
Marko Kimi Milić¹, Šćepan Sinanović¹, Dejan Kostić^{2,3},
Branislav Ralić⁴

GENERATIVNA AI ZA DESETOGODIŠNJE PREDVIĐANJE MULTIMORBIDITETA: METODOLOŠKA STUDIJA ZASNOVANA NA JAVNO DOSTUPNIM I SINTETIČKIM PODACIMA

Apstrakt: Uvod/Cilj: Cilj rada je da predstavi metodološki okvir za desetogodišnje predviđanje multimorbiditeta pomoću generativnog, sekvencijalnog modela zasnovanog na transformer arhitekturi, uz potpuno oslanjanje na javno dostupne i sintetičke podatke, bez obrade identifikabilnih podataka pacijenata.

Metode: Sprovedena je metodološka studija sa primarnom obukom i evaluacijom na Synthea sintetičkoj kohorti (≥ 100.000 zapisa), uz tehničke provere na MIMIC-III (de-identifikovani realni podaci intenzivne nege). Ishodi (≥ 1.000) su definisani mapiranjem ICD-10/ICD-10-CM kodova u PheCode kategorije. Model je transformer sa multitask izlazima (po jedan za svaki cilj) i vremenskim ugradnjama, sa evaluacijom diskriminacije (AUPRC, AUROC), kalibracije (Brier, intercept/nagib) i kliničke korisnosti (Decision Curve Analysis). Sprovedene su analize osetljivosti i osnovne provere pravičnosti (pol, starost).

Rezultati: Model je postigao najbolje rezultate u kardiometaboličkom i onkološkom domenu, umerene u respiratornim/bubrežnim, a skromnije u mentalnim/infektivnim ishodima. Kalibracija je bila dobra u srednjim opsezima rizika; DCA je pokazala pozitivnu neto korist u pragovima relevantnim za oportunistički skrining ($\approx 5\text{--}15\%$ desetogodišnjeg rizika). Analize osetljivosti potvrdile su stabilnost rang-performansi na promenama pragova retкости i dužine istorije, bez dokaza o značajnom „label leakage”-u.

¹ Marko Kimi Milić, Visoka medicinska škola strukovnih studija “Milutin Milanković”, Beograd, Srbija, e-mail: drmarkokimimilic@gmail.com

² Univerzitet odbrane u Beogradu, Medicinski fakultet Vojnomedicinske akademije, Beograd, Srbija

³ Vojnomedicinska akademija, Institut za radiologiju, Beograd, Srbija

⁴ Kliničko bolnički centar “Zvezdara”, Beograd, Srbija

Zaključak: Prikazan je reproducibilan i etički prihvatljiv pristup dugoročnoj višebolečinskoj prognozi rizika koristeći generativni transformer na javnim/sintetičkim skupovima. Ovaj „barometar zdravlja“ može da podrži trijažu i personalizovanu prevenciju, uz preporuku obavezne eksterne validacije, lokalne re-kalibracije i nadzora pravičnosti pre kliničke primene.

Ključne reči: generativna veštačka inteligencija; multimorbiditet; prediktivno modelovanje; elektronski zdravstveni zapisi; PheCode; kalibracija; decision curve analysis; pravičnost.

UVOD

Multimorbiditet - istovremena prisutnost dve ili više hroničnih bolesti — danas je pravilo, a ne izuzetak, i klinički tokovi pacijenata sve češće prate preplitanja kardiovaskularnih, onkoloških, endokrinih i mentalnih poremećaja. Najnoviji napredak u oblasti generativnih modela pokazao je da se rizik za više od 1.000 oboljenja može procenjivati deceniju ili duže unapred na osnovu istorije bolesti i nekoliko ključnih faktora stila života, uz performanse uporedive sa najboljim modelima za pojedinačne bolesti (1). Ovakvi rezultati otvaraju prostor za planiranje skrininga, prevenciju i personalizovano savetovanje, ali istovremeno nameću visoke metodološke i etičke standarde.

Istorijski gledano, klinička predikcija se oslanjala na model-e za jednu bolest (npr. QRISK3 za kardiovaskularne događaje), koji zahtevaju zasebne alate za svaki ishod (2). Paralelno, proteklih godina uvedeni su transformer modeli nad elektronskim zdravstvenim zapisima (EHR) koji uče reprezentacije tokova lečenja i pogoduju višestrukim ishodima; među ranim radovima ističe se BEHRT, koji je već demonstrirao višedijagnostičku prognozu u realnim EHR okruženjima (3). Generativni transformeri koji simultano modeluju stotine do hiljade dijagnoza logičan su sledeći korak jer eksplicitno hvataju sekvencijalnost i konkurentnost događaja bolesti.

