



Politechnika Łódzka

Wydział Fizyki Technicznej, Informatyki
i Matematyki Stosowanej

Informatyka Stosowana	Semestr I
Warsztaty Badawcze	2025/2026
dr hab. inż. Dariusz Puchała	Grupa 7

Sprawozdanie

Zadanie 6*

Agnieszka Sławik
Zuzanna Błaszowska
Emilia Stodulska
Michał Olecki
Konrad Pławik
Piotr Wosiak

* SVN: <https://github.com/kplawik/WarsztatyBadawcze>

Spis treści

1. Przekaz główny artykułu	3
1.1. Kontekst artykułu	3
1.2. Problem badawczy	3
1.3. Stan Wiedzy - Wyniki innych badaczy	3
1.4. Moje badania: Metoda badawcza	3
1.5. Najważniejsze wyniki	4
1.6. Kluczowe wnioski	4
1.7. Jaka z tego wynika historia	4
1.8. Jaki przekaz główny wynika z tej historii	4
1.9. Jak wzbogaci dorobek naukowy?	5
1.10. Co doda do dziedziny naukowej?	5
1.11. Jak pomoże osobom trzecim?	5
2. Szablon artykułu naukowego	6
2.1. Tytuł roboczy	6
2.2. Autorzy	6
2.3. Data	6
2.4. Historia i przekaz główny	6
2.5. Tytuł roboczy	6
2.6. Lista referencji	6
2.7. Wprowadzenie	6
2.8. Metody	7
2.9. Wyniki	7
2.10. Dyskusja	7
2.11. Rysunki/tabele	7
2.12. Wnioski	8
Literatura	9

1. Przekaz główny artykułu

“A Deep Learning Approach to Antibiotic Discovery”

1.1. Kontekst artykułu

Artykuł naukowy przedstawiający przełomowe odkrycie pierwszego nowego antybiotyku po 30 latach (halicyny, charakteryzującej się odmienną strukturą molekularną w porównaniu z aktualnie stosowanymi antybiotykami) poprzez przeprowadzenie najpierw analizy wielu różnych zbiorów danych z zastosowaniem głębokich sieci neuronowych, a następnie testów laboratoryjnych, które potwierdziły działanie bakteriobójcze. Dziedzina badawcza: Artykuł dotyczy zarówno medycyny jak i informatyki ze względu na wykorzystanie wiedzy domenowej o antybiotykach, uczenia maszynowego z dziedziny informatyki oraz elementów statystyki do potwierdzenia zawartych w nim wniosków.

1.2. Problem badawczy

Wytrenowanie głębokiej sieci neuronowej w celu znalezienia molekuly o własnościach bakteriobójczych, ale jednocześnie różniącej się pod względem struktury od znanych już antybiotyków oraz wykazującej jak najniższą toksyczność. Predykcje zostały przeprowadzone na bibliotece zatwierdzonej przez US Food and Drug Administration (FDA), zawierających 1,769 różnych cząsteczek powiększonych o 800 innych naturalnych (uzyskanych np. z roślin), które umożliwiły wyodrębnienie halicyny. Ponadto, przeprowadzono dodatkowe predykcje na bibliotece ZINC15, zawierającej ponad 107 milionów cząsteczek, uzyskując kolejne 23 cząsteczki, które mogą być potencjalnymi antybiotykami. Zbiór liczył ponad 107 milionów molekuł z bazy ZINC15 a do empirycznych testów sieć neuronowa wytypowała 23 molekuly, jedną z nich była właśnie halicyna.

