

**NASTAVNO-NAUČNOM VEĆU MEDICINSKOG FAKULTETA
UNIVERZITETA U BEOGRADU**

Na sednici **Nastavno-naučnog veća** Medicinskog fakulteta u Beogradu, održanoj dana 05.07.2022. godine, broj 11/VII-3/2-MC, imenovana je komisija za ocenu završene doktorske disertacije pod naslovom:

„Grafički prikaz podataka u biomedicinskim istraživanjima: sistematski pregled literature i razvoj interaktivnog softverskog rešenja“

kandidata Marka Savića, zaposlenog na Institutu za medicinsku statistiku i informatiku Medicinskog fakulteta Univerziteta u Beogradu. Mentor je prof. dr Nataša Milić.

Komisija za ocenu završene doktorske disertacije imenovana je u sastavu:

1. Prof. dr Dejana Stanisavljević, Medicinski fakultet u Beogradu
2. Doc. dr Zoran Bukumirić, Medicinski fakultet u Beogradu
3. Prof. dr Vesna Garović, Mejo klinika, Ročester, SAD

Na osnovu analize priložene doktorske disertacije, komisija za ocenu završene doktorske disertacije jednoglasno podnosi Naučnom veću Medicinskog fakulteta sledeći

IZVEŠTAJ

A) Prikaz sadržaja doktorke disertacije

Doktorska disertacija Marka Savića napisana je na ukupno 91 strani i podeljena je na sledeća poglavlja: uvod, ciljevi rada, materijal i metode, rezultati, diskusija, zaključci, literatura i prilozi. U disertaciji se nalazi ukupno 8 tabela i 37 ilustracija. Doktorska disertacija sadrži sažetak na srpskom i engleskom jeziku, biografiju kandidata i podatke o komisiji.

U **uvodu** je istaknut značaj grafičkog prikazivanja podataka u naučnim publikacijama i važnost pravilnog izbora odgovarajućeg tipa grafikona u zavisnosti od tipa podataka, veličine uzorka, odnosa među podacima i dr. Ukazano je na problem upotrebe neadekvatnih grafičkih metoda koje nisu prilagođene tipu podataka i istaknuta je učestala upotreba stubičastih dijagrama za prikazivanje numeričkih kontinuiranih podataka. Stubičasti dijagrami nisu namenjeni prikazivanju kontinuiranih podataka, a ukoliko se koriste na ovaj način, ne omogućavaju čitaocima da kritički procene prikazane

podatke, obzirom da veliki broj različitih raspodela podataka može dovesti do istih stubičastih dijagrama.

Pored značaja adekvatnog vizuelnog prikazivanja, u radu je istaknuta i potreba da se originalni podaci istraživanja učine javno dostupnim. Nedostatak originalnih podataka i neadekvatna vizuelna prezentacija u publikovanim studijama značajno umanjuje transparentnost i ponovljivost rezultata istraživanja u oblasti biomedicinskih nauka. Zbog toga se poslednjih godina posebna pažnja posvećuje razvoju softverskih alata za interaktivnu vizuelizaciju primarnih istraživačkih podataka koji omogućavaju korisniku prilagođen način prikaza, pri čemu integracija baze originalnih podataka u interaktivne grafikone doprinosi transparentnosti. Interaktivni prikaz podataka prikupljenih u istraživanjima u oblasti biomedicinskih nauka može biti koristan grafički alat za promociju transparentnosti, ponovljivosti rezultata istraživanja i otvorenog pristupa nauci, a kreiranje interaktivnih alternativa statičkom načinu prikazivanja sumarnih mera istraživanja predloženo je kao efikasna strategija za poboljšanje vizuelizacije podataka u biomedicinskim istraživanjima.

Ciljevi rada su precizno definisani. Sastoje se od procene učestalosti primene neadekvatnih metoda grafičkog prikaza podataka u biomedicinskim istraživanjima, razvoja softverskog rešenja koje omogućava integraciju baze podataka sa interaktivnim načinom vizualizacije i implementacije softverskog rešenja u obliku javno dostupnog veb servisa.