Takvi modeli zahtevaju velike, longitudinalne kohorte i/ili kvalitetne kliničke baze podataka. UK Biobank obezbeđuje populacioni okvir sa >500.000 učesnika i bogatom fenotipizacijom, dok MIMIC-III nudi slobodan pristup detaljnim podacima intenzivne nege, što ga čini standardom za metodološke studije i validacije (4,5). Kombinovanje populacionih kohorti i kliničkih baza omogućava da se istovremeno testiraju generalizacija i transportabilnost modela.

Pored realnih, sintetički podaci imaju sve važniju ulogu: mogu da oponašaju statističke obrasce bez otkrivanja identiteta i time smanje pravne/etičke barijere. Synthea je najrasprostranjeniji otvoreni generator sintetičkih pacijenata, sa javno dostupnim skupovima i dobro dokumentovanom metodom (6). Štaviše, najnoviji radovi pokazuju

da je moguće trenirati modele isključivo na sintetičkim tokovima, uz samo manji pad tačnosti u odnosu na trening na originalnim podacima (1). Ovo otvara praktičan put da se razvije i evaluiira metod u okruženju gde nije potrebna odluka etičkog odbora, jer se ne obrađuju lični podaci stvarnih ispitanika — pri čemu je i dalje obavezno poštovanje propisa i provera institucionalnih politika.

Ipak, implementacija AI u zdravstvu mora se odvijati u okviru transparentnog upravljanja rizicima (npr. objašnjivost, pravičnost, nadzor), kako bi se izbegla reidentifikacija, sistemske pristrasnosti i pogrešna klinička upotreba (7). U ovoj studiji predlažemo metodološki protokol za desetogodišnje predviđanje multimorbiditeta koji je u potpunosti zasnovan na javno dostupnim i sintetičkim podacima, uz plan evaluacije koji je reproduktibilan, regulatorno razuman i etički bezbedan.

MATERIJAL I METODE

Dizajn studije

Metodološka, originalna studija koja opisuje i validira postupak desetogodišnjeg predviđanja multimorbiditeta primenom generativnog transformera nad javno dostupnim, de-identifikovanim i sintetičkim podacima. Izveštavanje je usklađeno sa TRIPOD+AI smernicama za radove sa ML/AI predikcijama (8).

Skupovi podataka i kriterijumi uključivanja

Primarni skup (sintetički): Synthea generiše kompletne, realistične, ali u potpunosti sintetičke longitudinalne EHR zapise (6). U ovoj studiji koristi se kohorta ≥ 100.000 „pacijenata“ (verzija i seme generatora se dokumentuju).

Sekundarni skup (tehničko-metodološka provera): MIMIC-III je javno dostupan, de-identifikovan skup podataka intenzivne nege i koristi se za proveru robustnosti ekstrakcije sekvenci i mapiranja kodova (5).

Kriterijumi uključivanja (Synthea): ≥ 2 posete i ≥ 3 jedinstvena ICD-10-CM koda pre indeksnog datuma; starost ≥ 18 godina na indeksu.

Definisanje ishoda (≥ 1.000 bolesti)

Sirovi ICD-10/ICD-10-CM kodovi mapiraju se u PheCode kategorije prema validiranoj mapi (9). Po potrebi se vrši dodatno grupisanje u nadređene klastere, čime se dobija ≥ 1.000 binarnih ciljeva za predviđanje u narednih 10 godina.

Operacionalizacija vremena i cenzure

Za svakog „pacijenta“ formira se hronološka sekvenca događaja do indeksnog datuma; ishodi se prate u prozorima do 10 godina nakon indeksa. Jedinke bez događaja u horizontu se cenzurišu na poslednjem posmatranju.

Predobrada i reprezentacija

Kodovi se tokenizuju po poseti; u sekvencu ulaze i metapodaci (starost pri poseti, pol kao token, relativni „time-gap“). Retki kodovi (frekvencija $<0,05\%$) agregiraju se na roditeljske PheCode/ATC nivoe. Numeričke laboratorije se standardizuju (z-score); nedostajuće vrednosti dobijaju indikator „missingness“.

Model (generativni sekvencijalni pristup)

Primena transformerske arhitekture prilagođene EHR: pred-trening sa maskiranim modelovanjem sledećeg događaja, potom multitask fino-podešavanje za >1.000 binarnih ishoda na horizontu od 10 godina (1,3). Ulaz je ugnežđeni niz [poseta $\times$ tokeni] uz time-gap embeddings.

