



UNIVERZITET CRNE GORE
ELEKTROTEHNIČKI FAKULTET

PRIMJENA VJEŠTAČKE INTELIGENCIJE ZA AUTOMATSKU ANALIZU FLUOROSKOPSKIH RENDGENSKIH SNIMAKA



IVAN MARTINOVIĆ

DOKTORSKA DISERTACIJA

PODGORICA, 2025. GODINA



UNIVERZITET CRNE GORE
ELEKTROTEHNIČKI FAKULTET

Ivan Martinović

**PRIMJENA VJEŠTAČKE INTELIGENCIJE ZA
AUTOMATSKU ANALIZU FLUOROSKOPSKIH
RENDGENSKIH SNIMAKA**

-Doktorska disertacija-

Podgorica, 2025. godine

UNIVERSITY OF MONTENEGRO
FACULTY OF ELECTRICAL ENGINEERING

Ivan Martinović

**APPLICATION OF ARTIFICIAL INTELLIGENCE FOR
AUTOMATIC ANALYSIS OF FLUOROSCOPIC X-RAY
IMAGES**

- A thesis submitted for the degree of *Doctor of Philosophy*-

Podgorica, 2025. godine

“The greatest glory in living lies not in never falling, but in rising every time we fall.”

Nelson Mandela

PODACI O DOKTORANDU, MENTORIMA I ČLANOVIMA KOMISIJE

DOKTORAND

Ime i prezime: **Ivan Martinović**

Datum i mjesto rođenja: **03.08.1995. godine, Cetinje, Crna Gora**

Naziv završenog postdiplomskog Studijskog programa: **Elektrotehnički fakultet, odsjek *Elektronika, telekomunikacije i računari*, smjer *Elektronika – magistarske studije***

Godina završetka: **2020. godine**

MENTORI: **dr Ervin Sejdić**, redovni profesor, Univerzitet u Torontu, the Edward S. Rogers Sr. Department of Electrical & Computer Engineering
dr Milena Đukanović, vanredni profesor, Univerzitet Crne Gore, Elektrotehnički fakultet

KOMISIJA ZA OCJENU PODOBNOSTI TEZE I KANDIDATA: **dr Ljubiša Stanković**, redovni profesor, Univerzitet Crne Gore, Elektrotehnički fakultet
dr Ervin Sejdić, redovni profesor, Univerzitet u Torontu, the Edward S. Rogers Sr. Department of Electrical & Computer Engineering
dr Milena Đukanović, vanredni profesor, Univerzitet Crne Gore, Elektrotehnički fakultet
dr Goran Nikolić, redovni profesor, Univerzitet Crne Gore, Medicinski fakultet
dr. sc. Sven Lončarić, redovni profesor, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva

KOMISIJA ZA OCJENU DOKTORSKE DISERTACIJE: **dr Ljubiša Stanković**, redovni profesor, Univerzitet Crne Gore, Elektrotehnički fakultet
dr Ervin Sejdić, redovni profesor, Univerzitet u Torontu, the Edward S. Rogers Sr. Department of Electrical & Computer Engineering
dr Milena Đukanović, vanredni profesor, Univerzitet Crne Gore, Elektrotehnički fakultet
dr Goran Nikolić, redovni profesor, Univerzitet Crne Gore, Medicinski fakultet
dr. sc. Sven Lončarić, redovni profesor, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva

KOMISIJA ZA ODBRANU DOKTORSKE DISERTACIJE: **dr Ljubiša Stanković**, redovni profesor, Univerzitet Crne Gore, Elektrotehnički fakultet
dr Ervin Sejdić, redovni profesor, Univerzitet u Torontu, the Edward S. Rogers Sr. Department of Electrical & Computer Engineering
dr Milena Đukanović, vanredni profesor, Univerzitet Crne Gore, Elektrotehnički fakultet
dr Goran Nikolić, redovni profesor, Univerzitet Crne Gore, Medicinski fakultet
dr. sc. Sven Lončarić, redovni profesor, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva

DATUM ODBRANE: 18.02.2026.godine

PODACI O DOKTORSKOJ DISERTACIJI

Naziv doktorskih studija:	Doktorske studije elektrotehnike
Naslov doktorske disertacije:	Primjena vještačke inteligencije za analizu fluoroskopskih rendgenskih snimaka
Ključne riječi:	vještačka inteligencija, fluoroskopski rendgenski snimci, disfagija, sigurnost gutanja, efikasnost gutanja, mašinsko učenje, self-supervised learning, supervised learning, segmentacija, detekcija karakterističnih tačaka
Datum prijave doktorske teze:	28.01.2022. godine
Datum sjednice Senata UCG na kojoj je prihvaćena teza:	01.06.2022. godine
Naučna oblast:	Računarske nauke
Uža naučna oblast:	Biomedicinski inženjering

REZIME:

U ovom doktorskom radu prikazana je primjena mašinskog učenja u automatskoj analizi fluoroskopskih rendgenskih snimaka. Fokus je stavljen na analizu fluoroskopskih rendgenskih snimaka nastalih tokom videofluoroskopske studije gutanja, dijagnostičke metode koja se najčešće primjenjuje u procjeni poremećaja gutanja. Poremećaj gutanja, tj. disfagija, manifestuje se kroz uticaj na dva ključna funkcionalna aspekta gutanja, na sigurnost i efikasnost. Automatska analiza sigurnosti gutanja privlači više pažnje istraživača širom svijeta, dok oblast automatske analize efikasnosti gutanja ostaje nedovoljno istražena. U ovom doktorskom radu fokus je stavljen na oba funkcionalna aspekta. Prvo je razvijena superiorna metoda, koja se temelji na self-supervised learning pristupu, za automatsku analizu sigurnosti gutanja. Nakon toga kombinovanjem postojećih metoda baziranih na mašinskom učenju omogućena je automatska analiza efikasnosti gutanja. Razvojem automatske metode analize efikasnosti gutanja, uspostavlja se nova istraživačka oblast. Dobijeni rezultati jasno ukazuju da se mašinsko učenje sa velikom uspješnošću može koristiti za automatsku analizu fluoroskopskih rendgenskih snimaka kreiranih tokom videofluoroskopske studije gutanja.

INFORMATION ON DOCTORAL DISSERTATION

PhD study program:: **PhD studies in Electrical Engineering**

Dissertation title:: **Application of artificial intelligence for automatic analysis of fluoroscopic x-ray images**

Keywords: **artificial intelligence, fluoroscopic X-ray images, dysphagia, swallowing safety, swallowing efficiency, machine learning, self-supervised learning, supervised learning, segmentation, landmark detection task**

Thesis application date: **28.01.2022.**

Thesis acceptance date (UoM Senate Session): **01.06.2022.**

Scientific area: **Computer Science**

Specific scientific area: **Biomedical engineering**

ABSTRACT:

This PhD research presents the application of machine learning in the automated analysis of fluoroscopic X-ray images. The focus is placed on the analysis of fluoroscopic X-ray images obtained during a videofluoroscopic swallowing study, a diagnostic method most commonly used for the assessment of swallowing disorders. Swallowing disorders, or dysphagia, manifest through impacts on two key functional aspects of swallowing: safety and efficiency. Automated analysis of swallowing safety has attracted significant attention from researchers worldwide, while the area of automated analysis of swallowing efficiency remains underexplored. This PhD thesis addresses both functional aspects. First, a superior method based on a self-supervised learning approach was developed for the automated analysis of swallowing safety. Then, by combining existing machine learning-based methods, an automated analysis of swallowing efficiency was achieved. The development of an automated method for analyzing swallowing efficiency establishes a new research area. The results clearly demonstrate that machine learning can be successfully applied for the automated analysis of fluoroscopic X-ray images created during videofluoroscopic swallowing studies.

Sadržaj

Abstrakt	
Abstract.....	
1. Uvod	1
1.1. Motivacija za realizaciju doktorskog istraživanja	1
1.2. Anatomija i fiziologija gutanja	2
1.2.1. Procjena poremećaja gutanja	4
1.3. Cilj doktorskog istraživanja	8
1.4. Doprinos doktorskog istraživanja	9
1.5. Struktura doktorske teze	9
2. Pregled literature.....	10
2.1. Uvod	10
2.2. Pregled self-supervised learning pristupa.....	12
2.2.1. Generativni modeli	13
2.2.2. Konstruktivni modeli	14
2.3. Primjene u rendgenskim modalitetima	16
2.3.1. Konvencionalna radiografija	18
2.3.2. Kompjuterizovana tomografija (CT).....	20
2.3.3. Mamografija	23
2.3.4. Dentalna radiografija	25
2.3.5. Multimodalne aplikacije	26
2.3.6. Ostale primjene.....	28
2.4. Glavni benefiti i izazovi	31
2.5. Zaključak	33
3. Oblast istraživanja	34
3.1. Segmentacija bolusa koristeći self-supervised learning pristup	34
3.1.1. Motivacija.....	34
3.1.2. Strategija.....	34
3.2. Automatsko mjerenje faringealnog ostatka	35
3.2.1. Motivacija.....	35
3.2.2. Strategija.....	35
4. Segmentacija bolusa koristeći self-supervised learning pristup	36
4.1. Uvod	37
4.2. Materijali i Metode	38
4.2.1. Kreiranje skupa podataka	38
4.2.2. Self-supervised pretraining.....	39
4.2.3. Downstream zadatak.....	40
4.2.4. Ensemble model strategija.....	41

4.2.5. Statistička analiza	43
4.3. Rezultati.....	43
4.4. Diskusija	46
4.5. Zaključak	48
5. Automatsko mjerenje faringealnog ostatka	50
5.1. Uvod	50
5.2. Materijali i Metode	52
5.2.1. Kreiranje skupa podataka	52
5.2.2. Klinička analiza efikasnosti gutanja	52
5.2.3. Automatska analiza efikasnosti gutanja	53
5.2.4. Statistička analiza	54
5.3. Rezultati i Diskusija.....	54
5.4. Zaključak	56
6. Zaključak	57
Literature	
Biografija	

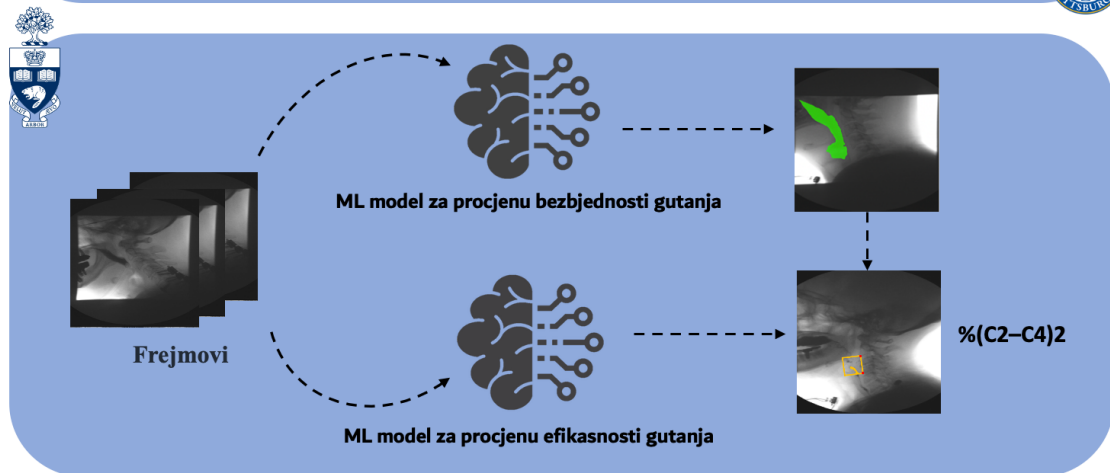
Spisak slika

Slika 1. Anatomske strukture uključene u proces gutanja.....	1
Slika 2. Oralna preparatorna faza	3
Slika 3. Oralna transportna faza	3
Slika 4. Faringealna faza	4
Slika 5. Ezofagealna faza.....	4
Slika 6. Videofluoroscopska studija gutanja/ modifikovana studija gutanja barijuma	7
Slika 7. Supervised vs SSL u medicinskom snimanju.....	12
Slika 8. SSL pristupi: a) generativni i b) kontrastivni modeli sematski dijagrami.....	15
Slika 9. Prikaz procesa skrininga studija u vodećim naučnim bazama	17
Slika 10. Contrastive Random Walk model	40
Slika 11. Schema zadatka segmentacije bolusa.....	42
Slika 12. Virtuelna ploča weighted average ensemble modela	42
Slika 13. Grad-CAM vizuelizacije koje ilustruju donošenje odluka modela kroz tri faze gutanja u zadatku segmentacije bolusa.....	46
Slika 14. Automatska kvantifikacija faringealnog ostatka koristeći kliničku metodu $\%(C_2-C_4)^2$	54

Spisak tabela

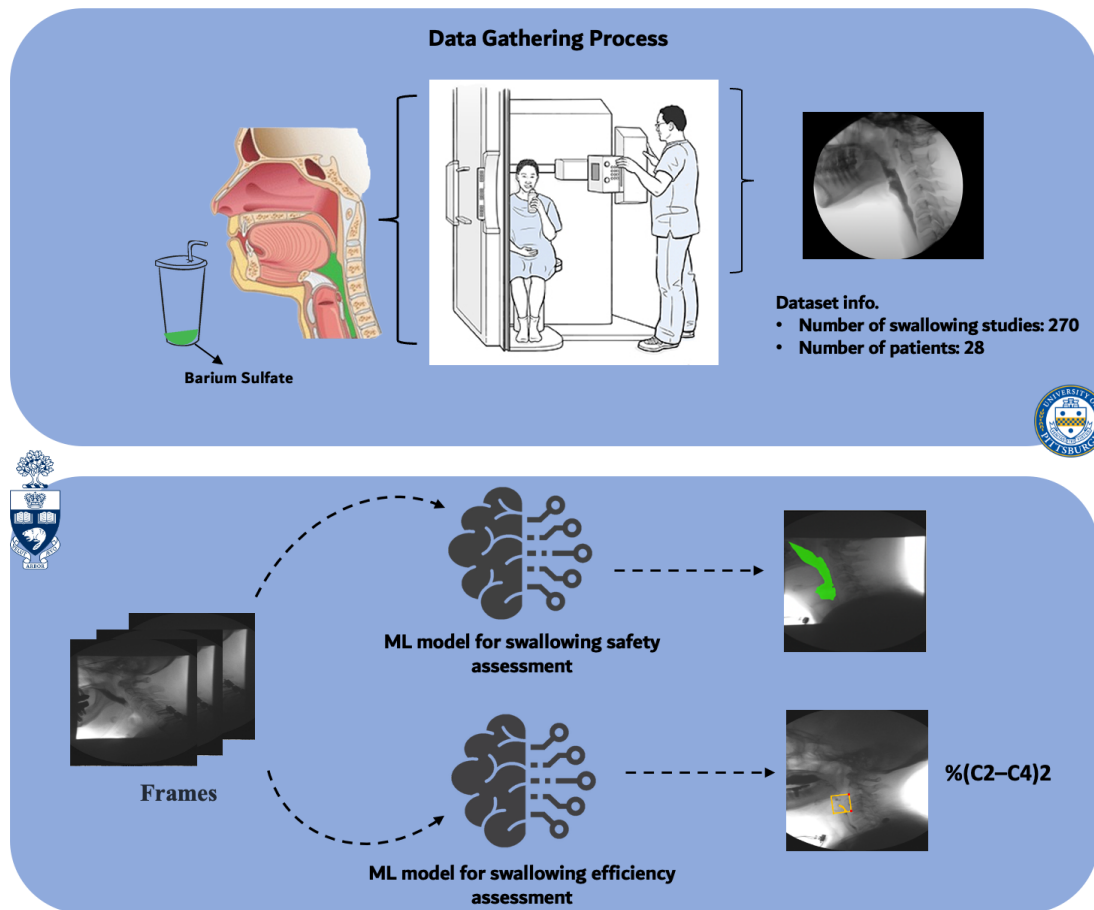
Tabela 1. Primjena SSL pristupa u konvencionalnoj radiografiji	19
Tabela 2. Primjena SSL pristupa u CT	21
Tabela 3. Primjena SSL pristupa u mamografiji	24
Tabela 4. Primjena SSL pristupa u dentalnoj radiografiji	25
Tabela 5. Primjena SSL pristupa u multimodalnim aplikacijama	27
Tabela 6. SSL pristup u fazi preprocesing i fazi arhiviranja	30
Tabela 7. Demografski podaci pacijenata čiji su podaci uključeni u skup podataka	38
Tabela 8. Performanse zadatka segmentacije bolusa na osnovu različitih vrijednosti hiperparametara u okviru pretext zadatka (F1 score)	44
Tabela 9. Performanse zadatka segmentacije bolusa kroz dva različita pristupa 9 nasumičnih podjela (split) skupa podataka (F1 score)	45
Tabela 10. Kvantitativne metode bazirane na mjerenje površine ostataka u pikselima .	53

Grafički Abstrakt



Proces prikupljanja podataka sproveden je u Prezbitarijanskoj univerzitetskoj bolnici Medicinskog centra Univerziteta u Pittsburgu. Prikupljeni skup podataka je korišćen u razvoju modela mašinskog učenja (ML) namijenjenih automatskoj procjeni bezbednosti i efikasnosti procesa gutanja.

Graphical Abstract



The data-gathering process was conducted at the Presbyterian University Hospital of the University of Pittsburgh Medical Center. The collected dataset was used in the development of machine learning models, including the creation of models for automatic swallowing safety and efficiency assessment.