U poglavlju **materijal i metode** je detaljno opisan način sprovođenja sistematskog pregleda literature, razvoja softverskog rešenja i njegove implementacije. Sistematski pregled literature sproveden je u skladu sa *PRISMA* protokolom, a obuhvatio je originalne naučne radove objavljene u 25% najbolje rangiranih časopisa iz oblasti cirkulatornog sistema. Detaljno su opisani strategija pretraživanja, kao i procesi skrininga publikacija i ekstrakcije podataka. Za strategiju pretraživanja dat je tačan izvorni kod pretrage sprovedene na *PubMed*-u. Skrining publikacija radila su dva nezavisna istraživača, a neslaganja su rešavana konsenzusom. Dat je tačan opis uslova za uključenje i isključenje radova iz studije kao i dijagram toka selekcije publikacija za uključenje u studiju, a nakon skrininga, u studiju je uključeno 206 originalnih naučnih radova. Ekstrakciju podataka radila su dva nezavisna istraživača, a neslaganja su rešavana konsenzusom. Opisane su korišćene metode deskriptivne statistike, računane pomoću softvera IBM SPSS Statistics Version 21.

Razvoj integrisanog softverskog rešenja za interaktivnu vizuelizaciju podataka u biomedicinskim istraživanjima opisan je detaljno, a podrazumeva definisanje memorijskih i fajl struktura za smeštanje, obradu i čuvanje istraživačke baze podataka; dizajn formi za unos i ažuriranje podataka; definisanje rutina za statističku analizu podataka; definisanje vrsta grafikona i pridruženih opcija koje treba da budu dostupne za interaktivnu vizuelizaciju, kao i dizajn formi za interakciju sa opcijama

prikazivanja grafikona. Softver je razvijen u obliku veb aplikacije korišćenjem informacionih tehnologija, programskih jezika i njihovih biblioteka. Programski jezici HTML, CSS i javaskript primenjeni su za razvoj korisničkog interfejsa, tj. za dizajn stranica sa formama za unos podataka i stranica za interakciju sa grafikonima. Za razvoj funkcionalnog dela softvera tj. skladištenje i analizu podataka, generisanje grafikona i preuzimanje rezultata korišćen je programski jezik PHP. Detaljno su opisane mere deskriptivne statistike koje se koriste u aplikaciji i date su precizne formule za njihovo izračunavanje. Deskriptivne mere uključuju aritmetičku sredinu, standardnu devijaciju, standardnu grešku, interval poverenja, koeficijent varijacije, mere iskošenosti i spljoštenosti raspodele, percentile, granice dijagrama kutije i procenu gustine raspodele verovatnoće.

U poglavlju **rezultati** detaljno su opisani i jasno predstavljeni svi dobijeni rezultati.

Diskusija je napisana jasno i pregledno, uz prikaz podataka drugih istraživanja sa uporednim pregledom dobijenih rezultata doktorske disertacije.

Zaključci sažeto prikazuju najvažnije nalaze koji su proistekli iz rezultata rada.

Korišćena **literatura** sadrži spisak od 282 reference.

B) Provera originalnosti doktorske disertacije

Na osnovu Pravilnika o postupku provere originalnosti doktorskih disertacija koje se brane na Univerzitetu u Beogradu i nalaza u izveštaju iz programa iThenticate kojim je izvršena provera originalnosti doktorske disertacije „Grafički prikaz podataka u biomedicinskim istraživanjima: sistematski pregled literature i razvoj interaktivnog softverskog rešenja”, autora Marka Savića, konstatujemo da utvrđeno podudaranje teksta iznosi 5%. Ovaj stepen podudarnosti posledica je citata, ličnih imena, bibliografskih podataka o korišćenoj literaturi, tzv. opštih mesta i podataka, kao i prethodno publikovanih rezultata doktorandovih istraživanja, koji su proistekli iz njegove disertacije, što je u skladu sa članom 9. Pravilnika.

C) Kratak opis postignutih rezultata

Prema kriterijumima za uključivanje, primenom strategije pretraživanja, identifikovano je 428 publikacija iz 13 časopisa uključenih u studiju. Nakon primene kriterijuma za isključenje (isključeno je 222 publikacije), u istraživanje je uključeno 206 originalnih naučnih radova.

Grafikon ili dijagram za prikaz podataka sadržalo je 180 (87.4%) od 206 pregledanih publikacija. Stubičasti dijagram kojim se prikazuju kontinuirani numerički podaci je korišćen u najvećem broju studija (86 od 180, 47.8%). Druge vrste dijagrama za prikazivanje kontinuiranih podataka koje daju uvid u distribuciju podataka su ređe korišćene (tačkasti dijagram: 21.7%, dijagram kutije: 15.6%, histogram ili prikaz gustine raspodele: 14.4%).