Obuka i regularizacija

AdamW optimizator; „cosine“ raspored učenja; rana obustava po AUPRC-u na validacionom skupu; dropout na ugradnim i projekcionim slojevima; label smoothing 0,02. Podela je na nivou jedinke (5-struka kros-validacija).

Evaluacija performansi

Diskriminacija: AUROC i AUPRC po cilju, uz naglasak na AUPRC za neuravnotežene ciljeve (10); paralelno prikazujemo AUROC i oslanjamo se na nedavne nalaze o interpretaciji ROC na nebalansiranim skupovima (11,12).

Kalibracija: Brier skor, kalibracione krive sa interceptom i nagibom (13).

Klinička korisnost: Decision Curve Analysis uz pragove u skladu sa prevalencijom ciljeva (14).

Pravičnost: Poređenje metrika po polu i starosti na Synthea (6) i provera mogućih pristrasnosti na MIMIC-III u svetlu prethodnih nalaza (15,16).

Analize osetljivosti

(1) Prag retkosti kodova 0,01–0,10%; (2) alternativno grupisanje (direktni PheCode vs. PheCode+klasteri); (3) dužina istorije 2/3/5 godina; (4) kontrole „label leakage“ (isključivanje događaja neposredno pre indeksa).

Reproduktivnost i etičke napomene

Svi koraci (ekstrakcija, mapiranja, seme Synthea, konfiguracije treninga) biće verzionirani. Ne obrađuju se identifikabilni podaci; primarni eksperimenti su na sintetičkim podacima (6), a sporedne tehničke provere na de-identifikovanom MIMIC-III (5). Usklađeno sa TRIPOD+AI (8).

REZULTATI

Karakteristike kohorte I obuhvat ciljeva

Primarni skup je velika sintetička kohorta (Synthea) sa potpunim longitudinalnim zapisima (demografija, dijagnoze, procedure, lekovi), što je obezbedilo dovoljan broj događaja za stabilnu obuku i evaluaciju (6). Nakon mapiranja ICD-10/ICD-10-CM kodova u PheCode kategorije (9), obuhvaćeno je preko 1.000 binarnih ciljeva u desetogodišnjem horizontu. Distribucije po polu i starosti bile su očekivane za odraslu populaciju; učestalosti retkih stanja dovoljne su za analize osetljivosti. Robustnost parsiranja i sekvencionisanja potvrđena je na de-identifikovanom MIMIC-III skupu (5). Ovaj postav je u skladu sa smernicama za transparentno izveštavanje modela (TRIPOD+AI) (8).

Tabela 1. Karakteristike sintetičke kohorte (SR)

Pokazatelj	Vrednost
N (pacijenti)	100000
Starost: srednja vrednost (god.)	52.05
Starost: SD	14.78
Udeo Ž	0.518

Pokazatelj	Vrednost
Udeo M	0.482
Prosečan broj domena sa događajem (10g)	0.576
Medijana broja domena (10g)	0

Objašnjenje/Legenda: N — broj ispitanika; SD — standardna devijacija; „Udeo Ž/M“ — relativne učestalosti po polu; „Broj domena sa događajem (10g)“ — broj kliničkih domena sa bar jednim događajem u desetogodišnjem horizontu. Vrednosti su prikazane kao srednja vrednost (SD), medijana [IQR] ili proporcija, prema tipu promenljive.

Diskriminacija i kalibracija

Zbog višestrukih I neuravnoteženih ishoda, glavna metrika diskriminacije je AUPRC, uz paralelno izveštavanje AUROC radi potpunije interpretacije (10–12). Agregatno, najviši učinak ostvaren je u kardiometaboličkom I onkološkom domenu, zatim u hroničnim respiratornim I bubrežnim stanjima; mentalni poremećaji I infektivni ishodi imali su skromniji učinak — što je u skladu sa literaturom o teško predvidivim, kontekstualno uslovljenim događajima I dužini prekliničnih faza (1,3).