Abstrakt

U ovom doktorskom radu prikazana je primjena mašinskog učenja u automatskoj analizi fluoroskopskih rendgenskih snimaka. Fokus je stavljen na analizu fluoroskopskih rendgenskih snimaka nastalih tokom videofluoroskopske studije gutanja, dijagnostičke metode koja se najčešće primjenjuje u procjeni poremećaja gutanja. Poremećaj gutanja, tj. disfagija, manifestuje se kroz uticaj na dva ključna funkcionalna aspekta gutanja, na sigurnost i efikasnost. Automatska analiza sigurnosti gutanja privlači više pažnje istraživača širom svijeta, dok oblast automatske analize efikasnosti gutanja ostaje nedovoljno istražena. U ovom doktorskom radu fokus je stavljen na oba funkcionalna aspekta. Prvo je razvijena superiorna metoda, koja se temelji na self-supervised learning pristupu, za automatsku analizu sigurnosti gutanja. Nakon toga kombinovanjem postojećih metoda baziranih na mašinskom učenju omogućena je automatska analiza efikasnosti gutanja. Razvojem automatske metode analize efikasnosti gutanja, uspostavlja se nova istraživačka oblast. Dobijeni rezultati jasno ukazuju da se mašinsko učenje sa velikom uspješnošću može koristiti za automatsku analizu fluoroskopskih rendgenskih snimaka kreiranih tokom videofluoroskopske studije gutanja.

Ključne riječi: vještačka inteligencija, fluoroskopski rendgenski snimci, disfagija, sigurnost gutanja, efikasnost gutanja, mašinsko učenje, self-supervised learning, supervised learning, segmentacija, detekcija karakterističnih tačaka

Abstract

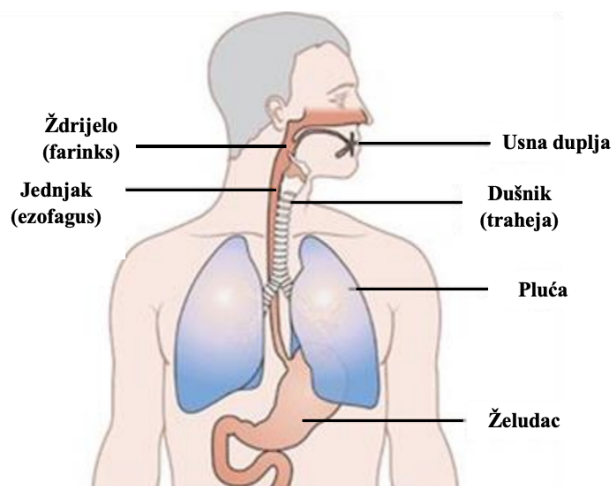
This PhD research presents the application of machine learning in the automated analysis of fluoroscopic X-ray images. The focus is placed on the analysis of fluoroscopic X-ray images obtained during a videofluoroscopic swallowing study, a diagnostic method most commonly used for the assessment of swallowing disorders. Swallowing disorders, or dysphagia, manifest through impacts on two key functional aspects of swallowing: safety and efficiency. Automated analysis of swallowing safety has attracted significant attention from researchers worldwide, while the area of automated analysis of swallowing efficiency remains underexplored. This PhD thesis addresses both functional aspects. First, a superior method based on a self-supervised learning approach was developed for the automated analysis of swallowing safety. Then, by combining existing machine learning-based methods, an automated analysis of swallowing efficiency was achieved. The development of an automated method for analyzing swallowing efficiency establishes a new research area. The results clearly demonstrate that machine learning can be successfully applied for the automated analysis of fluoroscopic X-ray images created during videofluoroscopic swallowing studies.

Keywords: Artificial Intelligence, fluoroscopic X-ray images, dysphagia, swallowing safety, swallowing efficiency, machine learning, self-supervised learning, supervised learning, segmentation, landmark detection task

1. Uvod

1.1. Motivacija za realizaciju doktorskog istraživanja

Gutanje predstavlja vitalan proces koji obezbeđuje nesmetan prolazak bolusa¹ od usne duplje do želuca kroz ždrijelo (farinks) i jednjak (ezofagus). Složenost ovog procesa ogleda se u činjenici da je gutanje usko povezano sa disanjem, što zahtijeva preciznu koordinaciju (Slika 1.) [1], [2]. U slučaju da proces gutanja nije precizno koordinisan dolazi do poremećaja u gutanju, koji se naziva disfagija [3]. Ovo zdravstveno stanje izaziva niz kliničkih posledica za pacijente, među kojima je najozbiljnija aspiracija bolusa u respiratorni sistem. Aspiracija bolusa je praćena gušenjem ili kašljanjem prije, tokom ili nakon gutanja. Aspiracija može rezultirati plućnom infekcijom, upalom pluća (pneumonijom) ili preranom smrću [4]–[7]. Takođe, disfagija se može manifestovati i blažim simptomima kakvi su bol tokom gutanja ili poteškoće pri započinjanju procesa gutanja. Pored opstrukcije disajnih puteva, disfagija dovodi do dehidracije, nehranjenosti, gubitka težine, kao i pada imunog sistema [8].



Slika 1. Anatomske strukture uključene u proces gutanja
ilustracija kreirana po uzoru na [9]

Iako se disfagija može javiti kod ljudi svih uzrasta, starija populacija je pod povećanim rizikom prvenstveno zbog oboljenja koja su karakteristična za ovo životno

¹ Bolus-potiče od grčke riječi bōlos (gruda) i odnosi se na meku masu sažvakane hrane ili tečnosti koja je spremna da se proguta.

doba, kao što su: moždani udar, Parkinsonova bolest, demencija, karcinom glave i vrata [10], [11]. U prilog tome govore podaci kanadskog društva za intestinalno istraživanje navodeći da je 19-33% osoba starijih od 80 godina i čak 50% korisnika usluga domova za njegu starijih osoba pogođeno disfagijom [12]. Dodatno, istraživanja pokazuju da je disfagija rasprostranjena kod 8.1-80% pacijenata sa moždanim udarom, 30% pacijenata sa povredom mozga, kao i 11-60% pacijenata sa Parkinsonovom bolešću [4], [7]. Računajući ukupan broj pacijenata u svijetu on se kreće oko 8% svjetske populacije što je negdje oko 590 miliona ljudi [10]. U budućnosti možemo očekivati povećanje incidencije disfagije usled starenja svjetske populacije. Iz tog razloga, dijagnostičke metode i liječenje disfagije potrebno je stalno unapređivati [13] fokusirajući se na multidisciplinarni pristup kakav je prikazan u ovoj doktorskoj tezi.

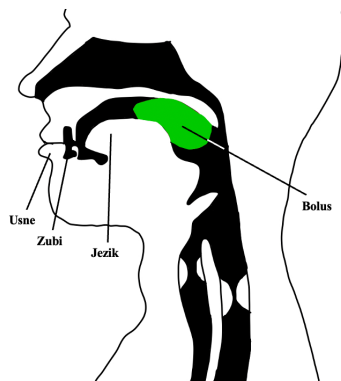
1.2. Anatomija i fiziologija gutanja

Gutanje je izuzetno složen neuromišićni proces koji uključuje više od 30 nerava i mišića [3], [14]. Može se opisati u tri faze (oralna, faringealna i ezofagealna), koje kod zdravih osoba traju po jednu sekundu [3], [15]. Oralnu fazu je moguće kontrolisati, dok su preostale dvije refleksne. Pomenute faze se odnose na uzimanje tečnosti, dok se kod uzimanja čvrste hrane može govoriti o četiri faze. Razlika u broju faza proističe iz činjenice da se oralna faza kod uzimanja čvrste hrane dijeli na oralnu preparatornu i oralnu transportnu.

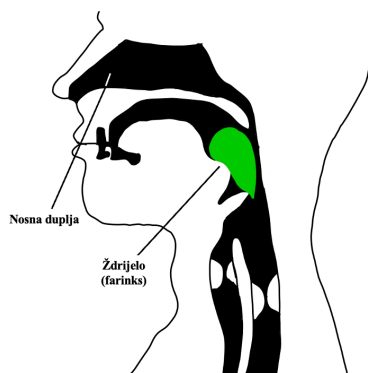
Uzimanjem tečnosti oralna faza otpočinje podizanjem zadnjeg dijela jezika kako bi se spriječilo prelivanje tečnosti u farinks. U literaturi se navodi da prelivanje tečnosti u farinks nije rijetka pojava i da se povećava sa procesom starenja [16]. Završetak oralne faze obilježen je aktivacijom mišića jezika i obraznih mišića, koji potiskuju tečnost ka orofarinksu.

Oralna preparatorna faza podrazumijeva sam postupak uzimanja hrane i formiranje bolusa adekvatne konzistencije postupkom žvakanja (Slika 2.). Uspješnost ove faze se obezbjeđuje aktivnošću usana, obraza, jezika i vilica. Nakon oralne preparatorne faze slijedi oralna transportna faza (Slika 3.) u kojoj se bolus postavlja na površinu jezika i potiskuje prema orofarinksu. Prelazak bolusa u farinks omogućuje se otvaranjem

zadnjeg glosovelarnog ventila, dok se vrijeme od ulaska bolusa u farinks do početka faringealne motoričke aktivnosti naziva vrijeme prelazne faze [4], [17].

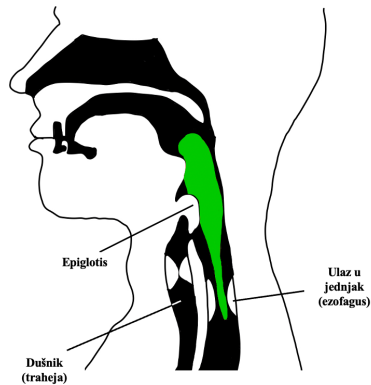


Slika 2. Oralna preparatorna faza
ilustracija kreirana po uzoru na [9]



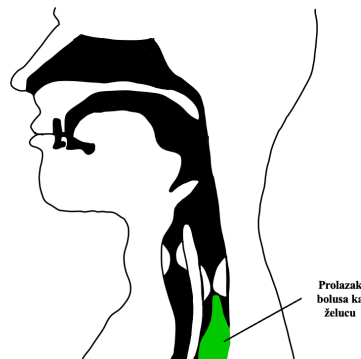
Slika 3. Oralna transportna faza
ilustracija kreirana po uzoru na [9]

Faringealna faza uključuje prolazak bolusa kroz farinks i gornji ezofagealni sfinkter do ezofagusa uz složenu koordinaciju mnogih struktura (Slika 4.). Jednom kada gutanje dođe u faringealnu fazu inicira se niz sekvenci koje se ne mogu zaustaviti. Važnu komponenta ove faze predstavlja i zaštita disajnih puteva od aspiracije bolusa. Zaštita se u prvom redu postiže zatvaranjem laringealnog vestibula, koji se smatra prvom linijom odbrane. Dodatno, zaštita se postiže i pomjeranjem hiodine kosti. Vertikalno pomjerenje hioidne kosti doprinosi zatvaranju epiglotisa i larinksa, dok anteriorno pomjerenje omogućava otvaranje gornjeg ezofagealnog sfinktera [4].



Slika 4. Faringealna faza
ilustracija kreirana po uzoru na [9]

Poslednja faza gutanja je ezofagealna faza koja uključuje niz peristaltičkih talasa sa parcijalnom relaksacijom i kontrakcijom mišića ezofagusa (Slika 5.). U okviru ove faze omogućava se prolazak bolusa od gornjeg ezofagealnog sfinktera do želuca. Prelazak bolusa u želudac se omogućava relaksacijom donjeg ezofagealnog sfinktera, čija se relaksacija inicira momentom ulaska bolusa u ezofagus.



Slika 5. Ezofagealna faza
ilustracija kreirana po uzoru na [9]

1.2.1. Procjena poremećaja gutanja

Procjena poremećaja gutanja zasniva se na skrining procedurama i dijagnostičkim procedurama [3]. Glavna razlika između ovih procedura leži u opsegu informacija koje se nude kliničarima u procesu procjene gutanja. Skrining procedure nude odgovor na pitanje „*da li postoje određeni simptomi koji ukazuju na poremećaje u gutanju?*“, dok dijagnostičke metode daju dublje informacije o eventualnim uzrocima poremećaja.

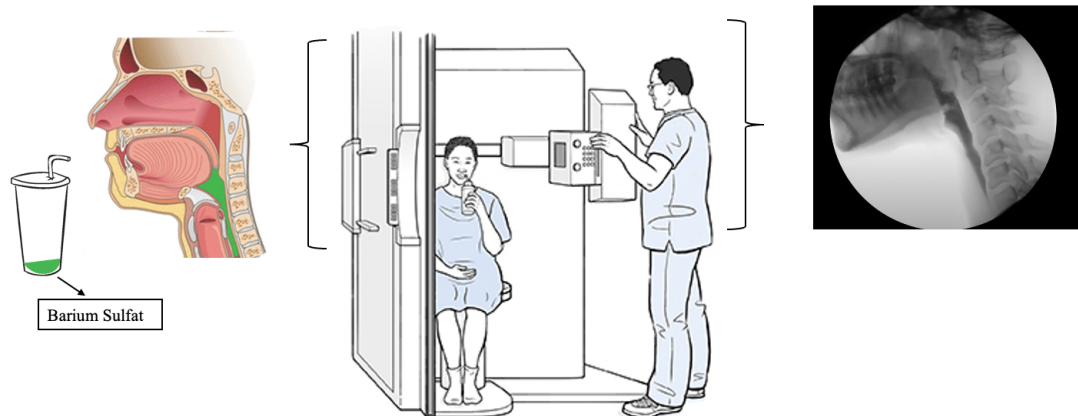
Prvi korak u procjeni procesa gutanja uključuje skrining procedure koje se uglavnom oslanjaju na non-imaging metode. Najrasprostranjenija skrining procedura je klinička ili bedside evaluacija. Značaj ove procedure je izuzetan budući da je ona često jedino dostupna metoda za procjenu procesa gutanja. Dodatno, u literaturi se navodi da bi klinička evaluacija trebalo da ponudi najbolje razumijevanje svakodnevne funkcije gutanja, jer ne uključuje modifikovanu hranu u procesu procjene kao neke dijagnostičke metode. Međutim, istraživanja su pokazala da su dijagnostičke metode neophodne s obzirom da koristeći skrining procedure kliničari mogu da otkriju svega 40%-60% pacijenata koji aspiriraju [3], [18], [19]. Glavni razlog za ovako lošu mogućnost procjene aspiracije leži u činjenici da je procjena faringealne faze, koristeći skrining procedure, značajno ograničena. Pored kliničke evaluacije kao pomoćne metode u procjeni gutanja koriste se cervikalna auskultacija i pulsna oksimetrija. Cervikalna auskultacija se zasniva na analizi zvukova koji se kreiraju tokom procesa gutanja. Snimanje zvukova se vrši uz pomoć stetoskopa ili mikrofona, dok se nastale vibracije mogu registrovati uz pomoć akcelerometra. Istraživanja pokazuju da su zvukovi gutanja kod zdravih osoba kratki (približno 0.4s) i stabilne jačine od 43dB sa srednjom frekvencijom od 2200Hz. Dodatno, na karakteristike zvukova gutanja kod zdravih osoba utiče starosna dob osobe, kao i količina i struktura bolusa. Prema tome, dužina trajanja zvukova gutanja se povećava kako starimo, dok se dužina skraćuje kako se povećava količina bolusa. Kod osoba sa disfagijom dužina zvukova gutanja je duža nego kod zdravih osoba iznoseći 0.92s kod pacijena sa dijagnostikovanom disfagijom i 0.6s kod pacijenata koji aspiriraju. Primjena pulsne oksimetrije, sa druge strane, omogućava identifikaciju pacijenata koji aspiriraju prateći nivo desaturacije kiseonika u arterijskoj krvi. Desaturacija kiseonika u arterijskoj krvi može nastati kao posledica aspiracije bolusa u disajnim putevima narušavajući na taj način sklad između ventilacije i perfuzije. Nivo desaturacije se određuje korišćenjem pulsno oksimetra koji se obično pozicionira na kaži prstu ili usnoj školjki. Koristeći pulsni oksimetar kliničarima se nude informacije o nivou saturacije kiseonika (SpO_2) u krvi, koja se kod zdravih osoba kreće u opsegu od 95% do 100%. Istraživanja su pokazala da se prilikom procjene poremećaja gutanja koristeći pulsnu oksimetriju prvo treba izmjeriti bazna vrijednost saturacije kiseonika u intervalu od oko 2 minuta [20]. Indikacija da se aspiracija dogodila donosi se ukoliko je zabilježena desaturacija viša od 2% od bazične vrijednosti. Dodatno, nivo desaturacije nije moguće registrovati u momentu kada

se aspiracija dogodila, već nakon nekoliko sekundi ili čak do nekoliko stotina sekundi. Istraživanja su pokazala da se kombinacijom kliničke evaluacije i pomoćnih metoda sa velikom vjerovatnoćom mogu otkriti pacijenti koji aspiriraju.

Ukoliko se upotrebom skrining procedura utvrde indicije o mogućem postojanju poremećaja tokom gutanja kliničari se odlučuju za dodatnu procjenu, koja uključuje dijagnostičke procedure. Tokom vremena razvijen je veliki broj dijagnostičkih procedura zasnovanih na imaging metodama, uključujući videofluoroskopsku studiju gutanja (VFSS), fibrooptičku endoskopsku procjenu gutanja, ultrazvuk i scintigrafiju [3], [14], [21]. Pored pomenutih procedura razvijene su specijalizovane procedure za analizu moždanih aktivnost tokom procesa gutanja kao što su: funkcionalna magnetna rezonanca (fMRI), elektromiografija (EMG), pozitronska emisiona tomografija (PET) i elektroencefalografija (EEG) [22]–[25].

