SVEUČILIŠTE U RIJECI

**TEHNIČKI FAKULTET**

Diplomski sveučilišni studij računarstva

Diplomski rad

**ANALIZA SLIČNOSTI PEPTIDA TEMELJENA NA LATENTNOM PROSTORU AUTOENKODERA**

Rijeka, rujan 2023. Bruno Novosel   
 0069078199

SVEUČILIŠTE U RIJECI

**TEHNIČKI FAKULTET**

Diplomski sveučilišni studij računarstva

Diplomski rad

**ANALIZA SLIČNOSTI PEPTIDA TEMELJENA NA LATENTNOM PROSTORU AUTOENKODERA**

Mentor: doc. dr. sc. Goran Mauša

Rijeka, rujan 2023. Bruno Novosel   
 0069078199

IZJAVA

Izjavljujem da sam ovaj rad izradio samostalno, uz vodstvo i stručnu pomoć mentora doc. dr. sc. Gorana Mauše i asistenta Marka Njirjaka.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Bruno Novosel

ZAHVALA

Zahvaljujem mentoru doc. dr. sc. Goranu Mauši i asistentu Marku Njirjaku na pomoći i stručnom vođenju kojim su me vodili kroz izradu ovog rada i što su mi uvijek bili na raspolaganju.

Sadržaj

[1. UVOD 1](#_Toc142396633)

[2. PEPTIDI 2](#_Toc142396634)

[2.1. Što su peptidi? 2](#_Toc142396635)

[2.2. Antimikrobna svojstva 4](#_Toc142396636)

[2.3. Antiviralna svojstva 6](#_Toc142396637)

[3. AUTOENKODERI 7](#_Toc142396638)

[3.1. Autoenkoderi općenito 7](#_Toc142396639)

[3.2. Varijacijski autoenkoderi 7](#_Toc142396640)

[4. MODEL 8](#_Toc142396641)

[4.1. Podaci 8](#_Toc142396642)

[4.2. Metode (sami model) 8](#_Toc142396643)

[5. REZULTATI 8](#_Toc142396644)

[6. ZAKLJUČAK 9](#_Toc142396645)

[LITERATURA 10](#_Toc142396646)

[POPIS KRATICA 13](#_Toc142396647)

[Sažetak 14](#_Toc142396648)

[Abstract 14](#_Toc142396649)

# UVOD

Zahvaljujući napretku medicine i znanosti, otkrivena su i proizvedena mnoga cjepiva i lijekovi koji su značajno doprinijeli suzbijanju raznih bolesti kod ljudi. Unatoč tome, mnoge bolesti i dalje predstavljaju veliku prijetnju ljudskom zdravlju te se neprestano traže nova rješenja za njihovo suzbijanje. U tom kontekstu, peptidi su se istaknuli kao obećavajuće novo rješenje pri liječenju. Iako se još uvijek smatraju pomalo nekonvencionalnim pristupom liječenju raznih bolesti i infekcija i relativno su brojčano ograničeni, peptidi su pokazali obećavajuće rezultate u mnogim istraživanjima [1]. U prirodi postoje peptidi s različitim funkcijama, ali ograničena raznolikost sekvenci unutar ovih molekula ograničava i spektar njihove primjene pa se zbog toga pokušava sintetizirati nove peptide koji će imati željena svojstva. Međutim, predviđanje svojstava novo sintetiziranih peptida i poboljšanje postojećih metoda predviđanja još uvijek ima prostora za napredak.

