

Наставно-научном већу  
Математичког факултета  
Универзитета у Београду

На 305. седници Наставно-научног већа Математичког факултета Универзитета у Београду одржаној 25.06.2012. године, одређени смо за чланове комисије за преглед и оцену рукописа **Примена правила придруживања и метода подржавајућих вектора за предвиђање Т-ћелијских епитопа** који је предат као докторска дисертација кандидата Даворке Јандрлић. Након прегледа рукописа подносимо Наставно-научном већу следећи

## Извештај

### 1 Биографија кандидата

Мр Даворка Јандрлић рођена је 12. фебруара 1981. године у Дубровнику. Основну школу и гимназију завршила је у Херцег Новом. Школске 1999/2000. године уписала је Математички факултет у Београду (смер Рачунарство и информатика), и дипломирала школске 2003/2004. године са просечном оценом 9,10. Школске 2004/2005. године је уписала магистарске студије на Математичком факултету, смер Рачунарство и информатика. Последипломске магистарске студије је завршила 2010. године одбраном магистарске тезе под насловом "Примена техника истраживања података на одређивање корелације између неуређених и антигених региона протеина" под руководством проф. др Ненада Митића.

Од 2004. је запослена на Машинском факултету Универзитета у Београду као асистент приправник, а од 2010. као асистент на катедри за математику. До сада је држала вежбе из следећих предмета: Програмирање, Рачунарски алати, C-C++, Веб пројектовање у машинству и Објектно оријентисано програмирање и Јава. Основне области интересовања су јој базе података, развој и примена техника истраживања података и биоинформатика. Тренутно је истраживач на пројекту Методе нумеричке и нелинеарне анализе са применама, бр. 174002 који финансира Министарство просвете, науке и технолошког развоја Републике Србије. До сада има 5 објављених радова у часописима (од којих четири рада у часописима на *SCI* листи и један самостални), и један самостални рад на рецензији у часопису на *SCI* листи. Учествовала је на више међународних и домаћих конференција на којима има 6 саопштења која су објављена у зборницима радова. Коаутор је једног универзитетског уџбеника.

### 2 Предмет и садржај дисертације

Предмет докторске дисертације припада области истраживања података. Истраживање података је једна од области рачунарства која се најбрже развијала у последње две деценије. Методе истраживања података укључују велики број различитих (најчешће математички заснованих) алгоритама помоћу којих се врши провера података и одређује модел који је по карактеристикама најближи карактеристикама података који се посматрају, при чему може

да се оцени квалитет добијених резултата. Истраживање података обухвата широк скуп метода које укључују класификацију, кластеровање, одређивање правила придруживања, истраживање образаца, методе које раде са подржавајућим векторима, итд.

Тема рада повезује истраживање података и биоинформатику и везана је за конструкцију модела за предвиђање Т-ћелијских епитопа помоћу метода подржавајућих вектора и одређивања правила придруживања. Т-ћелијски епитопи се користе у имунотерапијама и вакцинама које су засноване на пептидима. Њихово идентификовање има кључну улогу у разумевању функционисања имуног система човека. Експерименталне методе за препознавање Т-ћелијских епитопа су веома скупе и временски захтевне. Због тога је јако значајан развој нових рачунарских модела и метода који скраћују време обраде и повећавају прецизност одређивања позиције Т-ћелијских епитопа у протеинима. У дисертацији је представљена нова метода помоћу које може да се повећа прецизност одређивања позиције Т-ћелијских епитопа у протеинима у односу на постојеће методе које се користе у те сврхе. Метода је заснована на две врсте машина заснованих на подржавајућим векторима, при чему су један од улаза физичко-хемијске карактеристике аминокиселина. У раду је такође приказана и употреба правила придруживања на одређивање односа Т-ћелијских епитопа и уређених/неуређених региона протеина.

### **3 Кратак приказ дисертације и оригиналних доприноса**

Рукопис се састоји од 164 странице (XVII+147) и има следећу структуру:

1. Увод
2. Методе истраживања података
3. Предвиђање Т-ћелијских епитопа
4. Нове методе за предвиђање Т-ћелијских епитопа
5. Резултати примене предложених модела
6. Неуређена структура протеина и Т-ћелијски епитопи
7. Закључак и даљи рад

уз Резиме (на српском и енглеском језику), Садржај, Списак слика, Списак табела, Додатке (додатак А, Б, В), Списак литературе и Биографију кандидата.

У уводном поглављу је дат приказ основних појмова и проблема који је обрађен у дисертацији.

Друго поглавље садржи приказ метода истраживања података које су коришћене у дисертацији: класификације, регресије, и методе засноване на подржавајућим векторима које су коришћене у дефиницији предложених метода за откривање Т-ћелијских епитопа. Такође, дат је и приказ метода откривања правила придруживања и кластеровања помоћу к најближих суседа које се користе за одређивање односа Т-ћелијских епитопа и уређених/неуређених региона протеина, односно при провери коректности предложених метода за откривање Т-ћелијских епитопа.

У трећем поглављу је дат преглед база података које садрже експериментално утврђене Т-ћелијске епитопе и МНС везујуће пептиде, као и детаљан приказ методологије рада постојећих метода за откривање Т-ћелијских епитопа и њихових недостатака. Као улазни подаци при

развоју новопредложених метода коришћени су подаци из IEDB 3.1 базе, а за тестирање модела подскуп података из MNCBN базе који нема пресека са IEDB базом.

У четвртом поглављу *Нове методе за предвиђање T-ћелијских епитопа* које представља централни део рада, дефинисани су и детаљно описани новоразвијени модели за предвиђање T ћелијских епитопа. Развијени модели су намењени за квантитативно (пептид јесте/није епитоп) и квалитативно предвиђање (предвиђа се афинитет везивања пептида на основу кога се даље закључује да ли јесте или није епитоп). Модели су засновани на техници подржавајућих вектора примењеној за решавање проблема бинарне класификације и регресије. Подаци који су коришћени за развијање и тестирање модела су преузети из IEDB базе. За припрему података у одговарајући облик за улаз у SVM и SVR моделе развијене су три нове шеме за представљање пептидне секвенце. У оквиру прве шеме за представљање пептида је коришћена  $\Delta$ -BM25-IDF техника за рачунање учесталости аминокиселина у пептиду на одговарајућој позицији која се успешно користи у класификовању докумената дуги низ година, али до сада није коришћена за ову врсту проблема. Техника рачунања учесталости је комбинована са већ постојећим представљањем пептида матрицама супституције. Друга шема представљања пептидних секвенци користи комбинацију физичко хемијских особина аминокиселина за представљање пептида. Физичко хемијске особине за представљање пептида су коришћене и у неким од постојећих метода за предвиђање T-ћелијских епитопа, које међутим имају знатно мању тачност у поређењу са осталим методама. У тези је установљен нови приступ у избору и комбинацији физичко хемијских особина и представљању пептида изабраним физичко хемијским особинама. За избор адекватних („најбољих“) физичко хемијских особина су развијени помоћни модели бинарне класификације засновани на кластер анализи. Сврха помоћних модела је првобитно била издвајање најважнијих атрибута за SVM и SVR моделе, међутим касније се показало да су и помоћни модели довољно добре тачности и упоредиви са постојећим методама за предвиђање T-ћелијских епитопа. Трећа развијена шема за представљање пептида и припрему улаза за SVM и SVR моделе представља проширење прве шеме  $z$ -дескрипторима.  $z$ -дескриптори представљају нумеричке вредности молекуларних особина аминокиселина, добијене применом технике анализе главних компоненти на неколико стотина молекуларних особина аминокиселина. За сваку од три описане шеме представљања пептида су направљена по два модела за скуп од 15 изабраних алела (скуп чине алеле са највећим бројем познатих експерименталних података), један за бинарно предвиђање епитопа други за регресионо предвиђање афинитета везивања пептида. Провера тачности направљених модела је рађена 10-унакрсном провером. Алгоритам који је коришћен за прављење модела заснованих на техници подржавајућих вектора је алгоритам секвенцијалне минималне оптимизације. За бинарне моделе су коришћене полиномијалне функције језгра а за регресионе моделе радијално засноване функције. Параметри модела су оптимизовани методом похлепног претраживања.