Najpoznatija i najrasprostranjenija imaging procedura za procjenu poremećaja gutanja je videofluoroskopska studija gutanja (VFSS), takođe poznata kao modifikovana studija gutanja barijuma (Slika 6.). Ova dijagnostička procedura omogućava kliničarima da odrede važne parametre gutanja kao što su: oralno i faringealno tranzitno vrijeme, količinu aspiriranog bolusa, kao i količinu ostatka bolusa nakon gutanja. Međutim, ova procedura je više od dijagnostičkog alata, ona takođe pruža mogućnost da se počne sa planiranjem liječenja i procijeni efikasnost liječenja. Pošto VFSS stvara dinamičke slike na osnovu slabljenja rendgenskih zraka, da bi procedura bila uspješna, pacijentu se uz hranu ili tečnost dodaje kontrastno sredstvo barijum. Barijum omogućava kliničarima da precizno razlikuju faze gutanja prateći kretanje bolus na videofluoroskopskim (fluoroskopskim) snimcima. Uglavnom ova procedura se sprovodi uključujući više od jednog gutanja uzimajući hranu/tečnost različite konzistencije. Na ovaj način se želi odabrati ona konzistencija bolusa koja će na najbolji način omogućiti praćenje procesa gutanja. Dodatno, žele se izbjeći neželjene posledice po pacijente koji aspiriraju u slučaju da odabrana konzistencija bolusa prelazi dozvoljenu vrijednost. Pored izuzetnih prednosti videofluoroskopsku studiju gutanja karakterišu i određeni nedostaci kao što su: korišćenje jonizujućeg zračenja za dobijanje fluoroskopskih snimaka; korišćenje hrane i tečnosti tokom pregleda koje ne oslikavaju svakodnevni proces gutanja; unaprijed predodređeno vrijeme pregleda (zbog zaštite od jonizujućeg zračenja) koje može biti nedovoljno za neke pacijente; fokusiranje na analizu motoričkih funkcija zanemarujući analizu senzornih

funkcija; nepristupačnost procedure zbog njene cijene u manje razvijenim i rurarnim predjelima; nemogućnost korišćenja procedure za pacijente sa određenim oboljenima, kao npr. hemipareza; visoki zahtjevi u pogledu vremena za interpretaciju snimaka.



Slika 6. Videofluoroskopska studija gutanja/ modifikovana studija gutanja barijuma

Fibrooptička endoskopska procjena gutanja predstavlja proceduru koja je komplementarna videofluoroskopskoj studiji gutanja. Istraživači navode da ni jednu od ove dvije procedure ne treba označavati kao zlatni standard u analizi procesa gutanja jedino u slučaju ako se upoređuju sa skrining metodama. Glavne prednosti fibrooptičke endoskopske procjene gutanja uključuje njena portabilnost, mogućnost korišćenja hrane i tečnosti koja se koristi u svakodnevnom životu, njena cijena koja je znatno manja od VFSS, mogućnost ponavljanja procedure, mogućnost evaluacije senzorne funkcije. Ova procedura se obično koristi u onim situacijama gdje VFSS nije moguće iskoristiti dajući kliničarima direktni pregled unutrašnjih struktura uključenih u postupak gutanja iz horizontalne ravni. Slika koja se koristi za ocjenu procesa gutanja je u boji i dvodimenzionalna. Dobijanje slike se obezbjeđuje koristeći laringoskop, izvor svjetlosti, kameru, digitalni rekorder i monitor. Prečnik endoskopa koji se uvlači u nos je izuzetno mali i iznosi 3mm-4mm za odrasle i 2.2mm za djecu. Ova procedura omogućava pregled struktura kao što su meko nepce, orofarinks, hipofarinks, larinks, kao i ulaz u dušnik (traheja). Jedan od glavnih nedostataka ove procedure je činjenica da ona ne može pružiti informacije o ezofagealnoj fazi. Pored pomenutog nedostatka ova procedura dijeli karakteristiku svih ostalih imaging metoda, a to je loša pouzdanost između interpretatora (intrarater reliability). Iako su ove metode razvijene da pruže objektivnu procjenu procesa

gutanja interpretaciju dobijenih snimaka sprovode kliničari, koji ne mogu da garantuju potpunu objektivnost.

1.3. Cilj doktorskog istraživanja

U prethodnom izlaganju mogli smo da se uvjerimo u kompleksnost procesa gutanja i njegove procjene. Takođe, uvidjeli smo da su dijagnostičke procedure neizostavan dio liječenja pacijenata oboljelih od disfagije. Oslanjajući se na pomenuta zapažanja, glavni cilj ovog doktorskog istraživanja je unapređenje dijagnostičkih procedura razvojem automatske metode čije će glavne odlike biti objektivnost i efikasnost. Konkretno, fokus će biti na unapređenju videofluoroskopske studije gutanja. Odabirom ove dijagnostičke procedure želi se kreirati metode koja će imati široku primjenu u kliničkoj praksi.

Glavni izazov za kliničare u primjeni videofluoroskopske studije gutanja predstavlja vrijeme koje je potrebno za interpretaciju fluoroskopskih snimaka. Fluoroskopski snimci predstavljaju rendgenski video zapis kreiran sa frekvencijom od 30 frejmova u sekundi. Interpretacija fluoroskopskih snimaka između ostalog uključuje ocjenu mehaničkih funkcija procesa gutanja što zahtijeva detaljnu analizu snimaka, frejm po frejm [26]–[28]. Iz tog razloga u istraživačkim krugovima pojavila se ideja da se proces intepretacije automatizuje. Uzimajući u obzir složenost procesa gutanja kao najbolje rješenje za proces automatizacije nametnula se tehnologija vještačke inteligencije. Budući da se radi o vizuelnim podacima u okviru vještačke inteligencije posebno je pogodna primjena dubokih neuralnih mreža kakve su npr. konvolucione neuralne mreže.

Istraživačke aktivnosti na ovom polju su intezivirane u poslednjoj deceniji. Na svjetskom nivou izdvaja se rad istraživačke grupe iMED Lab [29] u kojoj je sprovedeno ovo doktorsko istraživanje. Pregledom dostupne literature jasno se može uočiti da se vještačka inteligencija sa velikim uspjehom može primijeniti u automatskoj analizi fluoroskopskih rendgenskih snimaka [26], [30], [31]. Međutim, pored uspjeha u primjeni vještačke inteligencije ipak postoji prostor za poboljšanje postojećih metoda, kao i razvoj novih. Kao glavni nedostatak postojećih metoda uglavnom se navodi nedostatak robusnosti i generalizacije što sprečava široku primjenu metoda u svakodnevnoj kliničkoj

praksi [32]. Dodatno, sve postojeće metode se zasnivaju na supervised learning pristupu gdje postoji nesaglasje između pretraining seta podataka (ImageNet) i downstream seta podataka (fluoroskopski rendgenski snimci) [33]. Analizom postojećih metoda takođe se može primijetiti da su one uglavnom fokusirane na automatskoj analizi parametara povezanih sa sigurnošću gutanja, dok nedostaju studije čiji je fokus automatska analiza parametrima povezanim sa efikasnošću procesa gutanja.

1.4. Doprinis doktorskog istraživanja

U ovom doktorskom radu postavljena je hipoteza da je moguće automatski odrediti parametre koji opisuju sigurnost i efikasnost gutanja. Koristeći mogućnosti vještačke inteligencije predlaže se da se parametri određuju analizom fluoroskopskih rendgenskih snimaka. Automatska analiza fluoroskopskih rendgenskih snimaka ima za cilj da se ljekarima olakša postupak uspostavljanja dijagnoze i praćenja procesa liječenja pacijenata. Da bi se postigli definisani ciljevi u okviru ovog doktorskog rada realizovana su sledeća istraživanja:

- Razvoj metode za automatsko segmentiranje bolusa koristeći self-supervised learning pristup.
- Razvoj metode za automatsko mjerenje faringealnog ostatka što je rezultiralo uspostavljanjem nove istraživačke oblasti.

1.5. Struktura doktorske teze

Poglavlje 2 obrađuje primjenu self-supervised learning pristupa u rendgenskim modalitetima. Poglavlje 3 predstavlja pregled tema koje su dio doktorskog istraživanja. Poglavlje 4 prikazuje razvoj metode za automatsko segmentiranje bolusa koristeći self-supervised learning pristup. U poglavlju 5 prikazan je istraživački rad koji je rezultirao uspostavljanjem nove istraživačke oblasti, koja obuhvata automatsko određivanje parametara koji opisuju efikasnost gutanja. Poslednje poglavlje uključuje zaključak i budući pravci u istraživačkoj oblasti.

2. Pregled literature

Pregled literature obuhvata istraživanje koje je sastavni dio naučnog rada Martinović I., Shitong M., Dousty M., Li W., Đukanović M., Colak E., Sejdić E. (2025), *X-ray Modalities in the Era of Artificial Intelligence: Overview of Self-Supervised Learning Approach, FACETS.*

Self-supervised learning pristup omogućava kreiranje algoritama koji nadmašuju supervised pre-training metode u velikom broju zadataka kompjuterskog vida (computer vision). U ovom poglavlju prikazan je sveobuhvatan pregled self-supervised learning pristupa u rendgenskim modalitetima uključujući: konvencionalnu radiografiju, kompjuterizovanu tomografiju, mamografiju i dentalnu radiografiju. Pored primjene self-supervised learning pristupa u fazi interpretacije rendgenskih snimaka u ovom poglavlju se naglašava važna uloga ovog pristupa u preprocesing fazi, kao i fazi arhiviranja radioloških snimaka. Štaviše, govori se o primjeni self-supervised learning pristupa u multi-modalnom scenariju, koji predstavlja budući ključni pravac u primjeni vještačke inteligencije u medicini. Na kraju, naglašavaju se glavni izazovi u razvoju self-supervised learning algoritama namijenjenih rendgenskim modalitetima. Na osnovu pregleda literature proističe da self-supervised learning pristup ima potencijal da bude „game changer“, prevazilazeći trenutnu situaciju u kojoj se samo mali broj od mnogobrojno razvijenih metoda baziranih na vještačkoj inteligenciji primjenjuje u svakodnevnoj kliničkoj praksi.

2.1. Uvod

Radiologija je uspostavljena prije više od 125 godina i omogućila je značajno smanjenje stope mortaliteta od različitih oboljenja kao što su pneumonija, karcinom, koronarna bolest srca i nefatalni infarkt miokarda [34]–[38]. Trenutno radiolozi se suočavaju sa izazovima u tumačenju ogromne količine radioloških snimaka što za posledicu može imati povećanje rizika od greške [39]. U cilju prevazilaženja pomenutog izazova kao rješenje se nameće razvoj kompjuterskih algoritama namijenjenih automatskoj analizi radioloških snimaka. Automatskom analizom radioloških snimaka

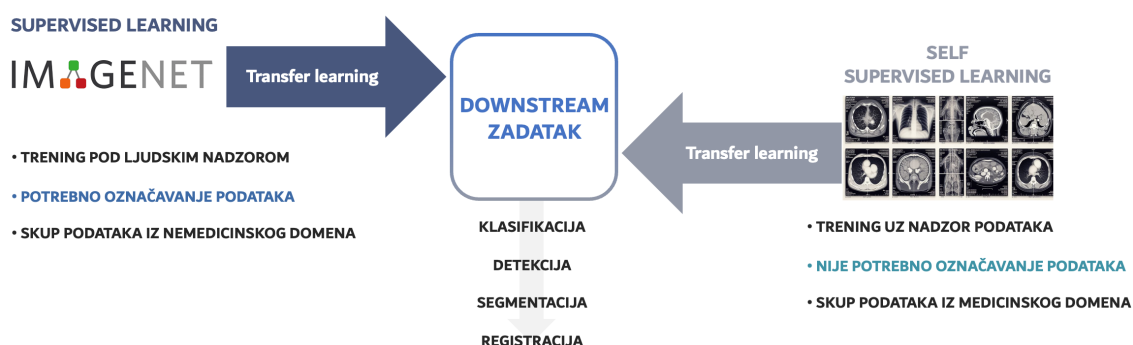
žele se postići sledeći ciljevi: povećanje osjetljivosti za suptilne promjene, davanje prioriteta hitnim slučajevima, automatizacija rutinskih zadataka, kao i ublažavanje nedostatka radiologa u zemljama u razvoju [40].

Istraživanja pokazuju da primjena mašinskog učenja (ML) u radiologiji može unaprijediti cijeli radni tok radiologije [41], omogućavajući povećanje dijagnostičke preciznosti i efikasnosti, personalizaciju plana liječenja, poboljšanje kvaliteta zdravstvene njege i potencijalno smanjenje izdataka za zdravstvenu zaštitu. Postojeći sistemi dubokog učenja namijenjeni automatskoj analizi radioloških snimaka postižu preciznost uporedivu iskusnim radiolozima [42]. Studije takođe pokazuju da korišćenje ML može pomoći ljekarima da efikasnije indentifikuju patološke promjene ili odstupanje u strukturama prikazanim na radiološkim snimcima [43]. Pored navedenih prednosti u primjeni ML, postoje otvorena pitanja koja je potrebno dodatno razmotriti. Koncept dubokog učenja temelji se na pristupu ogromnoj količini označenih podataka [44]. Ova odlika dubokog učenja je posebno problematična u medicinskim aplikacijama budući da su skupovi radioloških snimaka značajno manji (stotine/hiljade) u poređenju sa standardnim skupovima podataka (milion uzoraka). Dodatni izazov takođe predstavlja označavanje radioloških snimaka zahtijevajući značajno vrijeme, trud, kao i angažovanje eksperata [45]. Kako bi se prevazišao izazov u pogledu nedovoljne količine podataka uglavnom se koristi transfer learning tehnika. Transfer learning podrazumijeva obuku ML modela na ogromnim standardnim skupovima podataka, pa tek nakon toga obuku na radiološkim snimcima. Glavni nedostatak ove metode leži u činjenici da postoji veliki jaz u domenu između pomenuta dva skupa podataka. Prije svega zato što radiološki snimci imaju potpuno različite karakteristike od standardnih slika, kao što su nizak kvalitet, distribucija latentnih karakteristika, niska rezolucija, trodimenzionalna struktura i visoka sličnost anatomskih struktura među snimcima [46]–[48]. Neujednačenost podataka predstavlja još jedan problem u razvoju algoritama baziranih na dubokom učenju. Neujednačenost u podacima posebno je primjetan u kontekstu rijetkih oboljenja, gdje su većina snimaka u skupu podataka bez promjena. Dodatan problem predstavlja i situacija kada se ML model kreira koristeći podatke sakupljene iz različitih medicinskih centara i sa različitih uređaja. Na ovaj način unose se varijacije u domenu, kao što je odnos signal-šum. Svi navedeni izazovi utiču na robusnost i generalizaciju ML modela.

Stoga postavlja se logično pitanje: *Kako je moguće ublažiti ili prevazići navedene izazove?* Da bi istražili moguće odgovore na ovo pitanje, istraživači su razvili sisteme za automatsku analizu radioloških snimaka koristeći relativno novi self-supervised learning (SSL) pristup. Ovaj pristup omogućava izdvajanje smislenih reprezentacija iz neoznačenih podataka smanjujući na taj način zavisnost od velikih, označenih skupova podataka i poboljšavajući efikasnost i tačnosti analize radioloških snimaka. Sprovedeno istraživanje ukazuje da SSL pristup unapređuje primjenu ML u rendgenskim modalitetima. Iz tog razloga u ovom poglavlju se sumiraju doprinosi primjene SSL u rendgenskim modalitetima. Dodatno, prikazan je pregled trenutnog stanja, kao i glavnih izazova i budući pravci u ovoj oblasti.

2.2. Pregled self-supervised learning pristupa

SSL pristup omogućava učenje reprezentacija bez nadzora uključujući dva zadatka: pretext/proxy zadatak i downstream zadatak [49]. Pretext zadatak sprovodi se koristeći supervised pristup na neoznačenim podacima, omogućavajući modelu da nauči smislene reprezentacije kreiranjem pseudo oznaka (labela). U drugom koraku, naučene reprezentacije iz pretext zadatka se prenose u downstream zadatak radi fine-tuning-a [50]. Jedna od prednosti SSL pristupa u odnosu na supervised metode je to što se pre-training i fine-tuning mogu sprovesti koristeći setove podataka iz istog domena (Slika 7.) [51].



Slika 7. Supervised vs SSL u medicinskom snimanju.

SSL algoritmi pokazuju izuzetan uspjeh u oblasti kompjuterskog vida nadmašujući po performansama supervised pre-training metode [52]. Pitanje koje ostaje

otvoreno tiče se ponovljivosti ovih rezultata u oblasti radiologije, s obzirom na razlike koje postoje između analize standardnih slika i analize rendgenskih snimaka.

