

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Sekvenční homologie cirkulární RNA
Jméno autora:	Bsc. Eliška Širůčková
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Vedoucí práce:	Petr Ryšavý
Pracoviště vedoucího práce:	Katedra počítačů, FEL, ČVUT

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Cílem práce bylo navrhnout nový způsob predikce interakcí mezi miRNA a circRNA molekulami. Současné metody jsou většinou založené na využití vypočítaných vlastností známých interakcí, především pak na předpokladu o párování na kratším regionu, zvaném seed. V práci bylo třeba přijít s novým přístupem založeným na strojovém učení a k tomu vytvořit dostatečnou trénovací množinu a ohodnotit úspěšnost navržené metody. Hodnotím tedy zadání jako náročnější.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Zadání bylo splněno. Práce obsahuje rozsáhlou rešerši, popis funkce cirkulární RNA, návrh metody a experimentální evaluaci. Studentka navíc musela přijít s daty, která použila jako trénovací. V závěru by bylo vhodné například vyzkoušet například více přístupů pro učení výsledného klasifikátoru, nicméně zadání je i tak zcela splněno.	

Aktivita a samostatnost při zpracování práce	A - výborně
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
Studentka pracovala na práci samostatně a pravidelně konzultovala postup řešení. Sama přicházela s nápady, jak dosáhnout cíle a vylepšit existující postupy. Průběžné termíny nebylo nutné stanovovat, studentka pracovala průběžně během projektu i předmětu Diplomová práce, nicméně několikrát došlo k mírnému skluzu (například během načítání literatury či tvorby klasifikátoru založeném na sekundární struktuře), kdy studentka řešila do přílišného detailu některé z podcílů.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
V této části nemám výhrad. Práce je psaná informovaně, studentka musela pro její sepsání nastudovat velké množství literatury, která přesahuje vyučované předměty, a také načíst velké množství odborných článků. Kromě toho bylo třeba sestavit trénovací množinu, což představovalo vyřešení mnoha problémů, například nedostupnost jedné z databází v určité denní době či nutnost strojově číst primární struktury cirkulární RNA z webových stránek. Stejně tak bylo nutné přijít s vhodnými pozitivními a negativními příklady pro algoritmy strojového učení.	

Formální a jazyková úroveň, rozsah práce	A - výborně
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psaná výbornou a srozumitelnou angličtinou, formálně a typograficky správně. Především první verze trpěly na nevyrovnanost v úrovni detailu, do které jsou popisovány například existující metody, nicméně ve finální verzi jsou tyto nedostatky opraveny. V práci je jen málo překlepů, mohu například zmínit nekonzistentní kapitalizaci nadpisů.	

Výběr zdrojů, korektnost citací

A - výborně

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Práce obsahuje 72 citací, z velké části jde o vědecké články, doplněné částečně webovými odkazy, například u některých z použitých knihoven. Citace studentka sama aktivně vyhledávala a zahrnují všechny využití oblasti ze strojového učení, biologie a bioinformatiky. Převzaté části, například obrázky, jsou označené jako přetisky včetně uvedení zdroje.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Studentka kombinuje vlastní kódy v jazyce python s externími knihovnamí. Kódy jsou přehledné, nicméně některé věci jsou řešené nevhodně, například přepínání mezi datovými sadami je provedeno pomocí komentářů. Vhodnější by bylo mít dané nastavení jako součást parametrů programu nebo ve speciálním souboru s nastavením. Odevzdané kódy z KOSu bohužel nejde přímo spustit bez dalších dat, protože naráží na nastavený limit 50MB pro přílohy diplomových prací, nicméně s daty, které jsem měl k dispozici ve sdíleném repositáři fungují bez problémů.

III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.

Cílem práce bylo přijít s novou metodou pro klasifikaci interakcí mezi circRNA a miRNA. Během psaní práce bylo pracováno s hypotézou, že to, co odlišuje interakce circRNA a miRNA od mnohem lépe prozkoumaných interakcí mezi mRNA a miRNA, je sekundární struktura. To se nakonec v experimentálním vyhodnocení metody potvrdilo pouze z části, nicméně i tak se studentce podařilo přijít s klasifikátorem možných interakcí mezi circRNA a miRNA. Tento klasifikátor je založený na neuronové síti, která klasifikuje na základě zarovnání vstupního okna circRNA a miRNA. Kromě toho studentka přišla s klasifikátory, které se rozhodují na základě sekundární struktury, zda se na daném místě může nějaká miRNA vázat. Velký přínos práce vidím i v datasetu, který studentka sestrojila, protože předchozí ad-hoc metody nepotřebovaly pozitivní a negativní trénovací příklady. Studentka vyřešila i neočekávané problémy s přístupem na databáze či definicí negativních vzorků.

Studentka v řešení nepředpokládá znalost seedu, nicméně vstupní data obsahují určitý neodstranitelný bias, protože ověřených interakcí je málo a většinou jsou vytipovány především na základě předchozích metod. Nicméně i tak navržená metoda dosahuje dobrých výsledků, kdy se část predikcí shoduje s předchozími metodami a část neshoduje, čímž metoda navrhuje další kandidáty k ověření.

K obhajobě bych se rád zeptal, který z problémů, jež bylo nutné vyřešit pro dopsání práce byl nejméně očekávaný?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 8.6.2021

Podpis: