte u Danskoj u grupi profesora Bisgaarda.

Popis znanstvenih, stručnih i kongresnih radova iz teme doktorske disertacije:

1. Lovrić, M.; Malev, O.; Klobučar, G.; Kern, R.; Liu, J.J.; Lučić, B. Predictive Capability of QSAR Models Based on the CompTox Zebrafish Embryo Assays: An Imbalanced Classification Problem. Molecules 2021, 26, 1617, doi:10.3390/molecules26061617.
2. Malev, O.; Lovrić, M.; Stipaničev, D.; Repec, S.; Martinović-Weigelt, D.; Zanella, D.; Ivanković, T.; Sindičić Đuretec, V.; Barišić, J.; Li, M.. Toxicity prediction and effect characterization of 90 pharmaceuticals and illicit drugs measured in plasma of fish from a major European river (Sava, Croatia). Environ. Pollut. 2020, 115162, doi:10.1016/j.envpol.2020.115162.
3. Lovrić, M.; Molero, J.M.; Kern, R. PySpark and RDKit: Moving towards Big Data in Cheminformatics. Mol. Inform. 2019, 38, doi:10.1002/minf.201800082.
4. Lučić, B.; Batista, J.; Bojović, V.; Lovrić, M.; Sović Kržić, A.; Bešlo, D.; Nadramija, D.; Vikić-Topić, D. Estimation of Random Accuracy and its Use in Validation of Predictive Quality of Classification Models within Predictive Challenges. Croat. Chem. Acta 2019, 92, doi:10.5562/cca3551.
5. Babić, S.; Barišić, J.; Stipaničev, D.; Repec, S.; Lovrić, M.; Malev, O.; Martinović-Weigelt, D.; Čož-Rakovac, R.; Klobučar, G. Assessment of river sediment toxicity: Combining

empirical zebrafish embryotoxicity testing with in silico toxicity characterization. Sci. Total Environ. 2018, 643, 435–450, doi:10.1016/j.scitotenv.2018.06.124.

6. Lovrić, M.; Stipaničev, D.; Repec, S.; Malev, O.; Klobučar, G. Combined toxic unit: Moving towards a multipath risk assessment strategy of organic contaminants in river sediments. In Proceedings of the The 10th Eastern European Young Water Professionals Conference; 2018.
7. Lovrić, M. Molekulsko modeliranje odnosa strukturnih svojstava i aktivnosti molekula s pomoću programskog jezika Python (prvi dio). Kem. Ind. 2018, 67, 409–419, doi:10.15255/KUI.2017.052.