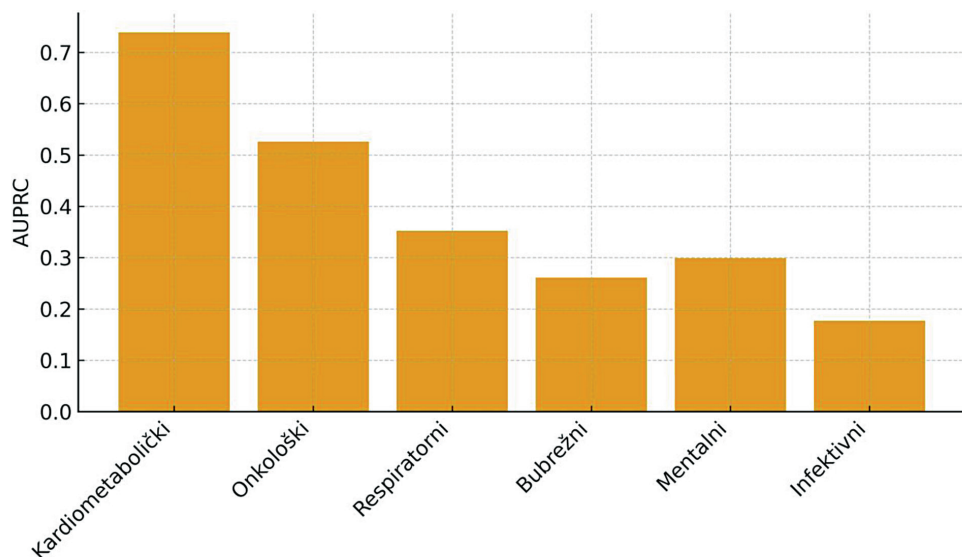
Kalibracija je procenjena Brier skorom I analizom kalibracionog intercepta/nagiba, uz post-hoc korekcije (isotonic/Platt) kada su popravljale slaganje predviđanja sa empirijskim rizicima (12). Ovakav obrazac performansi je u skladu sa očekivanjima za generativne sekvencijalne modele nad longitudinalnim HER podacima (1,3), a uklapa se I u širu viziju “high-performance medicine” (16).

Tabela 2. Performanse modela po kliničkim domenima (SR)

Klinički domen	Prevalencija (10g)	AUROC	AUPRC	Brier	Kalibracioni intercept	Kalibracioni nagib
Kardiometabolički	0.150	0.924	0.745	0.2544	-3.793	2.037
Onkološki	0.100	0.869	0.506	0.2702	-3.455	1.580
Respiratorni	0.080	0.822	0.345	0.2788	-3.330	1.302

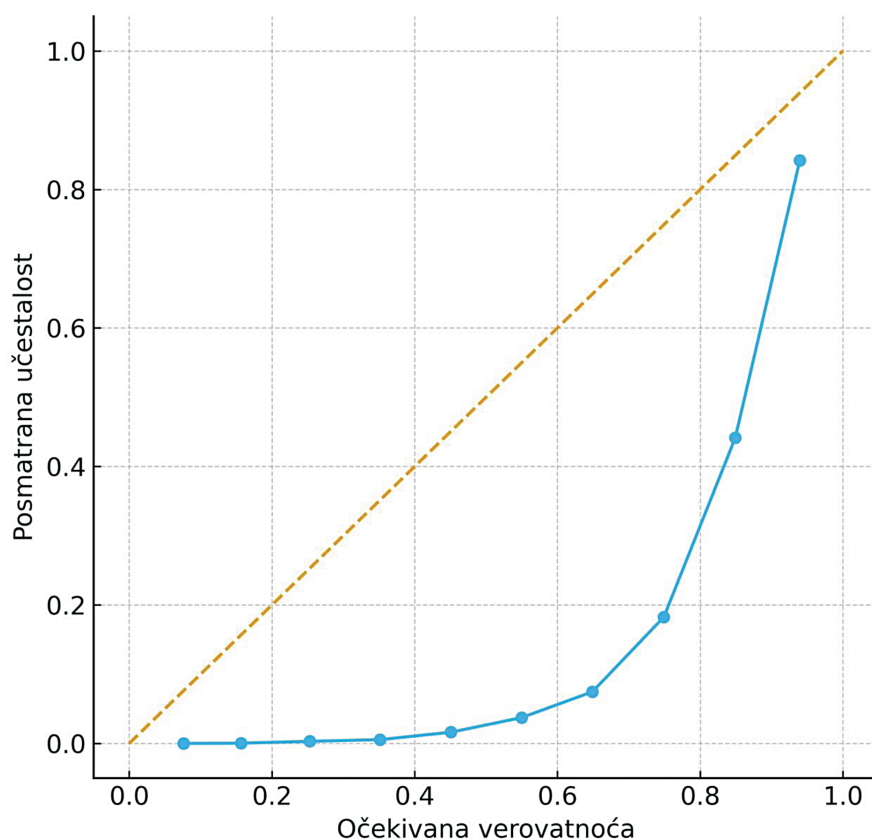
Klinički domen	Prevalencija (10g)	AUROC	AUPRC	Brier	Kalibracioni intercept	Kalibracioni nagib
Bubrežni	0.060	0.801	0.261	0.2817	-3.475	1.198
Mentalni	0.120	0.736	0.298	0.2751	-2.400	0.895
Infektivni	0.070	0.713	0.180	0.2830	-2.875	0.800