Većina istraživača u oblasti SSL fokus stavljaju na razvoj pretext zadataka budući da je to ključna komponenta u SSL pristupu i dio u kojem se zapravo SSL odvija. U literaturi se može naići na različite kategorizacije SSL pristupa, ali najčešće se kategorizuju na osnovu arhitekture modela i ciljeva [50]. Liu i koautori [50] su SSL podijelili na generativne, kontrastivne i generativno-kontrastivne, dok su ih drugi autori klasifikovali kao prediktivne, generativne i kontrastivne [49] ili enkodere i enkodere-dekodere [53]. U ovom poglavlju, fokus je stavljen na generativnim i kontrastivnim modelima jer su oni najčešće korišćeni pristupi u rendgenskim modalitetima.

2.2.1. Generativni modeli

Generativni SSL pristup omogućava modelu da nauči latentne karakteristike iz neobilježenih podataka modeliranjem osnovne distribucije ulaznih podataka [49]. Tipična primjena je zadatak restauracije fotografija [54]. Osnovna arhitektura generativnih modela sastoji se od dvije komponente: enkodera i dekodera (Slika 8A.). Enkoder kompresuje ulazne podatke u latentni prostor (Z), dok dekodeer rekonstruiše komprimovani ulaz iz latentnog prostora. Dva najčešće korišćena generativna modela u rendgenskim modalitetima su auto-regresivni modeli i modeli auto-enkodera.

Auto-regresivne metode omogućavaju učenje smislenih reprezentacija iz slikovnih skupova podataka regeneracijom slike piksel po piksel. Nagoor i koautori [55] naglašavaju da je auto-regresivni model jedan od najstandardnijih modela za procjenu distribucije podataka i vjerovatnoće piksela. Auto-regresivni modeli se implementiraju u različitim arhitekturama kao što su PixelRNN [56], PixelCNN [56], PixelCNN++ [57] i Multiscale-PixelCNN [58]. Glavna prednost auto-regresivnih modela u kompjuterskom vidu je njihova sposobnost da efikasno uhvate zavisnost od konteksta, omogućavajući im da razumeju prostorne odnose i zavisnosti na slikama. Međutim, ograničenje ovih modela je njihova jednosmerna dostupnost kontekstu, što ometa njihovu sposobnost da inkorporiraju informacije sa budućih pozicija i uhvate dugoročne zavisnosti i globalne strukture u slikama [50].

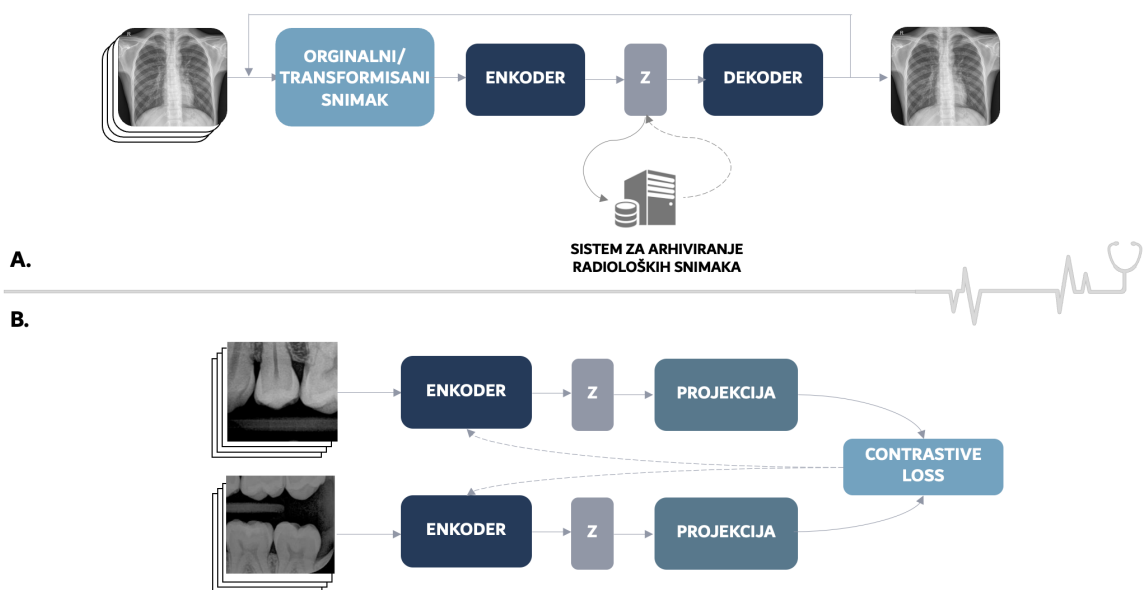
Auto-ekoder modeli mogu otkriti strukturu unutar podataka kako bi proizveli kompaktno predstavljanje ulaznih informacija [50]. Iako su auto-ekoderi prvi put korišćeni prije 40 godina, oni su i dalje najpopularniji generativni modeli zbog svoje fleksibilnosti i prilagodljivosti [59]. Uobičajno korišćene varijante opšte arhitekture auto-ekodera uključuju denoising autoekodere, konvolucione autoekodere i varijacione autoekodere [50]. Denoising autoekoderi uče reprezentacije otporne na šum tako što rekonstruišu izlaz bez šuma iz ulaznih podataka sa šumom (npr. Gausov šum, Poissonov šum, uniformni šum, impulsni šum, itd.) [50]. Međutim, denoising autoekoderi zanemaruju 2D strukturu slika, što dovodi do redundancije i globalnog prikaza karakteristika [60]. S druge strane, konvolucionni autoekoderi hvataju lokalizovane karakteristike dijeljenjem težina (weights) na svim ulaznim lokacijama, omogućavajući prostornu lokalizaciju i rekonstruišući slike pomoću linearne kombinacije osnovnih segmenata slike na osnovu latentnog koda [60]. Varijacioni autoekoderi kombinuju Bajesovo varijaciono zaključivanje sa dubokim učenjem da bi naučili distribuciju vjerovatnoće podataka putem amortizovanog varijacionog zaključivanja [61] i trika reparametarizacije.

2.2.2. Konstruktivni modeli

Konstruktivni modeli uče poredeći osnovni uzorak (anchor) sa pozitivnim i negativnim instancama dobijenim putem augmentacije (Slika 8B.). Metodologija kontrastivnog učenja je izuzetno intuitivna i odražava mehanizme učenja koji se mogu primijetiti kod djece [62]. U dječjoj igri matching [63], cilj kontrastivnog učenja je grupisati slične uzorke, dok se različiti uzorci drže na distanci. Framework-ovi kontrastivnog učenja se mogu podijeliti na različite načine [64]–[67], ali najčešća podjela je na kontraste između konteksta i instance (context-instance), kao i kontraste između instanci (instance-instance contrasts) [50].

Metode koje se zasnivaju na kontraste između konteksta i instanci (globalno-lokalni kontrast) fokusiraju se na modelovanje odnosa između lokalnih karakteristika uzorka i njegove globalne reprezentacije konteksta [50]. Dvije osnovne strategije globalno-lokalnog kontrasta su predviđanje relativne pozicije i maksimiziranje međusobne informacije. Prva kategorija naglašava relativni položaj lokalnih komponenti,

što neizbježno sugerije razumjevanju globalnog konteksta. Na primjer, predviđanje relativnog položaja dva organa unutar ljudskog tijela zahtijeva sveobuhvatno znanje o ukupnoj strukturi tijela. Shodno tome, naučnici su razvili nekoliko modela koji bi se mogli koristiti kao pretext zadaci. Ovi modeli rešavaju zadatke kao što su slagalica, predviđanje ugla rotacije i zadaci relativnog položaja [50], [65], [66]. S druge strane, druga strategija zanemaruje relativnu poziciju između lokalnih djelova i umjesto toga se fokusira na učenje odnosa pripadnosti između lokalnih djelova i globalnog konteksta. Ovaj koncept je izveden iz uzajamnih informacija u statistici [50]. U suštini, međusobne informacije procenjuju povezanost između dvije varijable koje su uzorkovane istovremeno. U pristupu globalno-lokalnog kontrasta, cilj je maksimizirati međusobne informacije. Korišćenje uzajamnih informacija u kontrastivnim zadacima učenja započeto je sa Deep Infomax [68] i dalje je evoluiralo uvođenjem novih modela kao što je Augmented Multiscale Deep InfoMax-AMDIM [69]. U međuvremenu, upotreba uzajamnih informacija nije rezervisana samo za globalno-lokalni kontrast, već se takođe primjenjuje u kontrastu između instanci.



Slika 8. SSL pristupi: a) generativni i b) kontrastivni modeli sematski dijagrami.

Metode kontrasta između instanci fokusiraju se na modelovanje odnosa između lokalnih reprezentacija na nivou instanci različitih uzoraka. Ove metode se mogu dalje podijeliti u dvije kategorije: diskriminacija klastera i diskriminacija instanci.

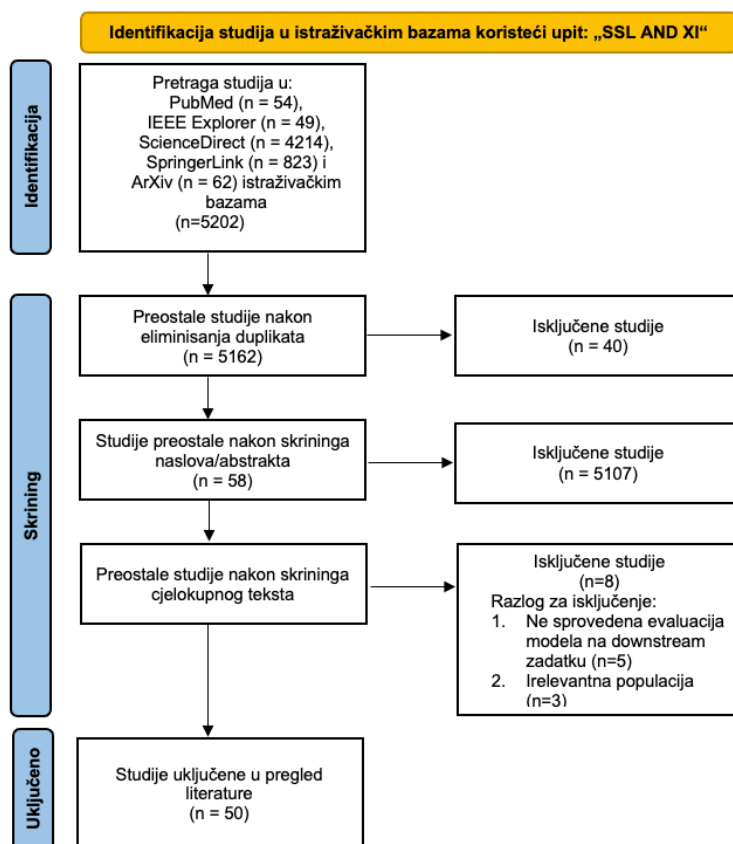
Početni uspjeh metoda kontrasta između instanci je demonstriran njihovom sposobnošću da postignu performanse uporedive sa AlexNet supervised modelom [50]. Pomenuti rezultati su postignuti korišćenjem tehnika zasnovanih na klasterovanju, kao što je Deep Cluster [70], koji koristi klasterovanje uzoraka za generisanje pseudo-labela (pseudo-oznaka). Napredna varijanta DeepCluster-a, poznata kao razmjena dodjela između višestrukih prikaza [71], dalje razvija ovaj koncept uvođenjem principa online klasterovanja i strategija za augmentaciju podataka sa više prikaza. Integracija augmentacije podataka u metodu razmjene dodjela između višestrukih prikaza inspirisana je pristupa zasnovanim na diskriminaciji instanci, koji koriste ove strategije kao zamjenu za vremenski intenzivan proces klasterovanja. Nedavno se razvoj metoda kontrasta između instanci uglavnom fokusirao na pristupe zasnovane na diskriminaciji instanci. Dobro poznate metode u ovom kontekstu uključuju CMC [72], MoCo [73], SimCLR [74], InfoMin [75], BIOL [76] i SimSiam [77]. Da bi se postigli kontrastni ciljevi učenja, koriste se tehnike augmentacije podataka za generisanje sličnih ili pozitivnih uzoraka. Međutim, generisanje i korišćenje različitih ili negativnih uzoraka variraju u zavisnosti od metode.

2.3. Primjene u rendgenskim modalitetima

U rendgenskim modalitetima, modeli bazirani na ML privukli su značajnu pažnju u procesu automatizacije analize rendgenskih snimaka. To je dovelo do razvoja sistema koji mogu obavljati različite ključne zadatke, kao što su segmentacija organa, lezija ili tumora, klasifikacija različitih bolesti ili stanja, detekcija abnormalnosti, praćenje napredovanja bolesti ili odgovora na liječenje, preprocessing-a rendgenskih snimaka i arhiviranje rendgenskih snimaka.

Pregled literature koji je sproveden u okviru ovog poglavlja realizovan je uz pomoć obimnih istraživačkih baza podataka uključujući Springer Link, ScienceDirect, IEEE Explore, ArXiv, and PubMed. Pretrage po ključnim riječima vršene su korišćenjem termina kao što su “medical imaging”, “deep learning”, “self-supervised learning”, “Computed tomography”, “X-ray imaging”, “self-supervised learning” i “X-ray imaging”. Budući da je fokus našeg istraživanja primjena SSL pristupa u rendgenskim modalitetima, najefikasnija strategija formulisanja upita prilikom pretrage baze podataka

podrazumijevala je korišćenje upita „self-supervised learning“ AND „X-ray imaging“. Pretraga sprovedena u martu 2023. godine ponudila je ukupno 5202 studije (PubMed=54; IEEE Explorer=49; ScienceDirect=4214; SpringerLink=823; ArXiv=62). Nakon uklanjanja studija koje se pojavljuju u dvije ili više baza (prvenstveno iz PubMed-a i arXiv-a) preostalo je 5162 studija. Nakon analize naslova i apstrakta isključeno je ukupno 5107 studija zbog irelevantnosti za cilj istraživanja. Zatim je uslijedio pregled cijelog teksta preostalih 58 studija, što je rezultiralo isključenjem dodatnih 8. Razlog za njihovo isključenje prvenstveno je bio zbog odsustva evaluacije predloženih modela na downstream zadatku, kao i zbog razvoja modela na podacima koji se ne odnose na ljudsku populaciju (skupovi podataka sakupljeni tokom dijagnostike različitih oboljenja kod životinja). Na slici 9. prikazana je sprovedena strategija koja je omogućila odabir relevantnih studija čiji sveobuhvatni pregled biće prikazan u sledećim poglavljima. Kako je od 2018. godine počeo da se pojavljuje značajno veći broj publikacija o primjeni SSL-a u rendgenskim modalitetima, za ovu sveobuhvatnu analizu uzeli smo u obzir literaturu počevši od 2018.



Slika 9. Prikaz procesa skrininga studija u vodećim naučnim bazama

2.3.1. Konvencionalna radiografija

Konvencionalna radiografija ima istorijski značaj u primjeni ML u radiologiji budući da je to jedan od prvih modaliteta u kojem je ML korišćeno da pomogne ljekarima u interpretaciji rendgenskih snimaka. SSL pristup je pronašao različite primjene u ovom modalitetu, pri čemu su glavne studije navedene u Tabeli 1.

Razvoj SSL algoritama namijenjenih automatskoj analizi rendgenskih snimka grudnog koša posebno je kompleksan zbog prirode bolesti/poremećaja, koje često zahtijevaju identifikaciju promjena (abnormalnosti) unutar ograničenog broja piksela [78]. U kontrastnim modelima, jedan od glavnih izazova je stvaranje pozitivnih parova snimaka kao što je naglašeno u MoCo-CXR studiji [78]. Ova studija razvila je specifičnu strategiju augmentacije podataka za rendgenske snimke grudnog koša, pošto tradicionalna augmentacija koja se koriste za standardne slike nije odgovarajuća zbog nedostatka smislenog prikaza na rendgenskom snimku grudnog koša. Korišćenjem nasumične rotacije (10 stepeni) i horizontalnog okretanja, ovaj pristup je stvorio unaprijed obučeni (pre-trained) model koji je generisao bolje reprezentacije i inicijalizacije za otkrivanje patologija. MoCo-CXR je pokazao da SSL modeli nadmašuju pre-trained modele obučene na ImageNet-u u postizanju reprezentacija višeg kvaliteta. Slični rezultati primijećeni su u slučaju modela C2L [79] i MUSCLE [80]. Dok je C2L isključivo koristio rendgenski skup podataka grudnog koša kao i MoCo-CXR, MUSCLE je razvijen korišćenjem šireg skupa podataka sa više regija koji obuhvata glavu, pluća i kosti. Drugačiji pristup je predložen u studiji [81], u kojoj su autori razvili SSL modele koji su koristili ImageNet skup podataka za pretext zadatak. Studija [82] pokušala je da kombinuje prednosti C2L, MoCo-CXR i [81] razvijajući pretext zadatak iz dvije faze. Prvo su koristili SimCLR na skupu podataka ImageNet. Za drugi korak, koristili su novi model nazvan kontrastivno učenje sa više instanci, dizajniran za rad sa podacima iz različitih projekcija (frontalnih i bočnih) u skupovima rendgenskih snimaka grudnog koša. Dodatno, studija [83] je pokazala da izveštaji radiologa mogu poboljšati pretraining za downstream zadatke na rendgenskim snimcima grudnog koša. Štaviše, multimodalni ReptsNet [84] model je prilagodio pre-trained modele kompjuterskog vida i jezičke modele za tumačenje rendgenskih snimaka i automatizovano generisanje izveštaja.

