

Znaczenie reprezentacji molekularnej w kontekście odkrywania nowych leków o aktywności serotoninergicznej

Natalia Czub, Adam Paćlowski, Aleksander Mendyk

Katedra i Zakład Technologii Postaci Leku i Biofarmacji

Wydział Farmaceutyczny Uniwersytet Jagielloński Collegium Medicum



UNIwersytet JAGIELLOŃSKI
COLLEGIUM MEDICUM



Fundusze Europejskie
Inteligentny Rozwój



Rzeczpospolita
Polska



Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego





Aktywność serotonergiczna

Receptory serotoninowe
Transporter serotoniny (SERT)

5-HT1A

Nastrój - Lęk - Apetyt - Sen



Sztuczna inteligencja uczenie maszynowe

Redukcja czasu oraz kosztów na badania wstępne

QSAR – ang. Quantitative structure–activity
relationship models

pKi

- ujemny logarytm ze stałej inhibicji
- wartość powinowactwa do receptora
- model regresyjny



Cel badań

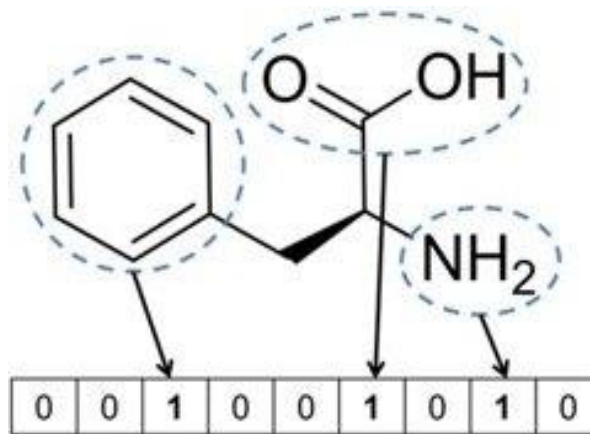
Znalezienie optymalnej reprezentacji molekularnej dla związków o aktywności serotoninerгіcznej

Reprezentacije molekularne

SMILES

NC(CC1=CC=CC=C1)C(O)=O

ABC	ABCGG	nAcid	nBase	SpAbs_A
17.056091	15.719753	0	1	0.0
22.127159	16.464308	0	1	0.0
19.096532	14.438169	0	1	0.0
15.274871	12.790612	0	3	0.0
21.557952	18.622813	2	1	0.0
...
28.369800	19.183183	0	1	0.0
22.615486	15.640544	0	3	0.0
26.886637	18.151908	0	1	0.0
27.484354	18.230660	0	1	0.0
23.989260	16.821512	0	3	0.0