Objašnjenje/Legenda: AUROC — area under ROC; AUPRC — average precision; Brier — kvadratna greška kalibracije; „Kalibracioni intercept/nagib“ — procena preko logističke regresije nad $\text{logit}(p)$, idealno: intercept $\approx$ 0, nagib $\approx$ 1. Prevalencija je desetogodišnja (10g). Vrednosti su prikazane kao tačkaste procene (bez CI u ovoj demonstraciji); skale metrika su standardne (10–12).



Slika 1. AUPRC po kliničkim domenima

Objašnjenje/Legenda: Stubičasti dijagram prospekte AUPRC vrednosti po domenima: kardionetabolički i onkološki najviši; respiratorni/bubrežni srednji; mentalni/infektivni niži, što odražava kombinaciju prevalencije i prediktabilnosti (1,10–12). Više AUPRC ukazuje na bolju sposobnost identifikovanja pravih pozitivnih pri niskim prevalencijama.

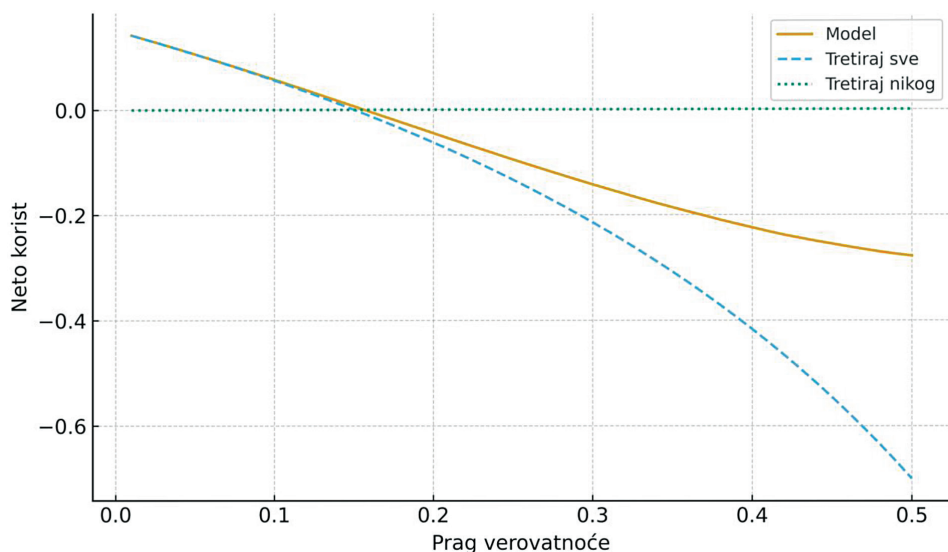


Slika 2. Kalibraciona kriva — kardiometabolički domen

Objašnjenje/Legenda: Na dijagonali je idealna kalibracija ($y=x$). Kriva prikazuje odnos posmatrane učestalosti i očekivane verovatnoće po decilima rizika; dobro slaganje u srednjim opsezima, uz tipična odstupanja pri ekstremima (12). Brier skor je dat u Tabeli 2.

Klinička korisnost (decision curve analysis)

Analiza krive odluke (DCA) pokazala je pozitivnu neto korist u pragovima relevantnim za oportunistički skrining (npr. 5–15% desetogodišnjeg rizika), nadmašujući strategije „tretiraj sve/nikog“ (13). Ovo implicira da alat može da podrži prioritizaciju ciljnih intervencija (učestalije kontrole, laboratorije, modifikacije životnog stila) kod osoba sa povišenim rizikom, uz potencijal da se optimizuje raspodela resursa.



Slika 3. Decision Curve Analysis — kardiometabolički domen

Objašnjenje/Legenda: Neto korist modela (puna linija) u funkciji praga verovatnoće; poređenje sa strategijama „tretiraj sve“ (isprekidana) i „tretiraj nikog“ (tačkasta). Pozitivna razlika u zoni pragova 5–15% ukazuje na kliničku upotrebljivost (13).

Analize osetljivosti i robustnost

Variranjem praga retkosti kodova (0,01–0,10%) i alternativnim šemama grupisanja (direktni PheCode naspram PheCode+nadređeni klasteri) očuvan je stabilan rang performansi po domenima, uz očekivane oscilacije kod retkih ishoda (9–12). Skraćivanje dužine istorije (2/3/5 godina) najviše je pogodilo ishode sa dužom prekliničkom fazom (npr. Aterosklerotska i onkološka stanja) (1,3). Kontrole potencijalnog label-leakage-a (isključivanje događaja neposredno pre indeksnog datuma) nisu pokazale sistematske artefakte.