Od 180 pregledanih publikacija koje su sadržale grafičko prikazivanje podataka, 129 (71.7%) je koristilo grafikone za prikaz numeričkih kontinuiranih podataka. Posmatrajući ovaj podskup publikacija, stubičasti dijagram je korišćen za prikazivanje kontinuiranih podataka u 86 od 129 radova (66.7%), od čega je 42 rada (32.6%) za prikazivanje kontinuiranih podataka koristilo isključivo stubičasti dijagram (u preostala 44 rada, pored stubičastog, korišćene su i druge vrste grafičkog prikaza za numeričke kontinuirane podatke).

Prilikom kreiranja stubičastih dijagrama za prikazivanje numeričkih kontinuiranih podataka, linije greške su predstavljale standardnu grešku u 57 publikacija (66.3%), standardna devijacija u 18 (20.9%), 95% interval poverenja u dve publikacije (2.3%), a u 8 publikacija (9.3%) nije bilo moguće utvrditi šta predstavljaju linije greške na stubičastim dijagramima. Jedna publikacija (1.2%) je sadržala više stubičastih dijagrama sa različitim merama prikazanim linijama greške.

Medijana veličine najmanje grupe podataka predstavljene stubičastim dijagramom iznosila je 3 (IQR: 3 – 5), a najveće grupe 10 (IQR: 7 – 17). Medijana veličine najmanje grupe podataka predstavljene tačkastim dijagramom iznosila je 6 (IQR: 4 – 10), a najveće grupe 12 (IQR: 8 – 17). Medijana veličine najmanje grupe podataka predstavljene dijagramom kutije iznosila je 20 (IQR: 12 – 52), a najveće grupe 66 (IQR: 18 – 280).

Softver za interaktivnu vizuelizaciju podataka razvijen je u vidu veb aplikacije i sadrži ukupno 5 sekcija: početna strana sa osnovnim informacijama o aplikaciji, sekcija sa uputstvima za upotrebu, sekcija za unos i ažuriranje baze podataka, sekcija za interaktivno prikazivanje grafikona i stranica za preuzimanje rezultata rada aplikacije. Početna strana veb aplikacije pruža korisniku osnovne informacije o njenoj nameni i mogućnostima upotrebe. Na početnoj strani su u kratkim crtama predstavljene ostale sekcije aplikacije i postavljeni su hiperlinkovi do prvih koraka u njenom korišćenju kao što su stranice za definisanje ili dodavanje podataka, ili preuzimanje ilustrativnih primera baze podataka. Sekcija za unos i ažuriranje podataka (*Data*) sastoji se od tri podsekcije: definisanje skupa podataka (*Define*), otpremanje fajla sa pripremljenim podacima (*Upload*) i podsekcija za prikaz i ažuriranje podataka (*View/Edit*). Sekcija za interaktivno prikazivanje grafikona (*Dotplot*) postaje dostupna tek nakon što se u aplikaciju učita baza podataka, bilo otpremanjem fajla

ili direktnim unosom. Grafikoni koji se kreiraju i čuvaju tokom rada u aplikaciji, nakon čuvanja postaju dostupni kao podsekcije unutar navigacije.

Vrste grafikona koje mogu da se generišu pomoću aplikacije su: tačkasti dijagram (*Dotplot*), dijagram kutije (*Box plot*), violina dijagram (*Violin plot*) i stubičasti dijagram (*Bar graph*). U odeljku za izbor vrste grafikona se prikazuje preporuka da je za male uzorke (manje od deset jedinica posmatranja po grupi) poželjno koristiti tačkasti dijagram. Forma sa opcijama za interakciju sa grafikonom sadrži odeljke za izbor podataka i podešavanja njihovog prikaza, tj. za podešavanje varijabli, grupa, podgrupa i klastera.

U okviru glavne navigacije dostupna je stavka za uključivanje tj. isključivanje prilagođavanja aplikacije i generisanih grafikona osobama sa slepilom za boje (*Color blind mode*). Ovu opciju je moguće aktivirati u bilo kom trenutku korišćenja aplikacije, a njenom aktivacijom se svi vizuelni elementi koji bojama obeležavaju pripadnost klasteru ili grupi prikazuju sa izmenjenom paletom kako bi bili jasno uočljivi i osobama sa smanjenom sposobnošću raspoznavanja boja.

Tokom unosa i ažuriranja baze podataka, kao i tokom kreiranja i interakcije sa grafikonima, svi podaci i podešavanja čuvaju se u privremenoj memoriji aplikacije. Kako bi rezultati rada u aplikaciji mogli biti trajno sačuvani, u glavnoj navigaciji je dostupna stavka za preuzimanje podataka (*Download workset*). Aktiviranjem ove stavke, u aplikaciji se prikazuje dijalog sa opcijama za preuzimanje samo podataka ili integrisanih podataka sa grafikonima.