Fingerprint

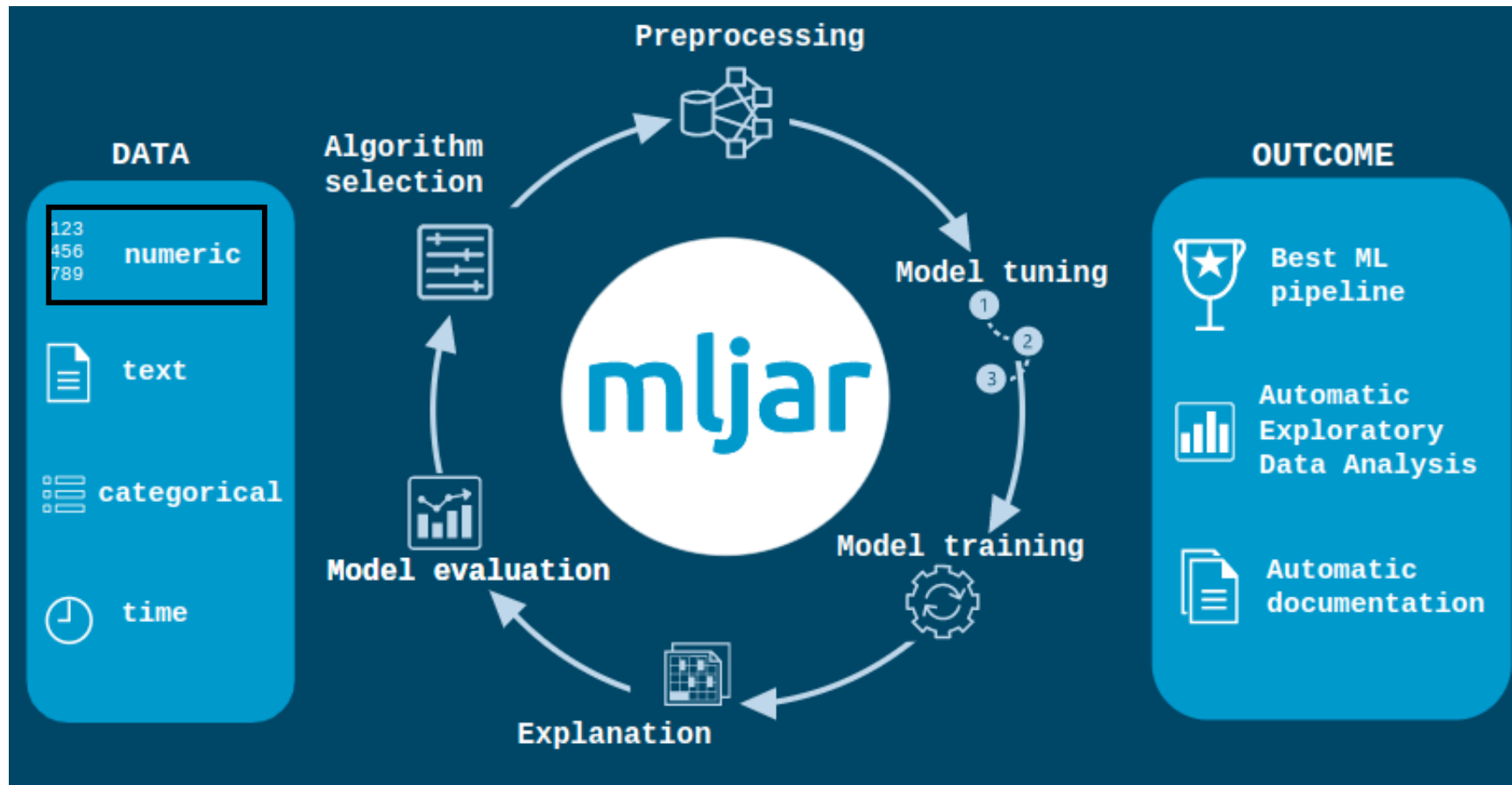
100 – 400 – 900 – 1600 – 2500 - 3600

mol2vec_1	mol2vec_2	mol2vec_3	mol2vec_4	mol2vec_5
-5.843384	-2.102663	9.327226	0.304115	1.493736
-6.755656	-1.459972	6.862629	-3.483530	-1.398833
-6.738349	-1.011717	6.998305	-0.378959	-0.795591
-9.111215	-2.203253	5.998557	0.241626	-1.343988
-9.111215	-2.203253	5.998557	0.241626	-1.343988
...
-10.130441	-4.646593	8.072098	2.555722	2.610086
-5.886534	-4.489444	6.648338	3.240062	1.921277
-11.971634	-3.657933	12.208331	-1.051232	-2.359875
-12.160875	-4.428664	16.099356	3.121636	-0.062192
-7.938127	-5.999627	14.812715	8.430059	3.075828

Mol2Vec

Mordred 2D

Mljar-supervised



<https://github.com/mljar/mljar-supervised/>

Bazy danych

ZINC



Wyselekcjonowana
baza danych

GLASS

ChEMBL

Zbiór uczący - 9440

Zbiór testowy - 735

Tworzenie i ocena modelu

10-krotne wzajemne sprawdzanie

	Data				
1.	Validate	Train	Train	Train	Train
2.	Train	Validate	Train	Train	Train
3.	Train	Train	Validate	Train	Train
...	Train	Train	Train	Validate	Train
k	Train	Train	Train	Train	Validate

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (pred_i - obs_i)^2}{n}}$$

Testowanie na zbiorze zewnętrznym

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^n (pred_i - obs_i)^2}{\sum_{i=1}^n (obs_i - \bar{obs})^2}$$