Napredak automatske analize rendgenskih snimaka posebno je ubrzan izbijanjem

Tabela 1. Primjena SSL pristupa u konvencionalnoj radiografiji

Ime modela / Arhitektura modela / Link za kod	Downstream zadatak	Performanse	Skup podataka	Ref.
SALAD / Auto-encoder / -	Detekcija anomalija	94.19% AUC (NIH), and 81.17% AUC (MURA)	NIH Chest X-rays i MURA	[85]
C2I / Arhitektura zasnovana na momentumu teacher-student / https://github.com/funnyzhou/C2L_MICCAI2020	Klasifikacija	89.3% AUROC (CheXpert)	ChestX-ray14, MIMIC-CXR, CheXpert, i MURA	[79]
- / Modified Models Genesis: U-oblikovana konvoluciona neuronska mreža u kombinaciji sa attention modulom zasnovanim na konvolucionim blokovima / -	Klasifikacija	98.6% prosječna tačnost	CXR	[86]
MoCo-CXR / Kontrastivno učenje: MoCo / -	Klasifikacija	(CheXper) MoCo-CXR 0.813 AUC, ImageNet 0.775 AUC. (Shenzhen) MoCo-CXR 0.974 AUC	CheXper, Shenzhen	[78]
- / Kontrastivno učenje. Sastoji se od tri dijela: glavne neuronske mreže, projekcionog sloja i stohastičkog modula za augmentaciju podataka / -	Klasifikacija rendgenskih snimaka grudnog koša	97.7% AUC	CheXper, ChestX-ray14, C19-Cohen, COVIDGR, Cell	[87]
MICLe / Kontrastivno učenje: prilagođeni SimCLR / -	Klasifikacija rendgenskih snimaka grudnog koša	0.7689 ± 0.0010 AUC, poboljšanje za 1.1% nadmašujući ImageNet pre-train	CheXpert	[82]
DVME / Kontrastivno učenje: SimCLR, SwAV, DINO; DVME- a tehnika nezavisna od modela za kombinovanje više unaprijed naučenih self-supervised learning karakteristika za downstream zadatke / -	Klasifikacija rendgenskih snimaka grudnog koša	DINO-0.6323 AUC (NIH Chest X-ray), SwAV - 0.5903 AUC. (APTOS), SwAV -0.6330 AUC, DVME 0.6566 AUC	NIH chest X-ray, Pneumonina chest X-ray; (APTOS, PatchCam)	[81]
- / Contrastive learning-MoCo / https://github.com/facebookresearch/CovidPrognosis	Predikcija nepovoljnih događaja iz pojedinačnih snimaka; Predikcija potrebe za kiseonikom iz pojedinačnih snimaka; i predikcija nepovoljnih događaja iz više snimaka	0.786 AUC za predviđanje i 0.848 AUC predviđanje mortaliteta	MIMIC-CXR, CheXpert, NYU COVID	[88]
- / BYOL [76] SimCLR [74], PixelPro [89] / -	Semantička segmentacija i detekcije objekata	Najbolje metode za četiri zadatka (BYOL za tri, a PixelPro za jedan zadatak)	MIMIC-CXR 2	[83]
MUSCLE / Sastoji se od dva dijela: učenja kontrastivne metode s momentom na više skupova podataka (Multi-Dataset MoCo) i kontinuiranog učenja više zadataka (Multi-Task Continual Learning)/-	Klasifikacija pneumonije, klasifikacija skeletnih abnormalnosti, segmentacija pluća i detekcija tuberkuloze.	99.72% AUC (klasifikacija pneumonije), i 88.37% AUC (Klasifikacija abnormalnosti skeleta)	NIHCC, China-Set-CXR, Montgomery-Set-CXR, Indiana-CXR, RSNA Bone Age; Pneumonia, MURA, Chest Xray Masks, TBX	[80]
RepsNet / Encoder-decoder model. Enkoder usklađuje snimka sa opisima na prirodnom jeziku putem kontrastivnog učenja, dok dekode predviđa odgovore na osnovu kodiranih snimaka/ https://sites.google.com/view/repsnet	VQA-Rad (vizuelno odgovaranje na pitanja) and IU-Xray (generisanje izvještaja)	81.08% tačnost klasifikacije na VQA-Rad 2018 i BLEU-1 rezultat od 0,58 na IU-Xray.	VQA-Rad, IU-XRay	[84]

Skraćenice: - model nije imenovan i/ili kod nije dostupan

pandemije COVID-19. Glavni razlog ubrzanog razvoja ogleda se u činjenici da je konvencionalna radiografija predstavljala prvu dijagnostičku proceduru za pacijente sa simptomima virusa. U cilju automatske analize rendgenskih snimaka pacijenata zaraženih virusom COVID-19 predloženo je niz aplikacija kao što su [86]–[88]. Među ovim studijama, studija [86] izdvaja se budući da poseban naglasak stavlja na problem overfitting-a. Overfitting je uobičajan izazov kada se primjenjuju modeli dubokog učenja u radiologiji, posebno u slučajevima rijetkih ili novih bolesti kao što je COVID-19, gdje je dostupni skup CXR podataka ograničen u poređenju sa drugim oboljenjima. Ograničenost skupa podataka utiče na generalizaciju modela, tj. mogućnost njegove primjene u svakodnevnoj kliničkoj praksi. Štaviše, razvijeni su specijalizovani modeli kao što je model SALAD [85], koji je posebno dizajniran da se pozabavi pitanjem overfitting-a u zadacima otkrivanja anomalija.

2.3.2. Kompjuterizovana tomografija (CT)

Kompjuterizovana tomografija (CT) igra značajnu ulogu u dijagnostikovanju mnogih oboljenja, što rezultira sa oko 300 miliona CT pregleda godišnje [90]. Analiza CT snimaka je zahtjevnija u poređenju sa snimcima konvencionalne radiografije, prije svega zbog obima slikovnih podataka, višeplanarne prirode CT-a i poboljšane rezolucije mekog tkiva [46], [91], [92]. Da bi se poboljšala tačnost i efikasnost donošenja odluka, razvijeni su algoritmi dubokog učenja. Međutim, zbog nedovoljne količine označenih podataka i razlika u domenu između ImageNet-a i CT snimaka, SSL pristup je poželjniji. Nedavne studije o primjeni SSL u CT su navedene u Tabeli 2.

Naučne studije su istakle potrebu za dubljim razumijevanjem CT snimaka prilikom kreiranja pretext zadataka. Na primer, primjena 2D neuronskih mreža na 3D medicinske snimke rezultirala je gubitkom ključnih 3D anatomskih informacija, što je dovelo do suboptimalnih performansi [47]. Da bi se prevazišao problem gubitka ključnih 3D anatomskih informacija, neke studije su predložile 3D SSL algoritme za volumetrijske CT snimke, kao što su Rubik's cube recovery [46], Models Genesis [47], Semantic Genesis [93], Rubik's cube++ [94], učenje 3D konteksta (3D context feature learning) [95] i Vol2Flow [96]. Performanse ovih pretext zadataka je potvrđena na downstream

Tabela 2. Primjena SSL pristupa u CT

Ime modela / Arhitektura modela / Link za kod	Downstream zadatak	Performanse	Skup podataka	Ref.
Rubik's cube recovery / Obuhvata dva operatora: cube rearrangement i cube rotation; i Siames - Octad mrežu / -	(1) klasifikacija krvarenja u mozgu i (2) segmentacija tumora mozga	(1) 83.8% tačnost (2) 76.2% mIoU koristeći U-Net i 77.3% mIoU koristeći 3D DUC	Brain Hemorrhage Dataset- interni skup podataka i BraTS-2018	[46]
Models Genesis / Enkoder-dekoder / https://github.com/MrGiovanni/ModelsGenesis	Klasifikacija i segmentacija (NCC, NCS, LCS, BMS)	NCC- 98.20±0.5% (AUC) NCS- 77.62±0.6% (IoU) LCS- 79.52±4.7% (IoU) BMS- 90.60±0.2% (IoU)	LIDC-IDRI, ChestX-ray8	[47]
- / REG2D i HEATMAP / -	Few-shot segmentacija CT	≈55% average Dice score (one-shot tačnost segmentacije)	VISCERAL Anatomy3	[95]
Semantic Genesis / Enkoder-dekoder, uvodi dvije nove komponente: self-discovery i self-classification / https://github.com/JLiangLab/SemanticGenesis	Klasifikacija i segmentacija (NCC, NCS, LCS, BMS)	NCC- 98.47±0.2% (AUC) NCS- 77.24±0.6% (IoU) LCS- 85.60±1.9% (IoU) BMS- 68.80±0.3% (IoU)	LUNA-2016, ChestX-ray14	[93]
- / Alternativne encoder-decoder CNN arhitekture: RS-AE, DE-UNET, RS-UNET / -	Rekonstruisati defekt lobanje uklonjen tokom dekompresivne kraniotomije izvedene nakon traumatske povrede mozga na postoperativnim CT snimcima	Predložena metoda direktne procjene zasnovana na U-Net arhitekturi (DE-UNET) nadmašuje sve ostale strategije	Skup podataka je obezbijedila Univerzitet u Cambridge (Divizija za estetiku, odjeljenje za medicinu)	[97]
- / Uključuje fazu prostornog transformatora koja uzima kao ulaz NN-generisane tačke i upoređuje rezultujuće poravnate slike / -	Automatsko pozicioniranje i detekcija tačaka	Fantom dataset-0.01% prosječan registration loss 2D dataset-0.1% testiranje tačnosti registracije	Shepp-Logan (S-L) fantom (sintetički) skup podataka, 3D CT snimci lobanje novorođenčadi, 2D snimci diatomeja četiri morfološke klase	[98]
Rubik's cube++ / Sastoji se od dvije komponente generator i diskriminator / -	(1) Segmentacija pankreasa, (2) segmentacija moždanog tkiva	(1) 4-struka unakrsna validacija je dala 100%-84.08% DSC (2) 77.56% DSC	NIH Kompjuterska tomografija pankreasa, MRBrainS18	[94]
- / Sastoji se od 3D enkodera, RE/SE modula, modula za augmentaciju i 3D Siamese mreže/ https://github.com/hongweilibran/imbanced-SSL	Klasifikacija: razlikovanje visokog stepena (H-grade) i niskog stepena tumora (L-grade), predviđanje stadijuma kancera pluća (npr. I, II ili III)	BraTS: Osjetljivost/specifičnost 0.920/0.711 Stadijum kancera pluća: Ukupna/takmičarska tačnost 0.538/0.372	Multicentrični MRI skup podataka (BraTS), skup podataka CT snimaka pluća sa 420 pacijenata sa nesitnoćelijskim karcinomom pluća (NSCLC-radiomics)	[48]
PCL / Kontrastivno učenje: U-Net enkoder sa projekcionom glavom/ https://github.com/dewenzeng/positional_cl	Segmentacija više organa	Postignut prosječni (standardna devijacija) rezultat na 5-strukoj unakrsnoj validaciji: CHD- 0.774 (.03) i ACDC- 0.929 (.00) za M=51. M- br. pacijenata korišćenih za fine-tuning.	CHD, MMWHS, ACDC, i HVSMR	[99]
FCL / Kontrastivno učenje: MoCo [73] / -	Volumetrijska segmentacija medicinskih snimaka	0.656±0.052 DIC za N=1, N- br. Označenih pacijenata (fine-tuning)	ACDC MICCAI 2017 izazov skup podataka	[100]

Skraćenice: - model neimenovan i/ili kod nije dostupan; NCC- Smanjenje lažnih pozitivnih nodula; NCS- Segmentacija plućnih nodula; LCS- Segmentacija jetre; BMS- Segmentacija tumora mozga

Tabela 2. nastavak

Ime modela / Arhitektura modela / Link za kod	Downstream zadatak	Performanse	Skup podataka	Ref.
SAME / Pixel-level kontrastivnog učenja, razbija registraciju snimaka u tri koraka: fina transformacija, gruba deformacija i duboka deformabilna registracija / -	(1) Intra-fazna registracija i (2) Inter-fazna registracija	(1) 54.42 % Dice skor, and (2) 50.96% Dice skor	Interni skup podataka	[101]
- / Varijacioni autoenkodori/ -	Detekcija nepravilno segmentiranih organa	0.92 AUC (bubreg), 0.95 AUC (jetra), 0.82 AUC (slezina)	Medicinski segmentacijski dekalon (jetra i slezina), interni skup podataka za segmentaciju bubrega, abdominalni CT skup podataka sa Univerziteta u Wisconsinu i bolnica i klinika	[102]
MoCo-COVID-19 / Kontrastivno učenje-MoCo / -	Few-shot klasifikacija - automatska dijagnostika COVID-19	0.931 ± 0.013 AUC	COVID-19 CT, skup podataka obezbijedeno od strane Italijanskog društva za medicinsku i intervencijsku radiologiju i prethodno obrađeno od strane MedSeg-a	[103]
CMT-CNN / Multi-task kontrastivno učenje	Klasifikacija COVID-19	CT (5.49%-6.45%) i X-ray (0.96%-2.42%)	Interni skup podataka datasets (CT and X-ray)	[104]
DeSD / Sastoji se od online studentske mreže i momentum učiteljske mreže/ https://github.com/veerwen/DeSD	Sedam zadataka segmentacije (jetra, bubregi, hepaV, pankreas, kolon, pluća, slezina)	Prosječan Dice skor od 55.6% za 10% anotacija i 75.8% prosječan Dice skor za 100% anotacija	DeepLesion; LiTS, KiTS, Hepatic Vessel (HepaV), Medical Segmentation Decathlon (MSD)	[105]
Vol2Flow / Arhitektura bazirana na 3D-UNet strukture (slično VoxelMorph [106]) / https://github.com/AdelehBitarafan/Vol2Flow	Segmentacija više organa	82.20±6.23 prosječan DSC	C4KC-KiTS, CT-LN, CT-Pancrea, Sliver07, CHAOS, 3Dircadb-01, 3Dircadb-02	[96]
PrepNet / Auto-ekoder / -	COVID-19 klasifikacija	0.5343 prosjek između skupova podataka	SARS-CoV-2, UCSD COVID-CT, MosMed skup podataka	[107]
MAE / Maskirani autoenkodori/ -	Segmentacija više organa na CT snimcima abdomena, segmentacija tumora mozga na snimcima magnetne rezonance, klasifikacija bolesti na rendgenskim snimcima toraksa	83.5Avg DSC, 78.91 Avg DSC, 81.5% mAUC	BTCV, BRATS, CXR14	[108]
3DFPN-HS² / Sastoji se od modula za rotaciju i mreže za predviđanje rotacije/ -	Detekcija plućnih nodula	Osjetljivost od 90.6% pri 1/8 lažnih pozitivnih po skenu na LUNA16 skupu podataka	LUNA16, SPIE-AAPM, LungTIME, i HMS	[109]
- / Sastoji se od multimodalnog modula za detekciju ključnih tačaka sa pažljivom fuzijom za 2D lokalizaciju zglobova pacijenta, kao i samostalno nadgledanog 3D regresionog modula za mreže / -	Automatizovano postavljanje izocentra sa kliničkim CT skenovima	Srednje greške od 5.3/7.5/8.1 mm za abdomen/toraks/glavu, u poređenju sa 13.2 mm medijan greškom radiografa	SLP, interni RGBD podaci	[110]

Skraćenice: - model nije imenovan i/ili kod nije dostupan

zadacima klasifikacije i segmentacije, što je dalo obećavajuće rezultate. Štaviše, određeni pretext zadaci posebno se bave drugim izazovima u domenu radioloških snimaka, kao što je Vol2Flow, koji poboljšava generalizaciju korišćenjem 3D radioloških snimaka dobijenih u različitim uslovima akvizicije.

Nedavno je kontrastivno učenje postalo dominantna SSL metoda zbog svojih superiornih performansi [111]. Da bi se iskoristio potencijal ove metode, razvijen je veliki broj algoritama koji se posebno bave izazovima rada sa CT snimcima, kao što su PCL [99], FCL [100], SAME [101], DeSD [105], MoCo-COVID-19 [103] i CMT-CNN [104]. Izazov kreiranja kontrastivnih algoritama učenja za medicinski domen je naglašen u PCL publikaciji. U ovoj studiji, autori su se fokusirali na pitanje pojave brojnih lažno negativnih parova koji su rezultat korišćenja najsavremenijih kontrastnih framework-ova, prvobitno dizajniranih za ne medicinski domen. Sa druge strane, publikacije DeSD i FCL nude rješenja za neke od izazova svojstvenih kontrastnim metodama prilagođenim posebno za medicinski domen. DeSD se bavi problemom slabog nadzora na početnim slojevima (shallow layer), što ima negativne efekte na performanse downstream zadataka. Nasuprot tome, FCL nudi novo rešenje za federated learning pristup, koje rješava probleme sa ograničenom količinom medicinskih podataka. Ovaj pristup omogućava dijeljenje reprezentacija na nivou snimka među različitim medicinskim institucijama, uz zadržavanje privatnosti izvornih podataka. Većina automatizovanih sistema za analizu CT snimaka koristi skupove podataka CT grudnog koša. Nedavno su se mnoge studije fokusirale na korišćenje ovih sistema za otkrivanje upale pluća ili COVID-19, kao što su MoCo-COVID-19, CMT-CNN, PrepNet [107]. Pored toga, postoje i druge SSL studije za specifične primjene kao što su dekompresivna kranijektomija [97], statistička analiza oblika [98], radiomika [48], detekcija nepravilno segmentiranih organa [102], detekcija plućnih nodula [109] i 3D modeliranje tijela pacijenta [110].