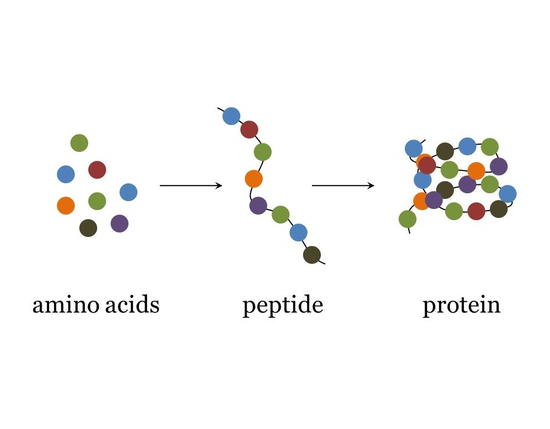
Temeljem dosadašnjih uspjeha metoda strojnog učenja na ovom području, u ovom radu koristi se varijacijski autoenkoder kako bi veliki broj peptida opisanih njihovim fizikalno-kemijskim svojstvima prikazali u latentom prostoru autoenkodera. S obizrom na to da su se brojna fizikalno-kemijska svojstva peptida pokazala kao potencijalni indikatori antimikrobne ili antiviralne učinkovitosti [2], u latentnom prostoru tada možemo promatrati sličnost među pojedinim vrstama i podvrstama peptida sa željenim značajkama. Također, može se postaviti pitanje postoji li korelacija između takve sličnosti, dobivene u latentnom prostoru autoenkodera, i sličnosti izračunate prema sekvencama pojedinih peptida.

Ovaj rad predstavlja metode i rezultate istraživanja koji prikazuju potencijal korištenja strojnog učenja za istraživanje peptida u svrhu otkrivanja novih lijekova. Korištenjem ovakvih računalnih metoda, cilj je ubrzati otkrivanje i dizajn antimikrobnih i antiviralnih peptida s poboljšanom učinkovitošću.

1. PEPTIDI

## Što su peptidi?

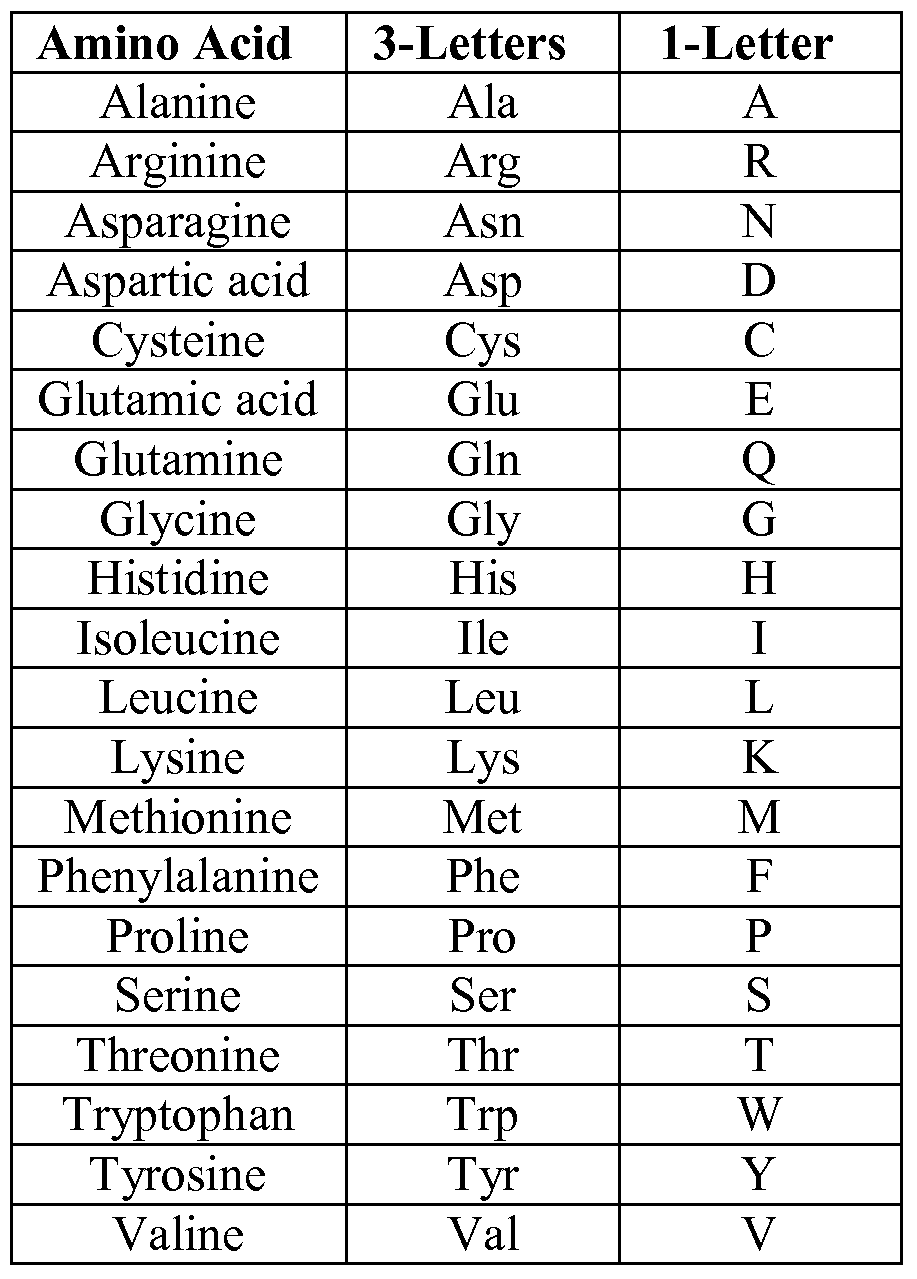
Peptidi su biomolekule prisutne u svim živim organizmima. Kao i proteini, sastavljeni su od aminokiselina povezanih peptidnim vezama u polipeptidne lance, a razlikuju se po broju aminokiselina od kojih se sastoje. Polipeptidne lance do 50 aminokiselina smatramo peptidima, a s više od 50 smatramo proteinima [3]. Pojednostavljeni prikaz aminokiselina, peptida i proteina nalazi se na slici 2.1.