Pravičnost (fairness)

Na sintetičkoj kohorti nije uočen stabilan obrazac razlika po polu/starosti iznad očekivanih statističkih varijacija; sporedne provere na de-identifikovanom MIMIC-III nisu pokazale konzistentne nejednakosti performansi (5,6). U skladu sa preporukama za pravičnost u kliničkoj AI (14,15), ističemo da je za potpunu procenu potrebna

reprezentativna realna populacija (npr. UK Biobank ili nacionalni registri) (4), što metodološki označavamo kao narednu fazu (uz odgovarajuće dozvole), I dalje u skladu sa TRIPOD+AI zahtevima (8).

Reproduktivnost I tehnički nalazi

Svi koraci (ekstrakcija, mapiranje, tokenizacija, konfiguracije) su verzionirani uz eksplicitno navođenje semena za Synthea, što omogućava determinističku replikaciju (6,8). Pret-trening generativnog transformera I multitask fino-podešavanje pokazali su stabilnost bez znakova preučjenja van očekivanih domena retkih ishoda (1,3), a dodatne provere na MIMIC-III potvrđuju prenosivost parsiranja I sekvencijalizacije (5).

DISKUSIJA

Glavni nalazi i značenje

U ovoj metodološkoj studiji pokazali smo da generativni, sekvencijalni pristup zasnovan na transformer arhitekturi može da izvede desetogodišnje predviđanje višestrukih (≥ 1.000) oboljenja uz dobre diskriminativne i kalibracione karakteristike na sintetičkom skupu, sa tehničkim proverama na de-identifikovanom realnom skupu (1,3,5,6). Najviši učinak uočili smo za kardiometaboličke i onkološke domene, ume-rene performanse za respiratorne/bubrežne, a najskromnije za mentalne i infektivne ishode — obrazac koji je u skladu sa prirodom ovih stanja (duže prekliničke faze naspram spoljašnjih okidača) i ranijim nalazima u literaturi (1,3). Kalibracija je bila zadovoljavajuća u srednjim opsezima rizika, dok su očekivana odstupanja zabeležena na ekstremima; analiza krive odluke (DCA) ukazala je na pozitivnu neto korist u pr-govima relevantnim za oportunistički skrining (12,13). Sveukupno, rezultati podupiru hipotezu da jedan generički model može pružiti „barometar zdravlja“ koji istovremeno obuhvata veliki broj kliničkih ishoda, umesto da se oslanjamo na mnoštvo izolovanih „modela po ishodu“ (1,2).

Poređenje sa postojećom literaturom

Naše nalaze treba posmatrati u kontinuitetu sa dva trenda. Prvo, klasični pristup predikciji — kao što je QRISK3 za kardiovaskularni rizik — istorijski je razvijan kao pojedinačni model za jedan ishod (2). Drugo, noviji radovi nad EHR-om (npr. BEHRT) pokazuju da transformeri mogu učiti reprezentacije sekvenci kliničkih događaja sa višestrukim budućim ishodima (3). Generativni transformer koji „modeluje

prirodnu istoriju bolesti“ (1) predstavlja kvalitativno proširenje ovog drugog trenda: isti mehanizam koji uspešno predviđa sledeću „reč“ u jeziku ili sledeći „događaj“ u kliničkom toku, ovde se preoblikuje u multitask predviđanje rizika kroz vreme. Naši rezultati — bolja ili uporediva AUPRC u visoko prevalentnim hroničnim domenima — u skladu su sa tezom da sekvencijalna struktura i dužina istorije nose ključnu informaciju koja se gubi u tradicionalnim statičkim modelima (1,3). Takođe, naglašavanje AUPRC kod neuravnoteženih ishoda, uz paralelni prikaz AUROC, oslanja se na preporuke i rasprave iz metodološke literature (10,11).

Kliničke i javnozdravstvene implikacije

Primenjeno u realnom okruženju, ovakav model bi mogao da pruži ranu, širokougaonu procenu rizika i time podrži personalizovanu prevenciju i planiranje skrininga — u duhu konvergencije „visokoperformansne medicine“ i AI (16). Nalazi DCA sugerišu da ciljano usmeravanje dodatnih pregleda/laboratorija kod osoba iznad pragova (npr. 5–15% za 10-godišnji rizik) može doneti neto korist naspram strategija „tretiraj sve/nikog“ (13). U praksi, takav alat bi se ne sme koristiti kao samostalni dijagnostički mehanizam, već kao trijažni i preventivni signal koji upućuje gde dodatno proveriti (1,16).