Brisanje podataka dostupno je iz glavne navigacije pod stavkom *Dismiss work*. Kada se aktivira ova opcija, korisniku se prikazuje dijalog u kome treba da potvrdi brisanje kako ne bi došlo do slučajnog gubitka rezultata prethodnog rada u aplikaciji.

Softversko rešenje za interaktivnu vizuelizaciju podataka u biomedicinskim istraživanjima implementirano je u vidu veb servisa koji je javno i besplatno dostupan na sajtu Instituta za medicinsku statistiku i informatiku Medicinskog fakulteta Univerziteta u Beogradu, na internet adresi: <http://statistika.mfub.bg.ac.rs/interactive-dotplot>.

Korisnicima je omogućeno da učitavaju, unose, pregledaju i ažuriraju podatke, pregledaju deskriptivne mere svojih podataka i da kreiraju i ažuriraju grafikone. Na kraju rada, korisnik može da preuzme fajl koji u sebi integriše bazu podataka i sačuvane grafikone, a koji može koristiti kao prilog publikaciji ili za ponovno učitavanje u aplikaciju radi dalje interakcije sa podacima i grafikonima. Nakon završetka rada, podaci korisničke sesije se brišu iz privremene memorije čime se onemogućava neautorizovani pristup istraživačkim podacima.

D) Usporedna analiza doktorske disertacije sa rezultatima iz literature

U doktorskoj disertaciji su adekvatno prodiskutovani dobijeni rezultati u odnosu na objavljenu literaturu. Neadekvatna upotreba stubičastih dijagrama za prikaz numeričkih kontinuiranih podataka bila je najčešći problem uočen u pregledanim radovima, što je u saglasnosti sa rezultatima drugih istraživanja dostupnih u literaturi (Cooper i sar., 2002, Weissgerber i sar., 2015.). Stubičasti dijagrami mogu doprineti pogrešnom zaključivanju jer stvaraju prividnu sigurnost u zaključke usmeravajući pažnju čitaoca na razlike u sumarnim statistikama, pogrešno sugerišući da su sve prosečne vrednosti podjednako značajne. Ova pretpostavka je naročito problematična kod prikazivanja sumarnih statistika malih grupa koje mogu značajno odstupati od stvarnih populacionih parametara. Pretklinička i bazična biomedicinska istraživanja najčešće imaju manje od 15 jedinica posmatranja po grupi, dok su veličine najmanjih grupa najčešće između 3 i 5 jedinica posmatranja prema rezultatima ove studije, što je u saglasnosti sa nalazima prezentovanim u drugim radovima (Weissgerber i sar., 2015, Holman i sar., 2016).

Prilikom neadekvatnog korišćenja stubičastih dijagrama za prikazivanje numeričkih kontinuiranih podataka ustaljena je praksa da se stubićima dodaju linije greške koje mogu predstavljati neku od sumarnih statistika koje imaju za cilj da daju uvid u varijabilitet podataka ili preciznost merenja. Ovim radom pokazano je da u oko dve trećine slučajeva (66.3%) linije greške predstavljaju standardnu grešku pre nego standardnu devijaciju (20.9%) ili 95% interval poverenja (2.3%). Ovo može biti problematično jer standardna greška predstavlja preciznost procene aritmetičke sredine a ne varijabilitet podataka (Davies 2016). Obzirom da vrednost standardne greške zavisi od veličine uzorka, ukoliko među prikazivanim grupama postoje razlike u veličini, može se stvoriti lažan osećaj (ne)sigurnosti u prikazane sumarne statistike. Takođe, ukoliko su veličine grupa veoma različite, može se izbeći stvaranje lažnog osećaja sigurnosti tako što bi se tačkastim dijagramima pridružili dijagrami greške, kutije ili violine samo kod većih grupa.