  
*Slika 2.1. Pojednostavljeni prikaz aminokiselina, peptida i proteina (preuzeto iz [4])*

U prirodi se pojavljuje dvadeset standardnih aminokiselina. Svaka aminokiselina ima jedinstvenu strukturu i svojstva te one među sobom mogu stvarati peptidne veze i tako se povezivati u lance. Popis svih prirodnih aminokiselina prikazan je slikom 2.2.

Ovisno o broju aminokiselina od kojih se sastoje, peptidi se mogu klasificirati kao dipeptidi, tripeptidi, oligopeptidi ili polipeptidi. Također se mogu podijeliti na linearne i cikličke, ovisno o načinu povezivanja aminokiselina.

Peptidi imaju raznolike uloge u biološkim sustavima, kao što su signalizacija, obrana od patogena, regulacija enzimskih aktivnosti i strukturna podrška. Jedna od posebno bitnih karakteristika za ovaj rad je da peptidi mogu imati antimikrobnu aktivnost, boreći se na taj način protiv bakterija, gljivica, parazita i virusa. Takvi peptidi, poznati kao antimikrobni peptidi (AMP), imaju široki spektar djelovanja i važni su u razvoju novih lijekova.



*Slika 2.2. Popis svih prirodnih aminokiselina i njihove oznake (preuzeto iz [5])*

Interes istraživača za razvoj novih peptida i proučavanje struktura i funkcija ciljnih receptora znatno je porastao posljednjih godina. Suvremeno medicinsko i biokemijsko istraživanje ima značajan fokus na peptide zbog njihove selektivnosti, specifičnosti i snažne interakcije s ciljanim proteinima. Svojom veličinom i površinom omogućuju preciznije vezanje na ciljne molekule. Mogu imati vrlo selektivno djelovanje, smanjujući rizik od nuspojava, ali brzo se metaboliziraju te imaju kratko djelovanje u tijelu. Aktivnost peptida može se produljiti uvođenjem raznih modifikacija.