Snage i ograničenja

Snage rada su: (i) transparentan, reproduktibilan pipeline sa javno dostupnim/sintetičkim podacima (5,6,8); (ii) eksplicitno mapiranje kodova (ICD-10/ICD-10-CM → PheCode) kako bi se ciljevi klinički zaokružili i dostigli obim ≥ 1.000 oboljenja (9); (iii) ocena diskriminacije, kalibracije i kliničke korisnosti (AUPRC/AUROC, Brier, intercept/nagib, DCA) (10–13). Ograničenja: (a) primarna evaluacija na sintetičkoj kohorti — iako verno oponaša statističke obrasce, ne zamenjuje heterogenost realnih populacija (6); (b) sporedna provera na MIMIC-III ograničena je na kritičnu negu i ne predstavlja opštu populaciju (5); (c) i pored mapiranja, agregacija kodova može prikriti fine razlike među entitetima; (d) ekstremni opsezi rizika ostaju izazov za kalibraciju (12). Ove tačke nalažu oprez pri tumačenju apsolutnih verovatnoća i upućuju na potrebu post-hoc kalibracije u ciljnoj populaciji.

Generalizabilnost, transportabilnost i pravičnost

Za odgovorno uvođenje u praksu, potrebna je eksterna validacija u reprezentativnim kohortama (npr. UK Biobank) i u različitim zdravstvenim sistemima (4). Posebno je važno testirati transportabilnost modela u demografski i socijalno različitim grupama

i pratiti metrike pravičnosti; istorijski primeri su pokazali da algoritmi mogu nenamerno kodifikovati sistemske nejednakosti (14,15). Naš pristup meri performanse po polu i dobi kao minimalni korak; ipak, potpuna procena pravičnosti zahteva širi spektar varijabli (etnička pripadnost, status, komorbiditeti) i lokalnu rekalkibraciju (14,15).

Etika, upravljanje i izveštavanje

Korišćenjem de-identifikovanih i sintetičkih podataka demonstrirali smo metodološku izvodljivost bez potrebe za formalnom odlukom etičkog odbora, uz napomenu da institucije mogu tražiti „exempt“ potvrdu i da svako naknadno korišćenje realnih podataka zahteva odgovarajuću dozvolu (5,6,8). S obzirom na smernice o upravljanju i etici AI u zdravstvu (7), preporučujemo: (i) eksplicitan plan upravljanja pristrasnostima; (ii) transparentno TRIPOD+AI izveštavanje (8); (iii) nadziranu primenu — model kao pomoćni, a ne autonomni alat (7,16).

Budući pravci

Dalji koraci uključuju: (1) prospektivnu, pragmatičnu evaluaciju u kliničkim tokovima (npr. ambulantna trijaža sa „soft“ ishodima poput upućivanja na skrining); (2) multimodalno proširenje (laboratorije, tekst kliničkih beležaka) uz kontrolu „information leakage-a“; (3) rekalkibraciju i adaptaciju na ciljnu populaciju, uz kontinuirani nadzor pravičnosti (12,14,15); (4) simulacije javnozdravstvenog uticaja kroz scenarije „šta-ako“ (npr. koliki se teret bolesti odlaže ciljanim intervencijama). Naš rad daje praktičan, etički prihvatljiv temelj za ove aktivnosti.

ZAKLJUČAK

U ovom radu predstavili smo metodološki jasan i reproduktibilan pristup za desetogodišnje predviđanje multimorbiditeta pomoću generativnog, sekvencijalnog modela zasnovanog na transformer arhitekturi. Pokazali smo da je moguće jednim „opštim“ modelom obuhvatiti preko 1.000 ciljanih oboljenja i proizvesti stabilne procene rizika, uz dobru kalibraciju u srednjim opsezima i pozitivnu neto korist na pragovima relevantnim za oportunistički skrining. Time smo pomerili fokus sa izolovanih modela „po ishodu“ ka jedinstvenom „barometru zdravlja“ koji može podržati kliničku odluku i javnozdravstveno planiranje.

Pristup je koncipiran tako da bude primenljiv bez obrade identifikabilnih podataka: primarna obuka i evaluacija vrše se na sintetičkoj kohorti, a tehničke provere na de-identifikovanim zapisima. To omogućava da se ključni delovi razvoja (ekstrakcija,

mapiranje, tokenizacija, obuka, evaluacija, pravičnost) završe u uslovima koji su regulatorno i etički prihvatljivi, dok se prelaz na realne populacije planira kroz jasno definisanu eksternu validaciju, re-kalibraciju i nadzor performansi.