Interaktivno softversko rešenje za vizuelizaciju podataka u biomedicinskim istraživanjima razvijeno i predstavljeno u doktorskoj disertaciji omogućava istraživačima bez programerskih veština da kreiraju grafikone prilagođene empirijskim podacima, uključujući tačkaste dijagrame, dijagrame kutije i violina dijagrame. Koristeći ovo softversko rešenje, istraživači mogu da na interaktivan način sa lakoćom porede vrednosti numeričkih kontinuiranih varijabli među istraživačkim grupama pregledom različitih dijagrama koji adekvatno prikazuju raspodelu podataka. Dijagramima mogu biti pridružene sumarne statistike ili tačke individualnih podataka, uz dodatnu mogućnost isticanja tačaka ili sumarnih statistika u zavisnosti od veličine grupa ili drugih potreba istraživača. Funkcionalnosti interaktivnog softverskog rešenja uključuju i vizuelizaciju podgrupa ili tzv. klastera podataka (podaci koji nisu u potpunosti međusobno nezavisni). Prikazivanje podgrupa u ilustracijama svakako može

biti korisno u mnogim situacijama, a ova funkcionalnost posebno dobija na značaju nakon inicijative Nacionalnih instituta zdravlja SAD (eng. *NIH: National Institutes of Health*) da se pol u svim istraživanjima finansiranim od strane ove organizacije tretira kao biološka varijabla.

Klasteri uparenih podataka, kao što su laboratorijske životinje iz istog legla, su veoma učestali u bazičnim istraživanjima, međutim, pokazano je da se ovakvi skupovi podataka često analiziraju neadekvatnim metodama (Lazic 2010). Informacije o odsustvu nezavisnosti podataka i vrsti uparenosti (klastera) su retko prikazane na statičkim grafikonima, dok se pomoću interaktivnog alata razvijenog i opisanog u doktorskoj disertaciji one mogu sa lakoćom vizuelizovati i razmatrati. Ove informacije mogu biti od naročitog značaja kada je potrebno odrediti koja od statističkih metoda je adekvatna za analizu datih podataka (Galbraith i sar., 2010). Prilikom interakcije sa grafikonima, istraživači mogu koristiti i opciju redukcije podataka (data reduction) koja prikazuje samo aritmetičku sredinu ili medijanu svakog pojedinačnog klastera. Ova jednostavna tehnika je često korišćena prilikom analize podataka u klasterima kod studija sa malim veličinama uzorka.

Preporuke proistekle iz ovog istraživanja opisane su detaljno, a uključuju preporuke za otklanjanje uobičajenih problema u vizuelizaciji podataka, zatim za pravilno strukturisanje dijagrama i njihovih delova za prikazivanje rezultata uobičajenih statističkih analiza koje se koriste u biomedicinskim istraživanjima, kao i preporuke za najbolju praksu prilikom korišćenja tipova grafikona dostupnih u razvijenom softveru.

E) Objavljeni radovi koji čine deo doktorske disertacije

Tracey L. Weissgerber, **Marko Savic** (autori dele prvo mesto), Stacey J. Winham, Dejana Stanisavljevic, Vesna D. Garovic, Natasa M. Milic. **Data visualization, bar naked: A free tool for creating interactive graphics.** J Biol Chem. 2017;292:20592–8. M21 IF: 4.01

F) Zaključak (obrazloženje naučnog doprinosa)

Doktorska disertacija „Grafički prikaz podataka u biomedicinskim istraživanjima: sistematski pregled literature i razvoj interaktivnog softverskog rešenja“ Marka Savića predstavlja originalni naučni doprinos u sagledavanju učestalosti primene neadekvatnih metoda grafičkog prikaza podataka u biomedicinskim istraživanjima, kao i u obezbeđivanju naprednog softverskog rešenja za interaktivnu vizuelizaciju podataka prilagođenom korisniku. Razvojem softvera dobijen je vredan

istraživački alat koji podržava otvorenost u nauci čineći originalne podatke integrisanim delom ilustracija naučnih istraživanja.

Ova doktorska disertacija je urađena prema svim principima naučnog istraživanja. Ciljevi su bili precizno definisani, naučni pristup je bio originalan i pažljivo izabran, a metodologija rada je bila savremena. Rezultati su pregledno i sistematično prikazani i diskutovani, a iz njih su izvedeni odgovarajući zaključci.

Na osnovu svega navedenog, i imajući u vidu dosadašnji naučni rad kandidata, komisija predlaže Naučnom veću Medicinskog fakulteta Univerziteta u Beogradu da prihvati doktorsku disertaciju Marka Savića i odobri javnu odbranu radi sticanja akademske titule doktora medicinskih nauka.

U Beogradu, 11.07.2022.

Članovi Komisije:

Prof. dr Dejana Stanisavljević

Doc. dr Zoran Bukumirić

Prof. dr Vesna Garović

Mentor:

Prof. dr Nataša Milić