Peptidi također imaju neke prednosti u usporedbi s proteinima kada govorimo o liječenju, iako proteini zauzimaju sve veći udio na farmaceutskom tržištu posljednjih nekoliko godina. Proteini, iako su često vrlo sigurni i učinkoviti, moraju se proizvoditi u bioreaktorima koji koriste cijele stanice, a njihovo pročišćenje i strukturna analiza često su složeni i skupi. S druge strane, peptidi se često mogu dobiti kemijski, a njihovo pročišćenje i analiza su mnogo jednostavniji. Također, sve je više primjera oralno djelotvornih peptida, što ih čini poželjnijima jer se lijekovi zasnovani na proteinima gotovo uvijek moraju ubrizgati u tijelo [3].

## Antimikrobna svojstva

Antimikrobni peptidi (AMP) su peptidi koji imaju antimikrobna svojstva i time igraju ključnu ulogu u obrani organizama od mikrobnih infekcija. Često se nazivaju i kationskim peptidima jer su kao cjelina pozitivno nabijeni, iako su zapravo samo podvrsta kationskih peptida. Simboličan prikaz podjele kationskih peptida u skupine prema njihovim funkcijama, među kojima se nalaze i AMP prikazan je slikom 2.4. Oni su dio prirodnog imunološkog sustava prisutnog kod gotovo svih živih organizama, uključujući ljude, životinje, biljke i mikroorganizme. AMP se ističu svojom sposobnošću da selektivno ciljaju i ubiju mikrobe, uključujući bakterije, gljivice, parazite i viruse.

A circular diagram of different types of bacteria

Description automatically generated

*Slika 2.3. Podjela kationskih peptida prema funkcijama (preuzeto iz [6])*

AMP imaju raznoliku strukturu koja može uključivati linearno raspoređene aminokiseline, cikličke strukture, kao i kombinaciju različitih strukturnih motiva. Ovi peptidi često sadrže hidrofobne i hidrofilne regije, što im omogućuje interakciju s membranama mikroba. Mnogi AMP imaju pozitivno nabijene aminokiseline, kao što su lizin i arginin, što im pomaže u vezivanju za negativno nabijene komponente mikrobnih membrana [7].

AMP djeluju na mikrobe na različite načine, ciljajući njihove vitalne strukture i funkcije. Neki peptidi djeluju na mikrobne membrane, oštećujući njihovu strukturu i uzrokujući curenje unutarstaničnih tvari. Drugi peptidi prodiru u unutrašnjost mikroba i ciljaju vitalne procese kao što su sinteza proteina i DNA replikacija. Pojednostavljeni primjer djelovanja peptida na ciljnu molekulu prikazan je slikom 2.3. Također, neki peptidi imaju imunomodulatorna svojstva, potičući imunološki odgovor organizma i potencijalno regulirajući upalne procese.

Kada govorimo o razvoju lijekova, jedna od ključnih prednosti AMP je to što su mikroorganizmi manje skloni razvoju otpornosti na njihovo djelovanje za razliku od ostalih lijekova. Otpornost na antimikrobne lijekove postala je globalni problem, ali mnogi AMP djeluju na mikrobe putem mehanizama koji se razlikuju od konvencionalnih antibiotika, što smanjuje vjerojatnost razvoja otpornosti.

Diagram of a cell membrane

Description automatically generated with medium confidence

*Slika 2.4. Primjer djelovanja AMP na membranu ciljne molekule (preuzeto iz [8])*

## Antiviralna svojstva

Antiviralni peptidi (AVP) su podvrsta AMP koja ima gotovo sve iste karakteristike, ali su specifični po tome što imaju sposobnost suzbijanja virusnih infekcija. Kao i kod AMP, AVP koriste razne mehanizme kojima suzbijaju infekcije, npr. mogu blokirati vezanje virusa, spriječiti spajanje virusa sa stanicama domaćina, prekinuti proces signaliziranja virusa ili spriječiti razmnožavanje virusa u stanicama domaćina [1]. Osim toga, AVP mogu modulirati imunološki odgovor domaćina i poticati stvaranje protuvirusnih citokina.