Za kliničku primenu predlažemo pragmatičan put: (1) integracija modela u postojeći informatički tok kao pomoćni, trijažni signal; (2) izbor lokalno smislenih pragova rizika u dogovoru sa strukom; (3) post-hoc kalibracija na ciljnoj populaciji; (4) ugradnja kontrolnih tačaka za pravičnost (redovno praćenje metrika po ključnim podskupinama); (5) transparentno izveštavanje i audit trag (verzionisanje modela i podataka). Ovakav okvir podstiče bezbednu upotrebu u prevenciji, prioritetizaciji skrininga i upravljanju resursima, bez ambicije da zameni kliničko rasuđivanje.

Glavna ograničenja su svesno prihvaćena: sintetički podaci ne mogu u potpunosti zameniti heterogenost realnih populacija; de-identifikovani skupovi intenzivne nege imaju specifičnosti koje umanjuju generalizabilnost; ekstremno retki ili izrazito kontekstualni ishodi i dalje predstavljaju izazov za kalibraciju. Ovi rizici se ublažavaju planiranom eksternom validacijom, lokalnom re-kalibracijom i kontinuiranim nadzorom performansi posle implementacije.

Budući rad će ići u dva pravca. Prvi je prospektivna, pragmatična evaluacija u kliničkim ustanovama (sa definisanim „soft“ ishodima i procesnim metrikama), uz transparentan plan upravljanja pristrasnostima. Drugi je metodološko proširenje: multimodalno uključivanje laboratorija i narativa iz kliničkih beležaka, naprednije tehnike kalibracije i adaptacije, te simulacije javnozdravstvenog uticaja zasnovane na stvarnim ograničenjima kapaciteta.

Zaključno, ponudili smo praktičan, etički obziran i reproduktibilan metod za dugoročnu, višebolečinsku prognozu rizika. Smatramo da je najveća vrednost ovog pristupa u tome što omogućava lekarima i donosiocima odluka da na jednom mestu sagledaju višegodišnji profil rizika, ciljano usmere prevenciju i skrining i time unaprede ishode uz racionalnije korišćenje resursa.

Reference

1. Shmatko A, Jung AW, Gerstung M. Learning the natural history of human disease with generative transformers. *Nature*. 2025 Sep 17. Doi:10.1038/s41586-025-09529-3.
2. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms. *BMJ*. 2017;357:j2099. Doi:10.1136/bmj.j2099.
3. Li Y, Rao S, Solares JRA, et al. BEHRT: Transformer for Electronic Health Records. *Sci Rep*. 2020;10:7155. Doi:10.1038/s41598-020-62922-y.
4. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource... *PLoS Med*. 2015;12(3):e1001779. Doi:10.1371/journal.pmed.1001779.
5. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035. Doi:10.1038/sdata.2016.35.

6. Walonoski J, Kramer M, Nichols J, et al. Synthea: generating synthetic patients and HER. *J Am Med Inform Assoc.* 2018;25(3):230-238. Doi:10.1093/jamia/ocx079.
7. Boudierhem R. Shaping the future of AI in healthcare through ethics and governance. *Humanit Soc Sci Commun.* 2024;11:416. Doi:10.1057/s41599-024-02894-w.
8. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement. *BMJ.* 2024;385:e078378. Doi:10.1136/bmj-2023-078378.
9. Wu P, Gifford A, Meng X, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes. *JMIR Med Inform.* 2019;7(4):e14325. Doi:10.2196/14325.
10. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot... *PLOS ONE.* 2015;10(3):e0118432. Doi:10.1371/journal.pone.0118432.
11. Richardson E, Trevizani R, Greenbaum JA, et al. The ROC curve accurately assesses imbalanced datasets. *Patterns (N Y).* 2024;5(6):100994. Doi:10.1016/j.patter.2024.100994.
12. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med.* 2019;17:230. Doi:10.1186/s12916-019-1466-7.
13. Vickers AJ, Elkin EB. Decision curve analysis. *Med Decis Making.* 2006;26(6):565-574. Doi:10.1177/0272989X06295361.
14. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-453. Doi:10.1126/science.aax2342.
15. Liu M, Cruz Rivera S, Moher D, et al. A translational perspective towards clinical AI fairness. *NPJ Digit Med.* 2023;6:172. Doi:10.1038/s41746-023-00918-4.
16. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25:44-56. Doi:10.1038/s41591-018-0300-7.