

МОЛБА
ЗА ОДОБРАВАЊЕ ТЕМЕ МАСТЕР РАДА

Молим да се одобри израда мастер рада под насловом:

„Класификација типова коронавируса на основу карактеристика протеина представљених структурним алфабетом”

Значај теме и области:

Један од значајних проблема у области биоинформатике односи се на анализу и класификацију вирусних протеина. Посебно интересовање научне заједнице последњих година привукли су коронавируси због глобалних здравствених последица вируса *SARS-CoV* и *SARS-CoV-2*.

Репрезентација биолошких секвенци, а нарочито протеинских структура, у облику погодном за примену различитих рачунарских метода и алата представља посебан изазов. У том контексту, структурни алфабети омогућавају представљање локалних просторних карактеристика протеина у дискретном облику погодном за анализу. Структурни алфабет под називом *Протеински блокови*, који ће бити коришћен у овом раду, често се примењује у савременим истраживањима за опис локалне структуре протеина и примену метода истраживања података.

Специфични циљ рада:

Циљ рада би био испитивање могућности класификације различитих типова коронавируса на основу структурних карактеристика шпицастих (спајк) протеина представљених структурним алфабетом *Протеински блокови*. У оквиру рада формирао би се скуп података који би обухватао експериментално одређене 3D структуре спајк протеина различитих типова коронавируса. Над овим протеинима применила би се трансформација у секвенце протеинских блокова, након чега би се издвајале карактеристике као што су учесталост појединачних блокова, n-грам шаблони и евентуално карактеристике засноване на понављајућим структурним мотивима (рипитима). Над добијеним карактеристикама примениле би се методе истраживања података, као што су дрвета одлучивања, метод потпорних вектора, неуронске мреже, ансамбл методе и друге релевантне методе класификације, са циљем изградње модела за класификацију типова коронавируса. Додатно би се испитало да ли проширење скупа података структурама предвиђеним помоћу алата *AlphaFold* може допринети побољшању резултата класификације код типова вируса са ограниченим бројем експериментално одређених структура.

У раду би се користиле базе података *Protein Data Bank* и *NCBI Protein* за добијање података о експериментално одређеним 3D структурама спајк протеина коронавируса. Систем *AlphaFold/ColabFold* за предвиђање 3D структуре протеина би се користио за повећање броја инстанци у скупу података. За трансформацију структура у протеинске блокове и примену метода машинског учења користиле би се библиотеке *PBxplore*, *scikit-learn* и други релевантни алати из области биоинформатике и машинског учења.

Литература:

1. Burrell, Christopher J., Howard, Colin R., Murphy, Frederick A.: *Coronaviruses*, Fenner and White's Medical Virology, Elsevier, 2017
2. Etchebest, Catherine, Benros, Cristina, Hazout, Serge, de Brevern, Alexandre G.: *A structural alphabet for local protein structures: improved prediction methods*, Proteins, 2005

(име и презиме студента, бр. индекса, смер и модул)

(својеручни потпис студента)

<датум>
(датум подношења молбе)

(својеручни потпис ментора)

Чланови комисије

1. Јована Ковачевић
2. Саша Малков

Катедра <катедра> је сагласна са предложеном темом.

(шеф катедре)

(датум одобравања молбе)