1.3. Stan Wiedzy - Wyniki innych badaczy

Artykuł ma bogatą bibliografię zarówno, jeśli chodzi o stan współczesnej wiedzy o antybiotykach (m.in. „New antibiotics from bacterial natural products” czy “Antibacterial drug discovery in the resistance era”) jak i o wykorzystaniu uczenia maszynowego w biomedycynie (m.in. „Next-generation machine learning for biological networks.”, „A graph-convolutional neural network model for the prediction of chemical reactivity” czy “Using machine learning to predict suitable conditions for organic reactions”). W artykule nie zostały jednak podane wyniki innych badaczy podejmujących próbę predykcji struktury nowych antybiotyków, zostało opisane jedynie ogólne podejście do tematyki oraz obiecujący rozwój głębokich sieci neuronowych.

1.4. Moje badania: Metoda badawcza

1. Dobór i przygotowanie danych: Zbiór uczący składał się z 2335 związków (po usunięciu duplikatów, będących wynikiem złączenia różnych bibliotek): leków zatwierdzonych przez FDA oraz naturalnych produktów,

- testowanych pod kątem zahamowania wzrostu *Escherichia coli*. Wytrenowany zaś model testował ponad 107 milionów molekuł z bazy ZINC15.
2. Reprezentacja molekuł: Cechy 2335 molekuł zostały zwektoryzowane a wiązania międzycząsteczkowe przedstawione w formach grafów skierowanych.
 3. Trening i walidacja modelu: Do treningu użyto sieci konwolucyjnej a konkretnie Directed Message Passing Neural Network (D-MPNN) a skuteczność modelu oceniono głównie za pomocą krzywych ROC-AUC (wartość 0.896)
 4. Predykcja i selekcja: z bazy ponad 107 milionów cząsteczek (molekuł) do testów empirycznych wybrano 23 związki.
 5. Walidacja empiryczna: Przeprowadzono zarówno testy in vitro (czyli wykazano bakteriobójczość na poziomie komórkowym) jak i testy in vivo (czyli wykazano skuteczność na organizmach żywych. W tym wypadku potwierdzono skuteczność Hialicyny na myszach)

1.5. Najważniejsze wyniki

Wyniki są rozległe i dokładnie opisane. Przedstawiają wyniki zarówno z perspektywy walidacji modelu uczenia maszynowego (m.in. analiza krzywych ROC-AUC, typowanie kandydatów z bazy) jak i przeprowadzonych testów in vitro i in vivo (pomiar skuteczności wg modeli Murine, BALB, A.Baumannii czy skuteczność wobec bakterii *C. difficile*). Wyniki bogato ilustrowane dokładnie opisanymi wykresami.

1.6. Kluczowe wnioski

Artykuł przedstawia cztery kluczowe wnioski:

1. Model sieci głębokiej został wytrenowany, aby przewidzieć antybiotyki na bazie już istniejących struktur
2. Hialicyna została wytypowana jako antybakteryjna cząsteczka z bazy Drug Repurposing Hub.
3. Testy na myszach wykazały skuteczność jako antybiotyku
4. Kolejne antybiotyki o odmiennych strukturach zostało przewidzianych na bazie danych ZINC15.

1.7. Jaka z tego wynika historia

Historia jest taka, że nawet jeśli przez 30 lat nie opracowano nowego antybiotyku to warto próbować, ponieważ Uczenie Maszynowe daje możliwość analizy ogromnych zbiorów danych i potrafi zawęzić poszukiwania tak że można zastosować testy empiryczne. Prócz odkrytej hialicyny w/w artykuł nadmienia, że kolejne związki (8 z wytypowanych 23 wykazało pożądane własności) będą również testowane pod kątem użyteczności medycznej.

1.8. Jaki przekaz główny wynika z tej historii

Przekaz jest taki, że interdyscyplinarność umożliwiła opracowanie nowego antybiotyku, co przypomnę, ostatni raz miało miejsce ponad 30 lat temu.

Dodatkowo 8 z 23 wytypowanych związków będzie jeszcze badanych laboratoryjnie, gdyż wykazały one pożądane właściwości.