2.3.3. Mamografija

Skrining zasnovan na mamografiji, zajedno sa drugim skrining programima, igra ključnu ulogu u sistemima preventivne njege. U Sjedinjenim Državama, mamografija je značajno doprinela konstantnom padu stope smrtnosti od karcinoma dojke u posljednje

dvije decenije [112]. Štaviše, SSL algoritmi su pokazali obećavajuće performanse u zadatku predviđanja maligniteta, o čemu svedoči nekoliko studija navedenih u Tabeli 3.

Tabela 3. Primjena SSL pristupa u mamografiji

Ime modela / Arhitektura modela / Link za kod	Downstream zadatak	Performanse	Skup podataka	Ref.
MSVCL / Kontrastivno učenje + CycleGAN / https://github.com/lizheren/MSVCL_MICCAI2021	Detekcija lezija	0.792 mAP (poznati domen), 0.789 mAP (nepoznati domen)	Interni skup podataka od četiri proizvođača medicinske opreme (GE, United Imaging Healthcare, Hologic, and Siemens); INbreast	[113]
- / Jigsaw puzzle zadatak / -	Klasifikacija	0.958 AUC (Finopodešen na polovini ili četvrtini skupa za obuku)	Interni skup podataka (the US, the UK)	[108]
- / Kontrastivno učenje: contrastive loss lezije i normal contrastive loss / -	Klasifikacija	(Lezija) 58.2±0.055% (gustoća tkiva dojke) 78.1±0.029%	Interni skup podataka	[114]
- / SimCLR, BYOL, SWaV, ViT-MAEs / -	Klasifikacija	SWaV- 0.757 AUC (Linearna evaluacija) 0.815 AUC (Finetuning)	CBIS-DDSM, CMMD	[115]

Skraćenice: - model nije imenovan i/ili kod nije dostupan

Prilikom kreiranja SSL modela namijenjenih mamografiji takođe treba uzeti u obzir jedinstvene karakteristike mamografa. U studiji fokusiranoj na detekciju karcinoma na mamografskim snimcima [108], autori su istakli da tradicionalni SSL zadaci kao što su predviđanje ugla rotacije i predviđanje kolorizacije možda nisu pogodni za otkrivanje prisustva karcinoma na mamografskim snimcima [108]. Ovo ograničenje proizilazi iz činjenice da su tumori karcinoma dojke i druge abnormalnosti na mamografima obično malih razmjera, prikazani u sivim tonovima i nemaju definisanu orijentaciju. Kao alternativu, autori su predložili rješavanje jigsaw puzzle zadatka kako bi se generisale karakteristike koje su pogodnije za klasifikaciju maligniteta. Druge studije su pokušale da koriste SSL kontrastivne algoritme za predviđanje maligniteta [113]–[115]. Studija [115] je identifikovala optimalne transformacije za mamografe, uključujući nasumično kropovanje sa promjenom veličine originalnog snimka, gama pomjeranje, promjena

kontrasta i izjednačavanje histograma. Autori primjenjuju ove augmentacije na segmente (patches) nastale dijeljenjem mamografa sa grid pattern, umjesto na cijeli snimak, kako bi se poboljšala lokalizacija interesne oblasti (npr. lezije). Druge studije [113], [114] koriste kontrastivno učenje za rješavanje specifičnih izazova u skupovima podataka mamografije, kao što su razlike u domenu između skupova podataka iz različitih institucija, razlike u domenu između mamografa različitih proizvođača, lažno pozitivne stope i varijacije unutar klase.

2.3.4. Dentalna radiografija

Izraz „zubi su pola zdravlja“ na najbolji način pokazuje značaj oralnog zdravlja za naše cjelokupno zdravlje. Stomatolozi obično koriste radiološke procedure za dijagnostikovanje i liječenje stomatoloških problema. Duboko učenje je našlo široku primjenu u ovoj oblasti, dok SSL ima potencijal da dodatno unaprijedi performanse ML sistema. Međutim, vrijedi napomenuti da SSL ne nalazi široku primjenu u dentalnoj radiografiji, a ključne studije su navedene u Tabeli 4.

Tabela 4. Primjena SSL pristupa u dentalnoj radiografiji

Ime modela / Arhitektura modela	Downstream zadatak	Performanse	Skup podataka	Ref.
LCD-Net / Modifikovani MoCoV2. LCD (faza samostalnog pretreniranja i faza obuke dvo-grane mreže)	Segmentacija i klasifikacija: dijagnostikovanje tumora i ciste	LCD-Net 91.45% ACC (klasifikacija) LCD - 71.26 IoU (detekcija) LCD-Net 70.84 mIoU (segmentacija)	Interni skup podataka	[116]
- / SimCLR, BYOL, Barlow Twins	Klasifikacija karijesa	(Osetljivost) Barlow Twins -57.9%, SimCLR-57.2%, and BYOL-54.6%	Interni skup podataka	[117]

Skraćenice: - model nije imenovan

Studija [117] je predložila pristup zasnovan na kontrastivnim metodama učenja (SimCLR, BYOL i Barlow Twins) koji je omogućio poboljšanje performanse u zadatku klasifikacije karijesa. Značajna prednost ove studije leži u efikasnosti označavanja skupa podataka. Označavanje skupa podataka za SSL pristup trajalo je približno 71 radni sat, dok bi za označavanje skupa podataka za supervised learning pristup bilo potrebno preko 7600 radnih sati (ekvivalentno 950 radnih dana). Da bi dodatno poboljšali performanse modela, naučnici su definisali prikladnu strategiju augmentacije podataka zasnovanu na prirodi radiografija ugriza. Ova strategija augmentacije podataka uključivala je:

nasumično kropovanje veličine između 50–100% originalne veličine snimka, nasumično horizontalno okretanje sa 50% vjerovatnoće, podešavanja boje (vjerovatnoće): osvjetljenost (20%) i kontrast (10%) i zasićenost (10%) , i nasumična rotacija od -20° do 20°. Kontrastivne metode učenja su takođe korišćene kao pretraining modeli za otkrivanje tumora i cisti vilice. U studiji [116], autori su predložili metodu koja se bavi dvijema glavnim nedostacima postojećih modela: njihovo veliko oslanjanje na broj uzoraka lezija i nedostatak pouzdanosti u rezultatima dijagnostike. Kako bi prevazišli ove nedostatke, koristili su veliki broj zdravih uzoraka za pretraining modela i dizajnirali dvo-granu mrežu u kombinaciji sa strategijom augmentacije podataka zasnovanom na prekrivanju djelova snimka (patch-covering), uz gubitak konzistentnosti lokalizacije.

2.3.5. Multimodalne aplikacije

Multimodalni pristup u medicini odnosi se na sisteme bazirane na ML koji mogu da obrade i integrišu informacije iz različitih tipova i izvora podataka, kao što su radiološki snimci i klinički izveštaji. Primjenom ovog pristupa obezbjeđuje se sveobuhvatnije razumijevanje zdravstvenog stanja pacijenta. Literatura sugeriše da bi primjena multimodalnog pristupa mogla poboljšati performanse automatizovane analize radioloških snimaka [114]. Iz ovih razloga, ovaj holistički pristup danas se široko primjenjuje u svakodnevnoj kliničkoj praksi. Međutim, implementacija multimodalnog pristupa u automatskoj analizi radioloških snimaka može biti izazovna zbog ograničenja u pogledu privatnosti medicinskih podataka, visoke neravnoteže u klasama i visokim troškovima označavanja medicinskih podataka. SSL pruža priliku da se prevaziđu ova ograničenja, čineći primjenu multimodalnog pristupa izvodljivijom. Najnovije studije o multimodalnim aplikacijama i SSL-u navedene su u Tabeli 5.

Multimodalne aplikacije koje su prikazane u ovom potpoglavlju mogu se podijeliti u dvije kategorije: prva kategorija uključuje integraciju skupova radioloških snimaka sa različitih radioloških modaliteta, dok druga kategorija kombinuje skupove radioloških snimaka sa tekstualnim skupovima podataka. Najobimnija integracija podataka sa različitih radioloških modaliteta predstavljena je u studiji [112]. Ova studija je koristila trening skup podataka koji se sastoji od preko 100 miliona radioloških snimaka kreiranih koristeći sledeće dijagnostičke procedure: konvencionalna radiografija, CT,

magnetna rezonansa i ultrazvuk. Autori su predložili SSL metodu zasnovanu na kontrastivnom učenju i online klasterovanju karakteristika. Ova metoda je pokazala povećanje tačnosti i otpornosti na različite augmentacije snimaka, kao i ubrzano konvergiranje modela tokom obuke/treninga. Pored toga, rezultati studije su pokazali efikasnost predložene metode u poboljšanju performansi različitih downstream zadataka, kao što je klasifikacija/detekcija.

Tabela 5. Primjena SSL pristupa u multimodalnim aplikacijama

Ime modela / Arhitektura modela / Link za kod	Downstream zadatak	Performanse	Skup podataka	Ref.
- / Kontrastivno učenje: framework sa dvojnim enkoderima/ https://github.com/rwindsor1/biobank-self-supervised-alignment	Nadzorom nezavisna rigidna registracija multi-modalnih snimaka i segmentacija među modalitetima sa anotacijama suprotnog modaliteta	0.927 Dice kor	the UK Biobank [118]	[119]
HSSL / Kontrastivno učenje. Sastoji se od tri hijerarhijska nivoa: nivo slike, nivo zadatka, nivo grupe/ -	Zadaci segmentacije: srce, prostata, slezina	srce-87.65% Dice skor, prostata-68.58% Dice score, slezina-88.45% Dice skor	8 različitih izvora	[120]
SAR / Arhitektura enkodera-dekodera. Sastoji se od: Multi-scale cube generator, modula za transformaciju, enkodera, modula svesti o skali, dekodera i modula za modalitetsko-invarijantno adversarijalno učenje/ -	Segmentacija tumora mozga, segmentacija tumora pankreasa	BraTS2018- 84.92% prosječan Dice. MSD- 75.68% prosječan Dice (segmentacija pankreasa), 33.92% prosjek Dice (segmentacija tumora)	LUNA2016, LiTS2017, BraTS2018, MSD	[121]
CPRD / Sastoji se od tri specijalizovana učiteljska modela koji se fokusiraju na različite regije tijela, a zatim podučavaju studentski model da uči kako intra- i inter-regionalne karakteristike za Med-VQA/ https://github.com/awenbocc/cprd	Med-VQA (Medicinsko vizuelno odgovaranje na pitanja)	SLAKE-EN tačnost 81.2% and 83.4% za “open-ended” i “closed-ended” pitanja	SLAKE, SLAKE-EN, Medical Segmentation Decathlon.	[122]
M³AE / Maskirani autoenkodera. Sastoji se od različitih enkodera i dekodera za kopjutersku viziju i jezik/ https://github.com/zhjohnchan/M3AE	Medicinsko vizuelno odgovaranje na pitanja, klasifikacija medicinskih slika i teksta, pretraga medicinskih slika uz opise	87.82% tačnost za closed-ended” pitanja na SLACK, 78.50% tačnost on MELINDA, i najveća tačnost 66.65% na ROCO test setu	ROCO, MedICaT, VQA-RAD, SLAKE, VQA-2019, MELINDA	[123]
- / Online klastering - optimizacija zamijenjenih predikcija, B. hibridno self-supervised – supervised učenje/ -	Detekcija abnormalnosti na rendgenskim snimcima toraksa, detekcija metastaza na mozgu na magnetnoj rezonanci, detekcija hemoragija u mozgu na CT snimcima.	Detekcija lezija - 0.94 AUC, detekcija metastaza na mozgu - 0.932 AUC. 85% akceleracija konvergencije modela tokom obuke	Interni i javni skup podataka	[112]

Skraćenice: - model nije imenovan i/ili kod nije dostupan

U studijama [120], [121] se naglašava da iako je SSL intenzivno istraživano u zadacima klasifikacije i detekcije, njegova primjena u zadacima segmentacije ostaje ograničena. Kako bi prevazišli ova ograničenja, autori su predložili novi hijerarhijski SSL framework koji stiče znanje iz različitih zadataka segmentacije radioloških snimaka koristeći objedinjene skupove podataka iz više domena. Pored toga, još jedna studija [121] imala je za cilj da poboljša tačnost zadataka 3D segmentacije tumora korišćenjem SAR SSL metode na CT i skupovima podataka magnetne rezonance. Ova studija je usmjerena ka rješavanju izazova povezanih sa segmentacijom tumorskih regiona, koji proizilaze iz varijacija u razmjeri, incidenci i geometrijskim svojstvima tumorskih regiona. Pored toga, razvijen je pristup koji koristi SSL kada su dostupni snimci jednog pacijenta sa dva modaliteta [119]. U ovoj studiji, autori su predložili multimodalni kontrastni framework uparivanje snimaka na kojima se ne nalaze promjene. Pomenuti pristup koristi snimke cijelog tijela nastali korišćenjem dva rendgenska modaliteta: magnetne rezonance i dvoenergetske rendgenske apsorpciometrije. Ovaj pristup omogućava obuku modela za segmentaciju anatomskih regija na snimcima magnetne rezonance bez potrebe za označavanjem snimaka magnetne rezonance.

Druga kategorija multimodalnih aplikacija uključuje modele ML specijalizovane za rad sa radiološkim snimcima i kliničkim izveštajima. Ovi modeli imaju za cilj da steknu generalizovane reprezentacije iz obimnih skupova radioloških snimaka i tekstualnih podataka. Ovim pristupom se omogućava da se generalizovane reprezentacije mogu primijeniti na različite zadatke, kao što su medicinsko vizuelno odgovaranje na pitanja (visual question answering), klasifikacija radioloških snimaka i teksta, te pretraga radioloških snimaka i teksta. Jedan primjer takvog modela predstavljen je u studiji [123], i to kao multimodalni maskirni autokoder koji stiče znanje o domenu rekonstruišući nedostajuće piksele i tokene iz nasumično maskiranih snimaka i tekstova.

2.3.6. Ostale primjene

Kada se kreira rendgenski snimak on prolazi kroz niz koraka, uključujući preprocesing, interpretaciju i arhiviranje. SSL je pronašao primjenu u svim ovim fazama, a nedavne studije su istraživale na koji način pretext zadaci unapređuju preprocesing

tehnike i doprinose rješavanju izazova arhiviranja podataka. Najznačajnije studije su navedene u Tabeli 6.

Upotrebu rendgenskih modaliteta prate određeni rizici zbog same prirode rendgenskih zraka, koji mogu izazvati kratkoročne i dugoročne negativne efekte na pacijente i kliničko osoblje [124]. Noviji doprinosi smanjenju negativnih efekata nastalih korišćenjem rendgenskih modaliteta ogledaju se kroz digitalizaciju postupka akvizicije, kao i razvoj naprednih preprocesing metoda [125]. Jedan primjer je prikazan u studiji Zha i koautora [126], gdje su autori predložili SSL rješenje za rekonstrukciju sparse-view konusne kompjuterizovane tomografije. U ovoj studiji, smanjenje doze zračenja je postignuto smanjenjem projekcijske perspektive pri akviziciji cone beam CT. Dodatno, vrijeme izvršavanja modela je relativno kratko, u rasponu od 10 do 40 minuta, u zavisnosti od korišćenog skupa podataka. Zbirno, ovi rezultati sugerišu da ovaj model može biti primenljiv u kliničkoj praksi. Druga primjena preprocesing-a CT-a predložena je u studiji [127] za izvođenje interpolacije slajsova u CT-u. PixelMiner model koristi generativni pristup za preciznu rekonstrukciju tekstone, značajno poboljšavajući performanse downstream zadatka. Ova SSL metoda pokazuje zadovoljavajuću robusnost i generalizaciju, jer ostvaruje dobre rezultate ne samo na trening skupu podataka, već i na spoljnim validacionim skupovima podataka. Smanjenje doze zračenja neizbježno dovodi do smanjenja kvaliteta (stvaranje šuma) radioloških snimaka što u konačnici rezultira povećanjem potreba za razvojem što efikasnijih metodama za smanjenje šuma (denoising metodama). Generalno govoreći smanjenje šuma u biomedicinskim snimcima je izazovan zadatak jer šum direktno odgovara jačini signala i prati Poissonovu distribuciju. Studija [128] je predložila metodu Poisson2Sparse za denoising pojedinačnih snimaka bez podataka o stvarnom stanju (ground-truth data). Međutim, potreba za denoising-om snimaka je posebno važna za dinamičke metode snimanja, jer one koriste brze tehnike snimanja kako bi poboljšale vremensku rezoluciju, što može smanjiti odnos signal-šum u svakom frejmu. Jedna moguća primjena za denoising dinamičkih medicinskih snimaka predložena je u studiji [129]. Ovaj model, nazvan Deformed2Self, kombinuje denoising pojedinačnih slika i višestrukih slika kako bi poboljšao kvalitet snimka, a koristi i mrežu prostornog transformatora (spatial transformer network) za modeliranje kretanja između različitih slajsova.

Tabela 6. SSL pristup u fazi preprocessing i fazi arhiviranja

Ime modela / Arhitektura modela / Link za kod	Zadatak	Performanse	Skup podataka	Ref.
- / Autoenkoder / -	Komprimovanje medicinskih slika	Konvolucioni autoenkoder; gubitak je dostigao plato na $L = 0.0032$. Konvolucioni autoenkoder + PCA: kvalitet slike, $L' = 0.0036$. Asimetrični autoenkoder: u poređenju sa CAE+PCA1024, rezultati su očigledno lošiji iako je faktor kompresije isti	CRO Aviano: patch level skup podataka, full image skup podataka	[130]
Deformed2Self / Sastoji se od tri modula: denoising mreže jednog snimka; mreže za prostornu transformaciju (STN); denoising mreže za više snimaka / -	Dinamičko denoising snimaka	31.77% na PINCAT sa 15% standardnom devijacijom Gausovim šumom buke i 28.22% na ACDC sa istom standardnom devijacijom Gausovog šuma.	PINCAT, ACDC	[129]
PixelMiner / Model baziran na PixelCNN++. PixelMiner uključuje dosta karakteristika od PixelCNN++, Uključujući vertikalno i horizontalno slagane maskirane konvolucije, kontrolisane konvolucije i logističku mešovitu verovatnoću gubitka (Loss) / -	Interpolacija slajsova medicinskih snimaka	EPR 82% ($p < .01$), NRMSE of 0.11 ($p < .01$), (CCC) ≥ 0.85 ($p < .01$)	Izazov detekcije plućne embolije Radiološkog društva Severne Amerike ©	[127]
NAF / Sastoji se od četiri modula: uzorkovanje zraka, enkodiranje pozicije, predviđanje koeficijenta atenuacije i sinteza projekcija / -	Sparse-view CBCT rekonstrukcija (Cone Beam Computed Tomography)	PSNR/SSIM 33.05/.96 (toraks) 34.14/.94 (vilica) 31.63/.94 (stopalo) 34.45/.95 (abdomen) 30.34/.88 (aorta)	LIDC-IDRI, Otvorena naučna vizualizacija, interne slike fantoma	[126]
Poisson2Sparse / Pristup zasnovan na sparsitetu i učenju rečnika / https://github.com/tacalvin/Poisson2Sparse	Denoising biomedicinskih snimaka	Premašuju postojeće najsavremenije metode za SSL pristupe na različitim skupovima podataka i pri različitim nivoima Poisson šuma, na primjer, sa prosječnim poboljšanjem od približno ~ 2 dB	Skup podataka za denoising fluorescentne mikroskopije, PINCAT skup podataka	[128]
FedFew / federated SSL / -	Klasifikacija sa više oznaka	Tačnost 0.75 (edema), 0.73 (pneumonia), 0.77 (hernia)	ChestX-ray14	[131]
- / Autoenkoders / https://github.com/ciwarren/CCAE-MI	Arhiviranje snimaka	Odnos kompresije 32:1, do 96 procenata SSIM-a i MSE manji od 0.003	CT mozga, MRI mozga, COVID CT, chest X-ray	[132]

Skraćenice: - model nije imenovan i/ili kod nije dostupan

U prethodno opisanim aplikacijama, SSL pristup je prvenstveno razmatran kao dio preprocessing faze. Jedan od glavnih izazova u fazi arhiviranja je obezbeđivanje adekvatnog memorijskog prostora, jer radiološki snimci moraju biti arhivirani određeni

broj godina u skladu sa propisima zemlje u kojoj je snimak kreiran. Na primer, u Sjedinjenim Američkim Državama, radiološki snimci se moraju čuvati u periodu od 5 do 10 godina [132]. Tradicionalni komercijalni algoritmi kompresije nijesu prikladni za radiološke snimke budući da dovode do gubitka kvaliteta snimaka što je neprihvatljivo u dijagnostičkoj radiologiji [133]. Među trenutnim tehnikama za kompresiju snimaka, autoenkoder zasnovan na dubokom učenju pruža najbolje rezultate. Autoenkoderi se sastoje od mehanizama kompresije i dekompresije. Autoenkoderi kodiraju radiološki snimak u komprimovani prikaz i rekonstruišu snimak pomoću dekodera. Kao i druge metode kompresije, ovaj proces uzrokuje gubitak informacija. Nažalost, ne postoji jasna smjernica o prihvatljivom nivou gubitka u algoritmima kompresije. Institucije kao što je Uprava za hranu i lijekove (The Food and Drug Administration) samo savjetuje da gubitak kvaliteta snimka prilikom kompresije „*treba minimizirati što je više moguće*“ [132]. Studije [130], [132] identifikovale su metode zasnovane na autokoderu, posebno one zasnovane na konvolucionim autoenkoderima, kao najefikasnije za kompresiju radioloških snimaka. Ovi algoritmi kompresije za radiološke snimke postižu obećavajuće rezultate u smanjenju zahtjeva za resursima za obuku (treniranje) modela i interpretaciju snimaka. Joon i koautori [133] su za ovu svrhu koristili dvoslojni framework varijacionog autoenkodera.

2.4. Glavni benefiti i izazovi

Na osnovu pregleda literature, primjena SSL-a u radiologiji nudi značajnu prednost eliminišući potrebu za velikim, označenim skupovima podataka. Ovo je posebno značajno kada se radi sa 3D volumetrijskim radiološkim snimcima, budući da je označavanje ovih snimaka vremenski veoma zahtjevno čak i za iskusne radiologe [94]. Kako bi se prevazišao ovaj izazov, prethodne studije su uvele strategiju pre-treninga na skupu podataka ImageNet, koja ne samo da poboljšava tačnost već i ubrzava konvergenciju modela [79]. Međutim, značajno ograničenje proizilazi iz nepostojanja medicinskog skupa podataka sličnog ImageNet-u, što dovodi do očiglednog domenskog jaza između prirodnih slika i radioloških snimaka [79]. Na primjer, dok se modeli trenirani na skupu podataka ImageNet obučavaju na relativno uravnoteženim skupovima, oni imaju poteškoće u prilagođavanju „urođenoj“ neuravnoteženosti koja je često prisutna

u medicinskim podacima [48]. Ova razlika predstavlja značajnu prepreku pri pokušaju prenosa naučenih reprezentacija sa prirodnih slika na radiološke snimke, s obzirom na velike razlike u njihovim osnovnim distribucijama i karakteristikama. Uvođenjem SSL pristupa ovo ograničenje je prevaziđeno, jer se i pretext zadatak i downstream zadatak sprovode na istom domenu, omogućavajući bolje usklađivanje i iskorištavanje suštinskih svojstava radioloških snimaka. Pored toga, studije su pokazale da učenje SSL-a može pomoći u stvaranju medicinskog ImageNet-a brzim generisanjem početnih grubih oznaka za neoznačene snimke, koje zatim mogu da pregledaju radiolozi [47]. Štaviše, prednosti SSL-a su značajno veće od eliminacije označenih skupova podataka. Oni takođe doprinose poboljšanju robusnosti modela i generalizacije, što ih čini veoma primjenljivim u svakodnevnoj kliničkoj praksi. Na primer, u studiji [109] autori su poboljšali robusnost detekcije nodula u skupovima podataka prikupljenih sa različitih proizvođači CT skenera. Istraživači su takođe postigli poboljšanje robusnosti modela i generalizacije kroz primjenu federated learning, u čemu je SSL našao svoju primjenu.

Pored značajnih prednosti u odnosu na supervised learning pristup, vrijedi napomenuti da kreiranje SSL metoda predstavlja izazovan zadatak. U stvari, SSL pristupi se oslanjaju na heurističke metode pri dizajniranju pretext zadataka [104]. Ovo može biti posebno izazovno kada se primjeni na zadatke segmentacije [46]. U segmentaciji, iako se prethodno trenirani parametri obično mogu prilagoditi enkoderu, dekodeu i dalje zahtijeva nasumičnu inicijalizaciju. Ovo može predstavljati problem jer nasumična inicijalizacija dekodeu može narušiti prethodno naučenu reprezentaciju karakteristika i poništiti poboljšanja postignuta pre-treningom. Primjena metoda zasnovanih na kontrastnom učenju može biti nepraktična, posebno kod 3D skupova podataka, jer zahtijevaju velike veličine serija i/ili negativne parove. Ovo predstavlja izazov za medicinske skupove podataka jer različiti snimci mogu imati slične strukture ili organe, što rezultira velikim brojem lažno negativnih parova [99]. Iz tog razloga, naučnici često koriste Siamese mreže za prilagođavanje postojećih modela ili predlažu nove kontrastne mreže. U studiji [104], autori čak navode da modeli u kontrastivnom učenju obično uključuju više parametara od supervised metoda. Na primer, da bi se postigla uporediva top-1 tačnost, parametri SimCLR-a su 16 puta veći od ResNet-50. Štaviše, SSL se suočava sa sličnim izazovima kao i druge metode, kao što je „black box“ priroda i pristrasnosti (biases) u skupovima podataka [87].

2.5. Zaključak

Ovo poglavlje ponudilo je sveobuhvatan pregled primjene SSL pristupa u rendgenskim modalitetima. Opšti tok rada SSL-a uključuje pretext/proxy zadatak i downstream zadatak. Fokus je stavljen na pretext/proxy zadatku, jer je to zapravo mjesto gdje se SSL odvija. Različite kategorizacije se mogu primjeniti na SSL pristupe, u ovom poglavlju fokus je stavljen na generativne i kontrastne modele. Dok kontrastivni modeli pretežno nalaze svoju primjenu u fazi interpretacije rendgenskih snimaka, generativni modeli preuzimaju vodeću ulogu u preprocesing fazi i fazi arhiviranja. Studije su pokazale da formulisanje pretext zadataka zahtijeva duboko razumevanje inherentne prirode rendgenskih snimaka. U okviru kontrastnih modela, značajan izazov leži u generisanju pozitivnih parova slika, što dovodi do razvoja specifičnih strategija augmentacije podataka prilagođenih karakteristikama rendgenskih snimaka. Pored toga, SSL je pokazao izuzetan uspjeh u multimodalnim scenarijima, gdje integrisanje rendgenskih snimaka sa tekstualnim podacima ili korišćenje rendgenskih snimaka iz različitih modaliteta rezultira formiranjem kombinovanih skupova podataka. Ovi skupovi podataka služe kao osnova za izgradnju SSL modela koji zajedno imaju za cilj da poboljšaju performanse downstream zadataka. Štaviše, SSL pristupi mogu potencijalno da ublaže ograničenja supervised learning metoda. Ograničenja uključuju zahtjev za pristup velikim količinama označenih podataka, probleme povezane sa neuravnoteženi podacima, razmatranja robusnosti i poboljšanja generalizacije. Na osnovu pregleda literature jasno je da SSL pristup ima transformativni potencijal, eliminišući trenutnu situaciju u kojoj se razvija veliki broj sistema zasnovanih na ML, ali se mali broj njih primenjuje u svakodnevnoj kliničkoj praksi.

3. Oblast istraživanja

Pregled literature u prethodnom poglavlju obuhvatio je studije koje se odnose na primjenu SSL pristupa u rendgenskim modalitetima. Primjetno je da primjena SSL pristupa nije istražena u svim rendgenskim modalitetima. Za ovo doktorsko istraživanje od posebnog interesa je rendgenski modalitet fluoroskopija, budući da videofluoroscopska studija gutanja predstavlja najčešće korišćenu dijagnostičku proceduru u procjeni procesa gutanja. Identifikovali smo ključne oblasti istraživanja u cilju stvaranja mogućnosti za automatsku analizu fluoroskopskih rendgenskih snimaka. Sledeća podpoglavljja predstavljaju istraživačke oblasti doktorskog rada, kao i preduzete strategije u cilju njihove uspješne realizacije.

3.1. Segmentacija bolusa koristeći self-supervised learning pristup

3.1.1. Motivacija

Lokalizacija bolusa predstavlja jedan od ključnih parametara u procjeni poremećaja gutanja [1]. Klinička praksa pokazuje da su posebno kritični poremećaji tokom faringealne faze, jer mogu povećati rizik od ulaska djelova bolusa u disajne puteve. U zavisnosti od vremena i prirode ovih poremećaja, posljedice mogu varirati od blažih epizoda aspiracije do ozbiljnijih stanja koja zahtijevaju hitnu medicinsku intervenciju [134]. Lokalizacija bolusa u kliničkoj praksi se zasniva na videofluoroscopskoj studiji gutanja, ručnom analizom frejm po frejm. Ovakav način određivanja lokacije bolusa je dugotrajan i naučnici su u cilju automatizacije ovog procesa kreirali metode ML bazirane na supervised learning pristupu. Rezultati koje postižu ove metode su zadovoljavajuće, ali se karakterišu određenim nedostacima navedenim u pregledu literature. U cilju prevazilaženja nedostataka kao jedno od mogućih rješenja nameće se primjena SSL pristupa.

3.1.2. Strategija

Prvi korak podrazumijeva formiranje skupa podataka. Podaci korišćeni u ovom doktorskom istraživanju nastali su tokom svakodnevnih kliničkih pregleda procesa

gutanja. Nakon ove faze slijedi faza koja se odnosi na izradu ML modela baziranog na SSL pristupu. U okviru pomenute faze najzahtjevniji korak predstavlja kreiranje odgovarajućeg pretext zadatka. Predmet ovog doktorskog istraživanja su fluoroskopski rendgenski snimci, iz tog razloga pretext zadatak mora uzeti u obzir i prostornu i vremensku dimenziju unutar podataka. Znanje o skupu podataka stečeno tokom pretext faze se koristi za inicijalizaciju zadatka segmentacije bolusa, koji se oslanja na modelu prilagođenom radiološkim snimcima. Dodatno, pored razvoja SSL modela razmatra se mogućnost dodatnog unapređenja performansi zadatka automatske segmentacije bolusa primjenom inovativnih ensemble metoda.

3.2. Automatsko mjerenje faringealnog ostatka

3.2.1. Motivacija

Lokalizacija bolusa u videofluoroskopskoj studiji gutanja omogućava kliničarima, prije svega, da procijene parametre koji se odnose na sigurnost gutanja. Iako se u literaturi primjećuje da procjena sigurnosti gutanja predstavlja glavni pravac istraživanja ne treba zanemariti analizu efikasnosti gutanja. Poremećaji u efikasnosti gutanja se manifestuju nakon završetka procesa gutanja kroz sakupljanje ostatka bolusa u farinks. Postojanje faringealnog ostatka može rezultirati aspiracijom, ugrožavajući na taj način funkcionalnost respiratornog sistema. Istraživanja u pravcu automatske analize efikasnosti gutanja do sada se temelje na samo jednom objavljenom naučnom radu od strane kolega iz iMED Lab istraživačke grupe. Iz tog razloga istraživanje u okviru ovog doktorskog rada ima za cilj uspostavljanje nove istraživačke oblasti, koja će obuhvatiti automatsku analizu ključnih parametara efikasnosti gutanja.

3.2.2. Strategija

Slično kao u prethodnom istraživanju prvi korak podrazumijeva kreiranje skupa podataka fluoroskopskih snimaka. Važan dio istraživanja podrazumijeva i odabir kliničke metode za procjenu efikasnosti gutanja koja će primjenom ML biti automatizovana. Predloženo istraživanje neće zahtijevati razvoj novih ML modela, već će počivati na prilagođavanju jednog ili više modela razvijenih u okviru istraživačke grupe iMED lab.

4. Segmentacija bolusa koristeći self-supervised learning pristup

Glavni doprinosi prezentovani u ovom poglavlju su sastavni dio istraživačkog rada Martinović I., Dousty M., Li W., Đukanović M., Coyle L. J., Sejdić E. (2025), *Leveraging Ensemble on Self-Supervised Techniques to Enhance Model Performance: A Regime Analysis*.

Disfagija se manifestuje kroz širok spektar uticaja na ljudsko zdravlje, počevši od blagih simptoma kao što je kašalj pa sve do težih kao što su opstrukcija disajnih puteva ili upala pluća (pneumonia). Pomenuti simptomi su uglavnom prouzrokovani penetracijom sažvakane hrane ili tečnosti, koja je naziva bolus, u disajne puteve. Tradicionalni način analize flurosopskih snimaka, frejm po frejm, u cilju lokalizacije bolusa veoma je dugotrajan. Iz tog razloga naučnici su otpočeli istraživanja u pravcu razvoja algoritama koji treba da omoguće automatsku segmentaciju bolusa. U cilju stvaranja pomenutih algoritama naučnici/istraživači se oslanjaju na supervised learning pristup. Međutim, ovo doktorsko istraživanje pravi otklon od pomenutog pristupa i usmjerava se ka razvoju sistema za automatsku segmentaciju bolusa koristeći SSL pristup. Predložena metoda oslanja se na Contrastive Random Walk model u pretext zadatku i U-Net++ modelu u downstream zadatku. Konkretno kao backbone mreža korišćena je ResNet-18 kako bi se naučene karakteristike u pretext zadatku iskoristili u downstream zadatku. Postignuti rezultati jasno ukazuju da se korišćenjem SSL pristupa može nadmašiti modeli bazirani na supervised learning pristup. Dodatno, koristeći inovativnu self-supervised learning weighted ensemble strategiju, kombinovanjem od jednog do osam top modela, F1-score U-Net++ modela se poveća sa 79.1% na 81.8% u poređenju sa ImageNet inicijalizacijom. Predloženi pristup u ovom doktorskom istraživanju predstavlja pionirski poduhvat u primjeni i istraživanju SSL pristupa u rendgenskom modalitetu fluoroskopije, sa ciljem da se unaprijedi proces procjene poremećaja gutanja.