У петом поглављу *Резултати примене предложених модела* приказани су резултати провере предложеног модела као и оцена квалитета добијеног модела за различите скупове материјала. Показано је да сви предложени модели имају јако висок проценат тачности предвиђања. У овом поглављу је дато и поређење резултата дефинисаних модела са резултатима добијеним постојећим методама за предвиђање T-ћелијских епитопа. Показује се да у готово свим случајевима један од новопредложених модела даје боље резултате.

У шестом поглављу су приказани резултати односа неуређене структуре протеина и T-ћелијских епитопа. Показано је да могу да се добију правила придруживања која указују на постојање T-ћелијских епитопа MNC класа I и II у неуређеним/уређеним регионима

протеина у којима је заступљеност аминокиселина А, W, L и N већа или мања од просечне заступљености тих аминокиселина.

У седмом поглављу *Закључак и даљи рад* је дат сумарни приказ садржаја дисертације и описани могући правци обједињавања све три предложене шеме у јединствен модел који би дао боље резултате за шири скуп алела.

У Додатку А су приказани резултати класификационих модела заснованих на кластеровању. Додатак Б садржи табелу најбољих (за употребу у процесу откривања Т-ћелијских епитопа) физичко-хемијских особина за монограме и биграме, док Додатак В садржи кратак опис EpDis-MassPred система.

Списак литературе се састоји од 119 библиографских јединица.

## 4 Радови

Резултате приказане у овом рукопису кандидат је публиковала у четири рада (три на СЦИ листи и један самосталан), док је још један самосталан рад је у процесу рецензије:

1. N. Mitić, M. Pavlović, D. Jandrlić (2014). *Epitope distribution in ordered and disordered protein regions - part A. T-cell epitope frequency, affinity and hydrophathy*, Journal of Immunological Methods, vol. 406 pp. 83-103 (2014), (SCI - IF 1.820)
2. M. Pavlović, D. Jandrlić, N. Mitić (2014). *Epitope distribution in ordered and disordered protein regions. Part B - Ordered regions and disordered binding sites are targets of T- and B-cell immunity*, Journal of Immunological Methods, vol. 407, pp. 90-107. (SCI - IF 1.820)
3. D. Jandrlić, G. Lazić, N. Mitić, M. Pavlović *Software tools for simultaneous data visualization and T cell epitopes and disorder prediction in proteins*, Journal of Biomedical Informatics, Vol. 60, pp. 120–131, 2016, (SCI - IF 2.126)
4. D. Jandrlić, *The rule based classification models for MHC binding prediction and identification of the most relevant physicochemical properties for the individual allele*. University thought - Publication in Natural Sciences. doi:10.5937/univtho6-10768 (2016)

Кандидат је коаутор у још једном раду на СЦИ листи који није повезан са темом дисертације: S. Radenović, Z. Kadelburg, D. Jandrlić, A. Jandrlić: *Some Results on Weakly Contractive Maps*. Bulletin of the Iranian Mathematical Society, 38 (3), 625-645. (2012) (SCI - IF 0.223)

## 5 Закључак

Рукопис **Примена правила придруживања и метода подржавајућих вектора за предвиђање Т-ћелијских епитопа** садржи вредан научни допринос у области истраживања података и његове примене у биоинформатици. У раду је разматран проблем предвиђања Т-ћелијских епитопа. У току рада су дефинисани нови модели од којих су најзначајнији они засновани на SVM и SVR техникама истраживања података, као и на класификацији на основу резултата добијених кластеровањем експериментално потврђених Т-ћелијских епитопа. Резултати теста модела су показали да је њихова прецизност у највећем броју случајева већа од прецизности модела који се тренутно користе за решавање овог проблема. У раду су приказана и правила придруживања која указују на повезаност постојања Т-ћелијских епитопа у неуређеним/уређеним регионима протеина и заступљености појединих

амино киселина у тим регионима.

Имајући у виду претходно наведено предлажемо Наставно-научном већу Математичког факултета да рукопис **Примена правила придруживања и метода подржавајућих вектора за предвиђање Т-ћелијских епитопа** кандидата Даворке Јандрлић прихвати као докторску дисертацију и одреди комисију за њену одбрану.

У Београду, 03.06.2016.

Чланови комисије за преглед и оцену

---

(проф. др Ненад Митић, ванр. проф.)

---

(проф. др Гордана Павловић-Лажетић, ред. проф.)

---

(др Мирјана Павловић, виши научни сарадник)