1.9. Jak wzbogaci dorobek naukowy?

- Artykuł został opublikowany w 2020 roku w czasopiśmie Cell, którego współczynnik impact factor jest bardzo wysoki i wynosi ok. 45,5.
- Artykuł został napisany przez 18 pracowników z USA i Kanady z tytułem doktora lub profesora, specjalizujących się w wielu różnych dziedzinach, w tym farmacji, genetyki, biotechnologii i sztucznej inteligencji.
- Liczba cytowań artykułu wynosi 1,312.

1.10. Co doda do dziedziny naukowej?

Nowy antybiotyk: Hialicynę i 8 związków do kolejnych badań laboratoryjnych.

1.11. Jak pomoże osobom trzecim?

Głównym celem przeprowadzonych badań było odkrycie antybiotyku na bakterie wykazujące odporność na znane już antybiotyki. W artykule wymieniono, że odkryta Hialicyna wykazało skuteczność w walce z bakteriami *E. coli*, *Clostridioides difficile* oraz *Acinetobacter baumannii* – a to już bezpośrednio przełoży się na niższą śmiertelność, krótszą hospitalizację czy zmniejszenie ryzyka występowania epidemii.

Istnie jest również 8 związków wytypowanych do dalszych badań.

Wytrenowany model również być pomocny w podobnych badaniach.

2. Szablon artykułu naukowego

2.1. Tytuł roboczy

“A Deep Learning Approach to Antibiotic Discovery”

2.2. Autorzy

Jonathan M. Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M. Donghia, Craig R. MacNair, Shawn French, Lindsey A. Carfrae, Zohar Bloom-Ackermann, Victoria M. Tran, Anush Chiappino-Pepe, Ahmed H. Badran, Ian W. Andrews, Emma J. Chory, George M. Church, Eric D. Brown, Tommi S. Jaakkola, Regina Barzilay, James J. Collins

2.3. Data

20 lutego 2020

2.4. Historia i przekaz główny

Przekaz jest taki, że interdyscyplinarność umożliwiła opracowanie nowego antybiotyku, co przypomnę, ostatni raz miało miejsce ponad 30 lat temu. Dodatkowo 8 z 23 wytypowanych związków będzie jeszcze badanych laboratoryjnie, gdyż wykazały one pożądane właściwości.

2.5. Tytuł roboczy

Prawdopodobnie taki jak wersji ostatecznej: “A Deep Learning Approach to Antibiotic Discovery”

2.6. Lista referencji

Bibliografia liczyła 70 pozycji. Zarówno dotyczących medycyny, biologii i sieci neuronowych. Najstarsza z pozycji jest z roku 1977 (A protonmotive force drives bacterial flagella) a najnowsze to kilka pozycji z roku 2019, czyli mniej niż rok przed publikacją. Pozwala to przypuszczać że autorzy dogłębnie zbadali zarówno najnowsze odkrycia z medycyny (A multiplexable assay for screening antibiotic lethality against drug-tolerant bacteria) jak i zagadnień uczenia maszynowego (A white-box machine learning approach for revealing antibiotic mechanisms of action)

2.7. Wprowadzenie

Artykuł wprowadza nas w problem jakim są szczepy bakterii odporne na działanie antybiotyków. Następnie nakreśla pomysł jakim było wykorzystanie baz danych cząsteczek (molekuł) i wyszukanie w nich, z pomocą Uczenia Maszynowego, związków o właściwościach. Do badań laboratoryjnych wytypowano 23 związki z ponad 107 milionów (baza ZINC15).

2.8. Metody

Cały opisany proces liczył wiele metod:

1. Dobór i przygotowanie danych: Zbiór uczący składał się z 2335 związków: leków zatwierdzonych przez FDA oraz naturalnych produktów, testowanych pod kątem zahamowania wzrostu *Escherichia coli*. Wytrenowany zaś model testował ponad 107 milionów molekuł z bazy ZINC15.
2. Reprezentacja molekuł: Cechy 2335 molekuł zostały zwektoryzowane a wiązania międzycząsteczkowe przedstawione w formach grafów skierowanych.
3. Trening i walidacja modelu: Do treningu użyto sieci konwolucyjnej a konkretnie Directed Message Passing Neural Network (D-MPNN) a skuteczność modelu oceniono głównie za pomocą krzywych ROC-AUC (wartość 0.896)
4. Predykcja i selekcja: z bazy ponad 107 milionów cząsteczek (molekuł) do testów empirycznych wybrano 23 związki.
5. Walidacja empiryczna: Przeprowadzono zarówno testy *in vitro* (czyli wykazano bakteriobójczość na poziomie komórkowym) jak i testy *in vivo* (czyli wykazano skuteczność na organizmach żywych. W tym wypadku potwierdzono skuteczność Hialicyny na myszach)

2.9. Wyniki

Wynikiem całego procesu opisanego w artykule jest niewątpliwie opracowanie Hialicyny. Dodatkową wartością są: 23 związki przeznaczone do dalszych badań laboratoryjnych, wytrenowany już model Uczenia Maszynowego oraz empiryczny dowód na to, że warto stosować algorytmy Uczenia Maszynowego w dziedzinach, gdzie pozornie nie widzi się już pola do odkryć.

Sama hialicina prócz skuteczności wobec *E. coli* okazała się jeszcze wysoce skuteczna wobec *A. baumannii*, *M. tuberculosis* oraz *P. aeruginos*.

2.10. Dyskusja

Wstępem do części poświęconej dyskusji było nakreślenie problemu jakim była odporność niektórych bakterii na wcześniej znane antybiotyki. Dyskusja uzasadniała dobór danych, zastosowane metody nauki czy wytypowane predykcje.

Dalsza dyskusja podkreślała niedoskonałości sieci neuronowych jak i potencjalne niskie zróżnicowanie danych. Kolejnym poruszonym problemem były już same predykcje i czy faktycznie będą one miały właściwości bakteriobójcze.

W ostatnim akapicie dyskusja dawała potwierdzenie założonym wnioskom m.in. poprzez badania laboratoryjne z dziedziny biomedycznej.

2.11. Rysunki/tabele

Każdy z etapów przedstawiał wyniki czyniąc je czytelnymi:

- Schemat blokowy całego postępowania udostępniony był w formie graficznej (rysunek 1.)

- Wyniki walidacji treningu sieci neuronowej i najistotniejszych badań laboratoryjnych opatrzone dziewięcioma wykresami m.in. krzywe ROC-AUC, wartości OD dla 99 najlepszych predykcji, spadek rozwoju bakterii E.coli zależny od stężenia Hialicyny.
- Szczegółowe wyniki badania opatrzone kilkudziesięcioma wykresami m.in. spadki liczebności pozostałych opisanych bakterii od stężenia Hialicyny.
- Tabelaryczne wykazy testowanych bakterii
- Tabelaryczne wykazy innych związków niż Hialicyna przeznaczonych do dalszych testów
- Tabelaryczny wykaz testowanych algorytmów i modeli (m.in. Chemprop, RDKit, BWA, edgeR)

2.12. Wnioski

Główne wnioski są takie:

- Zastosowanie Ucznia Maszynowego umożliwiło opracowanie nowego antybiotyku (30 lat po ostatnim opracowanym).
- Nawet wielkie zbiory danych nie muszą stanowić problemu, jeśli wytrenowany model okaże się skuteczny.
- Pomimo koncentracji badań na zwalczaniu bakterii E. Coli Hialicyna okazała się również wysoce skuteczna wobec A. baumannii, M. tuberculosis oraz P. aeruginosa (prócz tego kilka innych również zostało wymienionych)

Literatura

- [1] J. M. Stokes et al., “A Deep Learning Approach to Antibiotic Discovery,” *Cell*, vol. 180, no. 4, pp. 688-702.e13, Feb. 2020, doi: 10.1016/j.cell.2020.01.021.