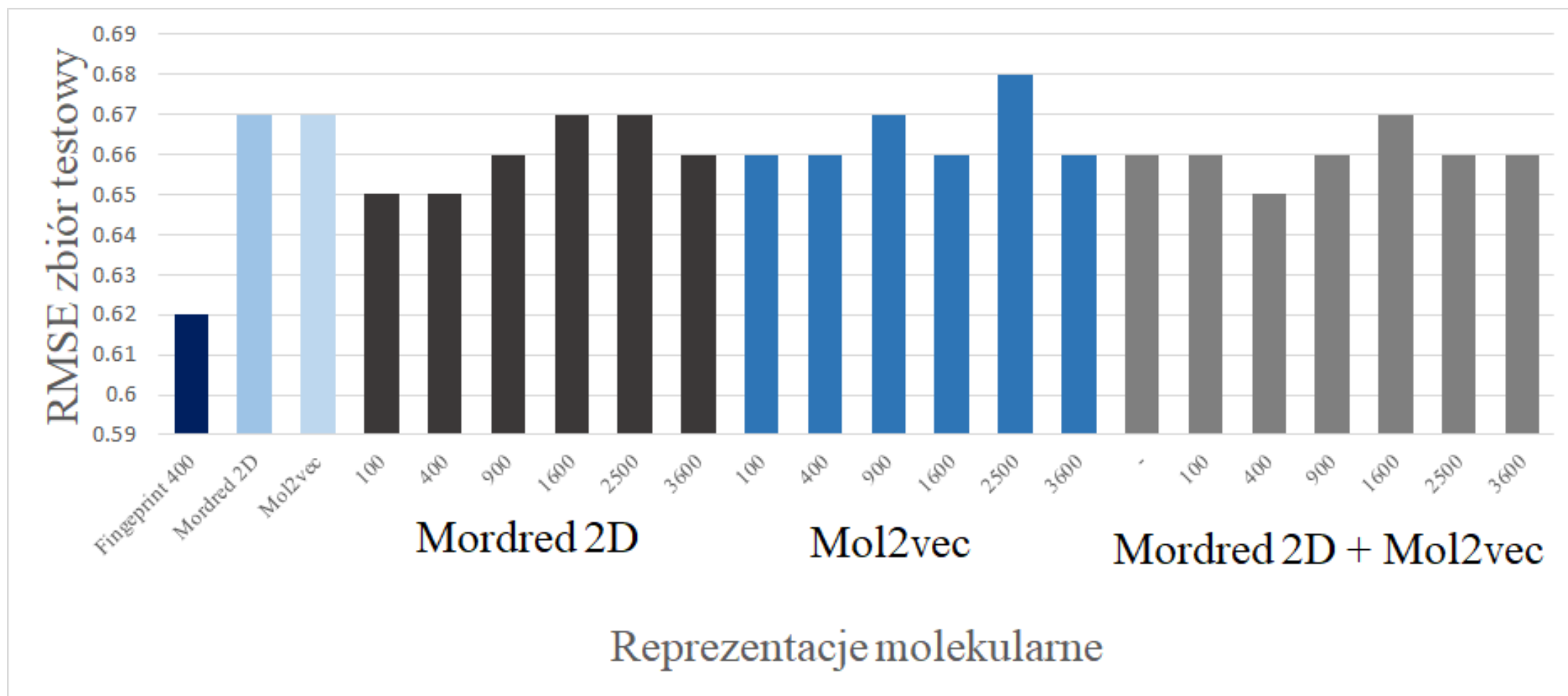
Wyniki - fingerprint

Rozmiar fingerprintu	RMSE 10-CV	R ² 10-CV	RMSE zbior testowy	R ² zbior testowy
100	0,55	0,74	0,66	0,63
400	0,52	0,77	0,62	0,67
900	0,51	0,77	0,65	0,63
1600	0,51	0,77	0,65	0,64
2500	0,50	0,78	0,68	0,61
3600	0,51	0,78	0,65	0,64

Wyniki – reprezentacje molekularne

Reprezentacja molekularna	RMSE 10-CV	R² 10-CV	RMSE zbiorów testowy	R² zbiorów testowy
Fingerprint - długość 400	0,52	0,77	0,62	0,67
Mordred 2D	0,57	0,72	0,67	0,61
Mol2vec	0,58	0,71	0,67	0,62

Wyniki – łączenie reprezentacji molekularnych





Wnioski

- Istnieje wiele rodzajów reprezentacji molekularnych do tworzenia modeli QSAR
- Dla związków o aktywności serotoninerгіcznej najlepszy model utworzony jest w oparciu o binarną reprezentację - fingerprint o długości 400 bitów
- Łączenie reprezentacji nie poprawia modelu

Dziękuję za uwagę!



UNIWERSYTET JAGIELLOŃSKI
COLLEGIUM MEDICUM



Fundusze Europejskie
Inteligentny Rozwój



Rzeczpospolita
Polska



CDT-CARD

Centrum Rozwoju Terapii
Chorób Cywilizacyjnych i Związanych z Wiekiem

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