Predviđanje antiviralnih svojstava prije sinteze novih peptida izuzetno je važno za razvoj lijekova i prevenciju virusnih infekcija. Mnoge metode strojnog učenja i računalnog modeliranja koriste se za predviđanje antiviralnih svojstava peptida na temelju njihovih fizičko-kemijskih svojstava i strukture. Ova područja istraživanja doprinose razvoju boljih modela za predviđanje antiviralne aktivnosti peptida i optimizaciju njihovih svojstava.

# AUTOENKODERI

## Autoenkoderi općenito

Autoenkoderi su posebna vrsta neuronskih mreža čiji je cilj što vjernije rekonstruirati podatke sa ulaza na izlazu. Sastoje se od tri glavna dijela: enkodera, latentnog prostora i dekodera, vrlo pojednostavljeno prikazano na slici 3.1. Možemo reći da autoenkoder uči kako kroz enkoder sažeti podatke u latentni prostor i prikazati ih pomoću manjeg broja značajki, a pritom minimizirati grešku rekonstrukcije na izlazu dekodera.

A network with blue and yellow circles

Description automatically generated

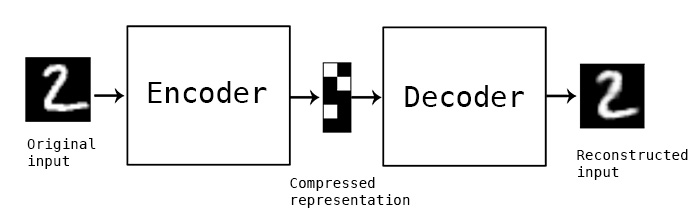
*Slika 3.1. Pojednostavljeni prikaz autoenkodera (enkoder – lijevi sloj, latentni prostor – srednji sloj, dekoder – desni sloj) (preuzeto iz [https://towardsdatascience.com/a-high-level-guide-to-autoencoders-b103ccd45924 LITR])*

Međutim, autoenkoder ne bi imao svrhu kada bi samo preslikavao podatke s ulaza na izlaz. Ono što ga čini korisnim je to da kroz proces rekonstrukcije podataka model uči koje su najvažnije značajke ulaznih podataka te kada je istreniran može se koristiti kako bi na temelju naučenog stvorio nove podatke uz dodatak ili izuzetak određenih značajki[https://www.unite.ai/what-is-an-autoencoder/ LITR]. Takav primjer možemo vidjeti i na slici 3.2. gdje autoenkoder uzima sliku znamenke napisane rukom, sažima ju u mali broj dimenzija, uči koje su najvažnije značajke te na kraju pomoću tih značajki rekonstruira sliku koja je vrlo slična originalnoj slici na ulazu.

Svaki dio autoenkodera ima svoju konkretnu ulogu u cijelome procesu. Enkoder je gusto povezana mreža i ulaz u autoenkoder. Zadaća mu je da uzme ulazne podatke te ih sažme i prikaže u latentnom prostoru, čineći tako novi prikaz podataka koji ima smanjenu dimenzionalnost.

Latentni prostor radi sa sažetim prikazom podataka. Konstruiran je tako da odredi najvažnije dijelove promatranih podataka odnosno značajke podataka koje su najvažnije za dobru rekonstrukciju. Cilj mu je odrediti koje značajke podataka trebaju biti očuvane, a koje se mogu ukloniti. Također treba uzeti u obzir i napraviti dobar omjer između veličine latentnog prikaza odnosno koliko će on biti sažet te relevantnosti značajki.

Dekoder je zadužen da uzme sažete podatke iz latentnog prostora koje je kreirao enkoder i pretvori ih ponovno u prikaz sa istim brojem dimenzija kao što su imali i početni ulazni podaci.



