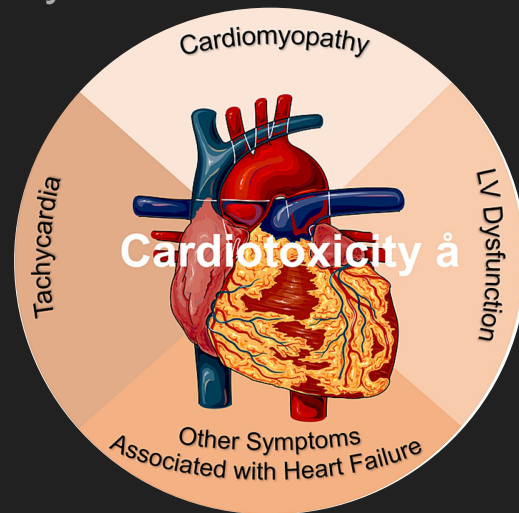


Przewidywanie kardiotoksyczności związków

Mateusz Poleski

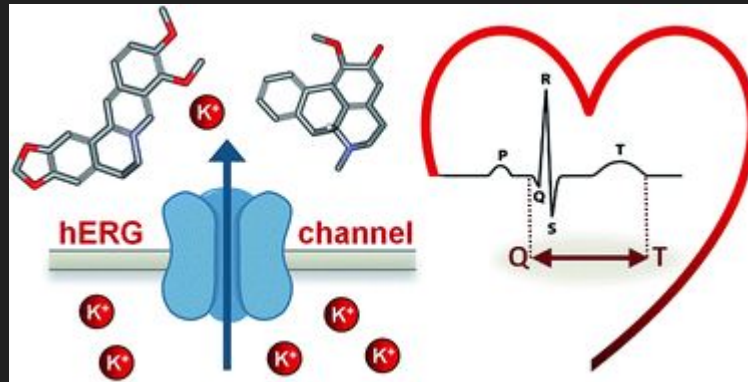
Kardiotoksyczność

- Kardiotoksyczność to różnego rodzaju uszkodzenia mięśnia sercowego lub zaburzenia rytmu jego pracy.
- Potencjalne skutki:
 - nieregularne bicie serca
 - wysokie/niskie ciśnienie krwi
 - zawał serca
- Objawy:
 - Dusznosc
 - Ból w klatce piersiowej
 - Palpitacje serca
 - Zawroty głowy



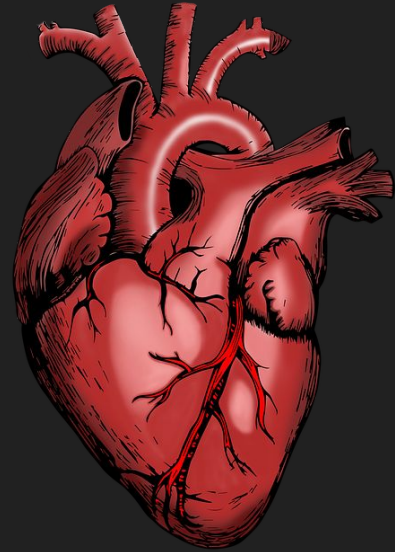
Cele projektu

- Przewidywanie kardiotoksyczności związków poprzez predykcje oddziaływania ich z kanałami potasowymi hERG.
- Identyfikacja cech związków odpowiadających za wiązanie z kanałem potasowym hERG.



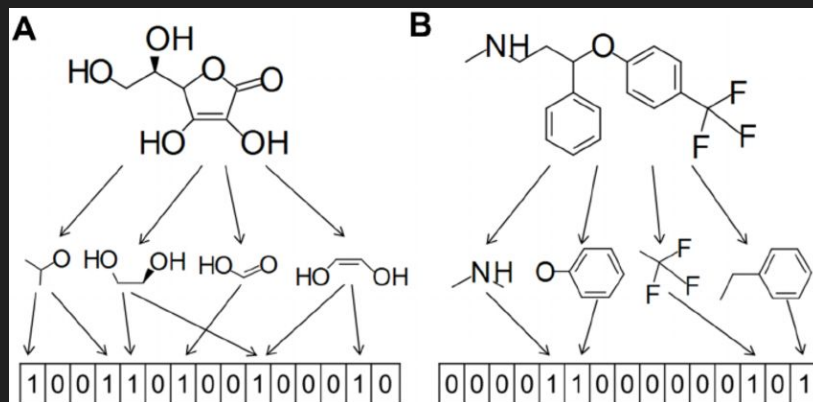
Znaczenie biologiczne

- Przewidywanie czy dana substancja lub lek nie będzie miała negatywnego wpływu na funkcjonowanie serca.
- Identyfikacja cech odpowiedzialnych za toksyczność może pomóc przy tworzeniu nowych leków lub modyfikowaniu starych, tak aby nie wykazywały toksyczności.



Dostępne dane

- **Fingerprinty podstrukturalne**
 - Klekota&Roth, 4860 bitów cech
 - MACCSFP (Molecular ACCess System), 166 bitów cech
- **Fingerprinty haszowane, 1024 bity**
- **11504 rekordów**
- Pliki .sdf



Wstępna analiza

- Stworzenie hybrydowych reprezentacji fingerprintów.
- Stworzenie fingerprintów podstrukturalnych 2D.
- W przypadku kardiotoksyczności mamy znacznie więcej danych (ponad 11 000 rekordów) oraz lepszą jakość w porównaniu do hepatotoksyczności.
- Analizujemy powinowactwa do kanału potasowego hERG, wyrażonego jako IC50, chcemy aby tego powinowactwa nie było.
- IC50 - minimalne stężenie, przy którym zachodzi połowa inhibicji (blokady działania), chcemy żeby IC50 było jak największe.

Planowana implementacja

- Implementacja i porównanie podejścia z klasyfikacją i regresją.
- Na wejściu model otrzymuje fingerprinty, na wyjściu zwraca wartość IC50 lub predykcje czy związek jest toksyczny.
- Klasyfikacja będzie wymagała ustalenia progu co do wartości IC50.

Metody:

Klasyfikacja:

- Regresja logistyczna
- SVM
- Drzewa losowe
- Lasy losowe
- kNN

Regresja:

- Regresja Liniowa
- SVR
- Proste sieci neuronowe

Metryki:

- Accuracy
- F1 score
- AUC
- Mean Squared Error

Powiązana literatura

<https://jcheminf.biomedcentral.com/articles/10.1186/s13321-021-00557-5>