*Slika 3.2. Slika prolazi kroz enkoder, postaje sažeti prikaz podataka te na izlazu postaje rekonstruirana ulazna slika (preuzeto iz [https://blog.keras.io/building-autoencoders-in-keras.html LITR]*

Postoje tri vrlo bitne značajke koje definiraju rad autoenkodera [https://blog.keras.io/building-autoencoders-in-keras.html LITR]:

1. Specifični su za podatke na kojima uče – bit će u mogućnosti raditi samo s podacima koji su vrlo slični onima na kojima je model treniran, npr. autoenkoder koji je treniran na slikama na kojima se nalaze ljudska lica će raditi vrlo loše sa slikama na kojima se nalazi drveće jer su značajke koje je autoenkoder naučio specifične za ljudska lica
2. Imaju gubitke – podaci na izlazu koji su prošli kroz dekompresiju će biti lošije kvalitete u usporedbi s ulaznim podacima
3. Uče automatski iz primjera podataka – vrlo je lako trenirati algoritam da radi dobro na specifičnom tipu ulaznih podataka, odnosno nije potrebno mijenjati algoritam već je dovoljno algoritmu dati prikladne podatke za trening

Autoenkoderi imaju vrlo široki spektar primjene, ali najčešće su sljedeće[https://www.unite.ai/what-is-an-autoencoder/ LITR]:

* Smanjenje dimenzionalnosti - prolaskom kroz enkoder podaci dolaze u latentni prostor gdje su prikazani pomoću manjeg broja dimenzija u odnosu na originalne ulazne podatke te se tako mogu lakše razumjeti ili vizualizirati; također može pomoći mrežama visokog kapaciteta naučiti korisne značajke slika što upućuje na to da autoenkoderi mogu biti korišteni za poboljšanje treninga drugih tipova neuronskih mreža
* Uklanjanje šuma iz podataka - često se koristi na slikama; osim uklanjanja šuma, autoenkoderi mogu ispraviti i druge tipove oštećenja na slikama kao što su mutne slike ili dopuniti dijelove slike koji nedostaju
* Ekstrakcija značajki - kao i smanjenje dimenzionalnosti može biti korisno za poboljšanje treninga drugih tipova neuronskih mreža jer se mogu koristiti za prepoznavanje značajki skupova podataka koji se koriste za treniranje drugih modela
* Generiranje slika - autoenkoderi mogu generirati nove, do sad nepostojeće, slike ljudi ili animiranih likova što pomaže pri konstruiranju sustava za prepoznavanje lica ili automatiziranja određenih aspekata animacije
* Predviđanje sekvence - mogu se koristiti za određivanje promjene strukture podataka kroz vrijeme, što znači da mogu generirati idući podatak u sekvenci i to dovodi do toga da autoenkoderi mogu generirati videozapise
* Stvaranje sustava preporuke – koriste se duboki autoenkoderi, a sustave stvaraju tako što će prepoznati uzorke korištenja i povezati ih s interesima korisnika; u ovoj primjeni enkoder analizira podatke o načinu interakcije, a dekoder stvara preporuke koje odgovaraju prepoznatim uzorcima

## Varijacijski autoenkoderi

# MODEL

## Podaci

## Alati (keras, tensorflow, jn, python, knjižnice)

## Metode (sami model)

# REZULTATI

# ZAKLJUČAK

Peptidi s antiviralnim svojstvima sve više obećavaju u liječenju bolesti izazvanih virusima, a da bi se mogli koristiti i primijeniti treba ih prvo otkriti. Kako ne bi morali testirati svaki peptid za svaki virus, što je vremenski zahtjevno, skupo i neefikasno, tu je vrlo obećavajuće postalo predviđanje antiviralne aktivnosti peptida prema njihovim fizikalno-kemijskim svojstvima.

Takvo predviđanje može biti vrlo učinkovito, ako se dobro izvede. Tako je i ovim jednostavnim modelom strojnog učenja postignuta točnost od preko 80% što je vrlo dobro za neke jednostavne potrebe. Ugađanjem nekih parametara i uključivanjem nekih dodatnih svojstava prema kojima bi se mogla predviđati antiviralna aktivnost, npr. sekundarnih struktura, koje su bile zamišljene kao dio ovog modela, ali zbog određenih problema na kraju nisu uključene u model, mogla bi se i povećati točnost predviđanja i tako bi ono moglo imati i širu primjenu i to čak za neka veća i važnija znanstvena istraživanja. Predviđanje antiviralne aktivnosti peptida predstavlja jedan težak, kompleksan, ali i zanimljiv zadatak te kada bi se shvatio i iskoristio potencijal koji nudi, to bi moglo vrlo pozitivno utjecati na budućnost i ostvariti velike korake u medicini, a samim time i ljudskim životima.

VAE designed to learn a meaningful latent representation of peptide sequences and compare peptide similarity based on both their original sequences and their representations in the latent space. It's a powerful approach for exploring structural relationships among peptides and can have applications in various fields such as bioinformatics, drug discovery, and molecular biology.

# LITERATURA

[1] Kuan Y. Chang, Je-Ruei Yang: Analysis and prediction of highly effective antiviral peptides based on random forests, s Interneta, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3734225/>, 7. kolovoza 2023.

[2] Vilas Boas LCP, Campos ML, Berlanda RLA, de Carvalho Neves N, Franco OL: Antiviral peptides as promising therapeutic drugs, s Interneta, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7079787/>, 8. kolovoza 2023.

[3] Peptides Guide, s Interneta, <https://peptidesguide.com>, 18. srpnja 2023.

[4] Peptide Sciences, s Interneta, <https://www.peptidesciences.com/information/peptides-vs-proteins/>, 18. srpnja 2023.

[5] Courtney Simons: Structure of amino acids and proteins, s Interneta, <https://cwsimons.com/structure-of-amino-acids-and-proteins/>, 5. srpnja 2020.

[6] Hitchhiking with Nature: Snake Venom Peptides to Fight Cancer and Superbugs - Scientific Figure on ResearchGate, s Interneta, <https://www.researchgate.net/figure/Principal-functions-of-cationic-peptides-ACP-anticancer-peptides-AMP-antimicrobial_fig1_340674862>, 7. kolovoza 2023.

[7] Sara Wilcox: Cationic peptides: A new hope, s Interneta, <https://www.scq.ubc.ca/cationic-peptides-a-new-hope/>, 8. kolovoza 2023.

[8] Protein Esterification: Design, Antibacterial and Safety Assessment - Scientific Figure on ResearchGate. Available from: <https://www.researchgate.net/figure/Fig-1-Model-for-the-mechanism-of-action-of-cationic-antimicrobial-peptides_fig1_282186504>, 7. kolovoza 2023.

----------

[1] Kuan Y. Chang, Je-Ruei Yang, s Interneta, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3734225/>, 25. lipnja 2020.

[2] Liana Costa Pereira Vilas Boas, Marcelo Lattarulo Campos, Rhayfa Lorrayne Araujo Berlanda, Natan de Carvalho Neves, Octávio Luiz Franco, s Interneta, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7079787/>, 26. lipnja 2020.

[3] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Peptide>, 25. lipnja 2020.

[4] World of Peptides, s Interneta, [world-of-peptides.com](http://www.world-of-peptides.com/what-is-the-difference-between-a-peptide-and-a-protein/), 5. srpnja 2020.

[5] Courtney Simons, s Interneta, <https://cwsimons.com/structure-of-amino-acids-and-proteins/>, 5. srpnja 2020.

[6] [Ali Adem Bahar](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bahar%20AA%5BAuthor%5D&cauthor=true&cauthor_uid=24287494),[Dacheng Ren](https://www.ncbi.nlm.nih.gov/pubmed/?term=Ren%20D%5BAuthor%5D&cauthor=true&cauthor_uid=24287494), s Interneta, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3873676/>, 26. lipnja 2020.

[7] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Protein_secondary_structure>, 27. lipnja 2020.

[8] ThoughtCo., s Interneta, <https://www.thoughtco.com/protein-structure-373563>, 5. srpnja 2020.

[9] Golifescience, s Interneta, <https://www.golifescience.com/secondary-structure-of-proteins/>, 27. lipnja 2020

[10] C.Mirabello, G.Pollastri, s Interneta, <http://distillf.ucd.ie/porterpaleale/>, 29. lipnja 2020.

[11] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Relative_accessible_surface_area>, 6. srpnja 2020.

[12] Claus AF Andersen & Burkhard Rost, s Interneta, <https://rostlab.org/papers/2003_rev_assign/paper_slow.html>, 6. srpnja 2020.

[13] Guy Yachdav, Peter Hoenigschmid, Laszlo Kajan, s Interneta, <https://github.com/Rostlab/reprof>, 29. lipnja 2020.

[14] RPBS Web portal, s Interneta, <https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::psipred>, 29. lipnja 2020.

[15] J Yang, R Yan, A Roy, D Xu, J Poisson, Y Zhang. The I-TASSER Suite: Protein structure and function prediction. Nature Methods, 12: 7-8 (2015),

A Roy, A Kucukural, Y Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5: 725-738 (2010),

Y Zhang. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, vol 9, 40 (2008), s Interneta, <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>, 29. lipnja 2020.

[16] Niklas Donges, s Interneta, <https://builtin.com/data-science/random-forest-algorithm>, 30. lipnja 2020.

[17] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Decision_tree>, 5. srpnja 2020.

[18] Tutorialspoint, s Interneta, <https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm>, 30. lipnja 2020.

[19] Tony Yiu, s Interneta, <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>, 6. srpnja 2020.

[20] Abilash R., s Interneta, <https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57>, 5. srpnja 2020.

[21] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Weka_(machine_learning)>, 30. lipnja 2020.

[22] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[23] Thakur N, Qureshi A, Kumar M (2012) [AVPpred: collection and prediction of highly effective antiviral peptides.](http://nar.oxfordjournals.org/content/40/W1/W199.long) Nucleic Acids Res. 2012 Jul;40(Web Server issue):W199-204. doi: 10.1093/nar/gks450.  
PMID: [22638580](http://www.ncbi.nlm.nih.gov/pubmed/22638580), s Interneta, <http://crdd.osdd.net/servers/avppred/>, 30. lipnja 2020.

[24] Modlamp dokumentacija, s Interneta, <https://modlamp.org/modlamp.html#module-modlamp.descriptors>, 6. srpnja 2020.

[25] Gufosowa - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/wiki/File:K-fold_cross_validation_EN.svg>, 1. srpnja 2020.

# POPIS KRATICA

AVP – antiviralni peptidi

AMP – antimikrobni peptidi

MSA – *multiple sequence alignment*

SVM – *support vector machines*

WEKA - *Waikato Environment for Knowledge Analysis*

# Sažetak

Virusi i dalje predstavljaju glavni uzrok bolesti kod ljudi usprkos masovnoj proizvodnji cjepiva i antivirusnih lijekova, a velik potencijal za nove pomake u tom području predstavljaju peptidi s antiviralnom aktivnošću. Kako bi efikasno otkrivali nove takve peptide potrebno je koristiti neku vrstu predviđanja, a pokazano je da je to moguće na osnovi njihovih fizikalno-kemijskih svojstava. Takvo predviđanje, ako se dobro izvede, može biti vrlo učinkovito. Posebno se istaknulo strojno učenje metodom *random forest*, kakvo je implementirano i u ovom radu. Daljnjim trudom i ulaganjem u ovakva istraživanja može se ostvariti velik napredak u medicini i poboljšati kvaliteta ljudskog života.

Ključne riječi: antiviralni peptidi, predviđanje sekundarne strukture, algoritam *random forest*, strojno učenje

# Abstract

Viruses are still the main cause of human diseases despite the mass production of vaccines and antiviral drugs. Antiviral peptides have a great potential to change things for the better in that particular field. For efficient discovery of such new peptides, some kind of prediction should be used and it was shown that this was possible based on their physicochemical properties. Such prediction, if done right, can be very effective. Machine learning using the random forest method was proven especially efficient, which was also implemented in this study. With further effort and investments in such research, a great progress in medicine can be achieved and the quality of human lives improved.

Keywords: antiviral peptides, secondary structure prediction, random forest algorithm, machine learning