4.1. Uvod

Jedan od ključnih i vremenski najzahtjevnijih zadataka u videofluoroskopskoj procjeni gutanja predstavlja lokalizacija bolusa, prije svega jer kompresovani bolus mijenja konfiguraciju i distribuciju tokom njegovog faringealnog puta. Studije pokazuju da metode za automatsku lokalizaciju bolusa, kao što su metode segmentacije zasnovane na dubokom učenju, mogu značajno da povećaju brzinu i tačnost procesa donošenja odluka u kliničkim uslovima [26], [30]. Pomenute metode su pokazale impresivne performanse oslanjajući se na supervised learning pristup, koristeći metode dubokog učenja kao što su U-Net i Mask-RCNN. Međutim, poređenje performansi ovih metoda predstavlja značajan izazov budući da su prilikom njihovog kreiranja korišćeni različiti skupovi podataka. Opšte je poznato da je prikupljanje velikih, označenih skupova podataka ključno za razvoj metoda ML. Ovo postaje posebno problematično u domenu medicine, tj. radiologije, gdje npr. postoje rijetke bolesti za koje ne postoji dovoljno podataka. Dodatno, u slučaju da nijesu u pitanje rijetke bolesti, tj. da imamo odgovarajući skup podataka neophodno je isti označiti i to isključivo od strane eksperta, radiologa [44]. Značajan problem u kreiranju adekvatnog skupa podataka predstavlja i činjenica da podaci sakupljeni isključivo od određenog proizvođača medicinske opreme ili određene zdravstvene institucije često unose pristrasnost u skupove podataka [135], [136]. Shodno tome, mnogi istraživači kao primarne nedostatke svojih metoda navode nedovoljnu robusnost i generalizaciju, što u značajnoj mjeri onemogućava primjenu ovih metoda u svakodnevnoj kliničkoj praksi [32]. Dodatno, ograničenje metoda dubokog učenja zasnovanih na supervised learning pristupu predstavlja neusaglašenost domena skupa podataka korišćenog za pretraining (ImageNet) i skupa podataka korišćenog za downstream zadatak (medicinski domen).

U cilju prevazilaženja gore navedenih izazova kao moguće rješenje nametnulo se korišćenje SSL pristup [137]–[140]. Primjena pomenutog pristupa u zadatku segmentacije bolusa još uvijek nije istražena. Prema tome, osnovni cilj ovog istraživanja je razvoj pouzdanog SSL modela za automatsku segmentaciju bolusa omogućavajući unapređenje performansi postignutih korišćenjem supervised learning pristupa. Predloženi metod se sastoji od Contrastive Random Walk modela u pretext zadatku i U-Net++ modela u downstream zadatku. Dodatno, postavili smo hipotezu da se primjenom

SSL ensemble strategije postignute performanse koristeći SSL model mogu još dodatno poboljšati. Svi naši naponi usmjereni su ka razvoju automatizovane metode za precizno lociranje bolusa u svakom frejmu fluoroskopskog snimka, sa tačnošću uporedivom manuelnim metodama lokalizacije bolusa.

4.2. Materijali i Metode

4.2.1. Kreiranje skupa podataka

Skup podataka korišćen u ovom istraživanju kreiran je u Presbyterian univerzitetskoj bolnici Medicinskog centra Univerziteta u Pitsburgu i sadži videofluoroskopske podatke iz 270 studija gutanja od 28 pacijenata sa sumnjom na disfagiju (Tabela 7). Protokol studije je odobrio Etički komitet Univerziteta u Pitsburgu (PRO19040040), a svi učesnici su dali pristanak da učestvuju u studiji. Tokom pregleda pacijentima su date instrukcije da gutaju tečnosti različite konzistencije i zapremine, kao i pasiranu hranu i kolače. Količinu i viskozitet supstanci datih pacijentima određivao je kliničar na osnovu kliničke istorije pacijenta i kliničkih indikacija. Tečni bolusi sa zapreminama u rasponu od 3-5 ml uzimani su pomoću kašike, a boluse koje su, pacijenti, mogli sami da uzmu davani su iz šolje zapremine od 10-20 ml. Fluoroskopija je podešena na 30 impulsa u sekundi, a snimci su dobijeni pri 60 ili više frejmova u sekundi. Snimci su snimljeni pomoću video kartice i softvera LabVIEW i konvertovani u dvodimenzionalne digitalne video snimke rezolucije 720×1080 . Broj frejmova u video snimci je smanjen na 30 frejmova u sekundi da bi se eliminisali duplirani frejmovi. Studija je pratila relevantne propise i smjernice.

Tabela 7. Demografski podaci pacijenata čiji su podaci uključeni u skup podataka

<i>Skup podataka</i>	<i>br. gutanja</i>	<i>br. pacijenata</i>	<i>ženski pol</i>	<i>godine (ženski pol)</i>	<i>Muški pol</i>	<i>godine (muški pol)</i>	<i>kriterijum za isključenje</i>
<i>Patients cohort</i>	270	28	13	57.08 ± 17.21	15	59.87 ± 14.88	<i>Traheostomija, prethodna operacija glave ili vrata, trudnoća i oni koji nisu u stanju da prate komande.</i>

4.2.2. Self-supervised pretraining

Uobičajna pretraining strategija za razvoj ML sistema namijenjenih za primjenu u radiologiji zasniva se na ImageNet supervised learning metodama. Međutim, glavno ograničenje ovakvog pristupa ogleda se u domenskom jazu između pretraining skupa podataka i skupa podataka downstream zadatka. Da bi se premostio ovaj jaz, kao jedno od mogućih rješenja nameće se primjena SSL pristupa, koji omogućava korišćenje neoznačenih skupova podataka iz medicinskog domena u pretraining fazi.

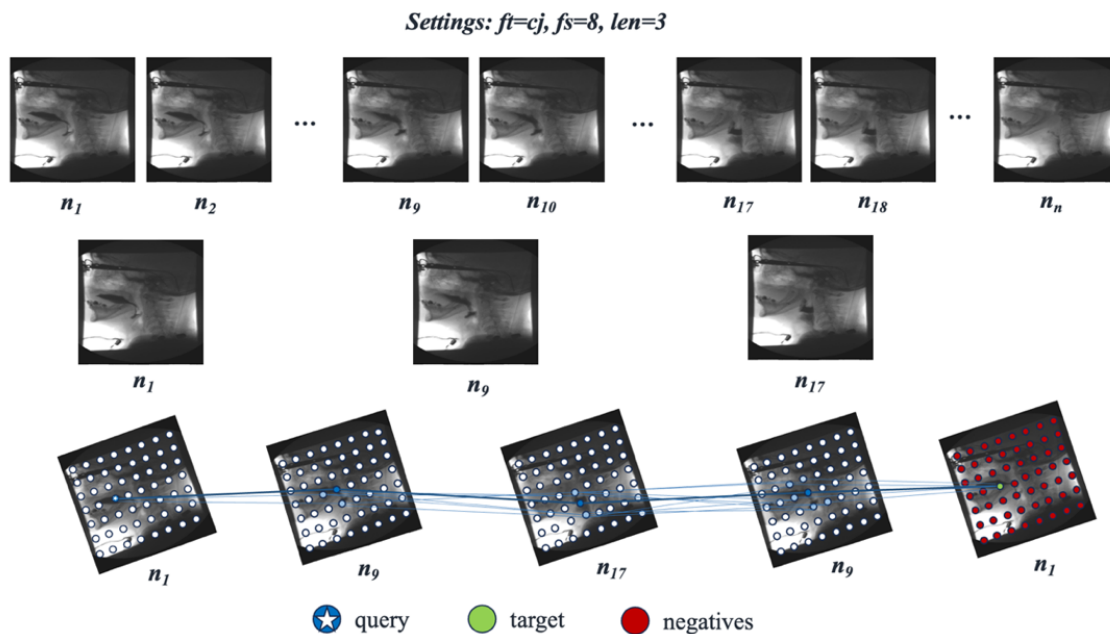
Odabir odgovarajućeg zadatka u pretraining fazi, takođe poznatog kao pretext zadatak, predstavlja jednu od najizazovnijih faza u razvoju SSL pristupa. Brojne istraživačke studije su naglasile važnost dubokog razumevanja podataka prilikom izrade pretext zadatka [137].

U ovom istraživanju fokus je stavljen na skup podataka fluoroskopije koji sadrži ograničeni skup rendgenskih snimaka snimljenih u vremenskim sekvencama. S obzirom na ovu karakteristiku, izabrani pretext zadatak mora uzeti u obzir i prostornu i vremensku dimenziju unutar podataka. Shodno tome, odabran je pretext zadatak koji omogućava učenje reprezentacija vizuelne korespondencije iz sirovog video zapisa, poznatog kao Space-Time Correspondence, koristeći Contrastive Random Walk model [141].

Model *Contrastive Random Walk* se može opisati kao prediktor veza u prostorno-vremenskom grafu konstruisanom iz video zapisa. Konkretno, video zapis je predstavljen kao graf gdje su čvorovi djelovi slike (image patches), a ivice predstavljaju sličnosti između čvorova susjednih frejmova [141]. Primarni cilj ove metode je da nauči karakteristike na način da su vremenske korespondencije predstavljene jakim ivicama. Model je obučen da se kreće kroz graf koristeći random walk princip i to između query i target čvorova. Ovo učenje se odvija bez potrebe za označenim podacima, jer se model trenira na palindromskoj sekvenci gdje se navigacija odvija od frejma n do $n + k$, a zatim nazad do n , pri čemu se početni čvor koristi kao target (cilj). Eksperimentalni rezultati pokazuju da reprezentacije modela na ne medicinskim skupovima podataka nadmašuju performanse supervised learning metoda u oblastima kao što su segmentacija objekata na video snimcima, propagacija ključnih tačaka položaja i propagacija semantičkih dijelova.

U toku ovog istraživanja sproveden je trening modela Contrastive Random Walk tokom 200 epoha koristeći istu arhitekturu kao u originalnoj implementaciji, gdje je ResNet-18 služio kao backbone mreža. Primarna funkcija backbone mreže je da izdvoji

karakteristike iz ulaznog snimka i napravi mape karakteristika na različitim nivoima rezolucije. Prikupljeni skup podataka je razdvojen po principu jedan pregled jedan video. Pored toga, prije treninga, veličina frejmova je prilagođena rezoluciji od 256x256, obezbjeđujući da je svaki frejm podijeljen na 49 čvorova (veličine 64x64). Proces obuke obuhvatao je više od 300 različitih konfiguracija, mijenjajući pet hiperparametara, uključujući brzinu učenja (learning rate), transformaciju frejmova, preskakanje između frejmova, broj frejmova po klipu i maksimalan broj klipova po videu. Na osnovu ovih hiperparametara može se zaključiti da je proces obuke podrazumijevao učenje reprezentacije iz određenog broja video klipova. Ovi klipovi su imali različite dužine i generisani su od transformisanih frejmova sa definisanim intervalom preskakanja (Slika 10.).



Slika 10. Contrastive Random Walk model

Skraćenice: ft-transformacija frejma, fs-preskakanje između frejmova, len-dužina frejmova po klipu, cj-varijacija boja.

4.2.3. Downstream zadatak

Downstream zadatak predstavlja centralni dio ML sistemima. Konkretno u ovom istraživačkom radu downstream zadatak treba da ponudi odgovor na pitanje „*Koje oblasti na frejmovima čine bolus i gdje su locirane*“. Semantička segmentacija se nameće kao logičan izbor za downstream zadatak i to ona koja je prilagođena radiološkim snimcima,

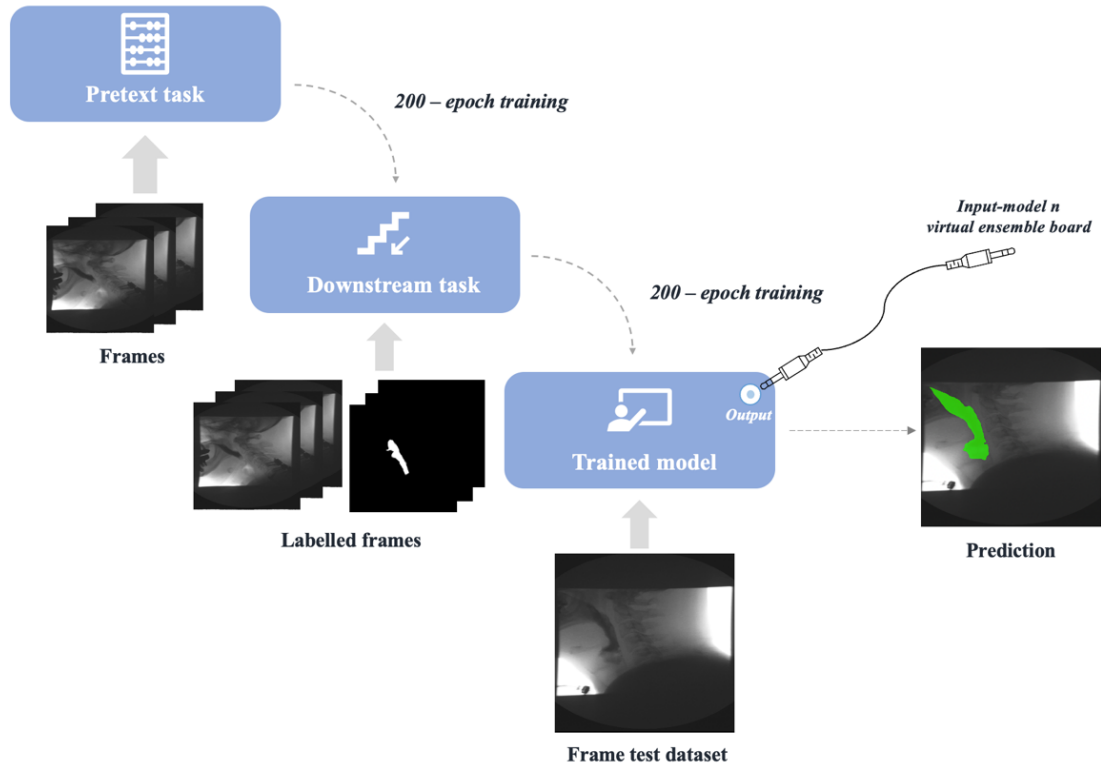
a koja počiva na korišćenju nadograđenog U-Net [142] modela pod nazivom U-Net++ [143]. U-Net++ predstavlja moćnu duboku supervised enkoder-dekoder mrežu, gdje su podmreže enkodera i dekodera povezane nizom ugniježdenih i gustih (nested, dense) skip puteva. Skup podataka fluoroskopije sastoji se od frejmova i maski za segmentaciju. Maske za segmentaciju su ručno kreirali kliničari. Nakon toga, podaci su nasumično podijeljeni po pacijentima u tri grupe: 70% pacijenata je dodijeljeno skupu podataka za obuku, koji sadrži približno 4.000 frejmova sa maskama; 20% pacijenata je formiralo skup podataka za validaciju, koji se sastoji od oko 1.000 frejmova sa maskama; a preostalih 10% pacijenata čini testni skup podataka, koji obuhvata otprilike 1.200 frejmova. ResNet-18 predstavlja backbone mrežu u zadatku semantičke segmentacije, koja je prethodno obučena na pretext zadatku (Slika 11.). Rezolucija ulaznog snimka je 256x256. Odabrana rezolucija snimka obezbjeđuje dobre performanse. Vrijedi napomenuti da bi veća rezolucija snimka potencijalno mogla da pruži još bolje rezultate hvatanjem više detalja. Međutim, važno je imati na umu da povećanje rezolucije snimka dovodi do većeg zahtjeva za računarskim resursima, što smo u ovom istraživačkom radu pokušali da minimizujemo.

Evaluacija zadatka segmentacije bolusa sprovodi se koristeći F1 score, takođe poznat kao Dice koeficijent. F1 score je široko korišćena metrika za evaluaciju zadataka klasifikacije, čak i kada se radi o neuravnoteženim skupovima podataka. Zadatak segmentacije bolusa uključuje binarnu klasifikaciju na nivou piksela, praveći razliku između pozadinskih i bolus piksela.

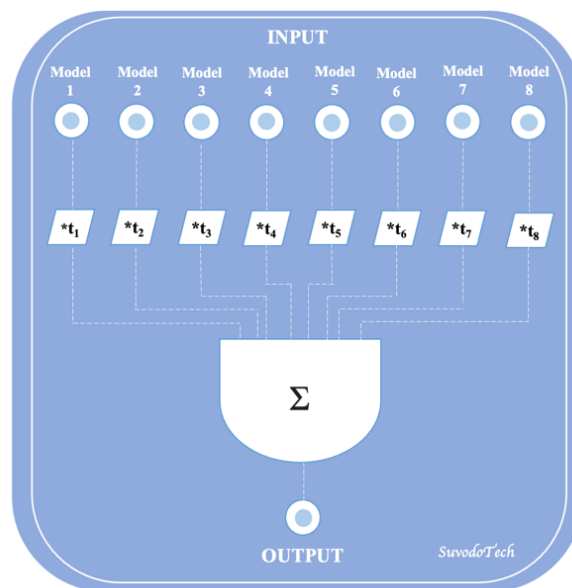
4.2.4. Ensemble model strategija

Dodatno poboljšanje performansi postignuto je primjenom weighted average ensemble metode (Slika 12.). Motivacija za korišćenjem ove metode proističe iz studije [144] koja pokazuje da bi model semantičke segmentacije zasnovan na weighted average ensemble metodi mogao nadmašiti standardni U-Net model. Međutim, u ovoj studiji se navodi da određivanje relativnih weights za svakog ensemble člana predstavlja izazov. U ovom doktorskom istraživanju korišćen je grid search metoda za određivanje odgovarajućih weights ($t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8$) za U-Net++ modele sa različitim SSL

inicijalizacijama. Ove weights su uzimale vrijednost u opsegu od 0 do 1, uz pridržavanje ograničenja $\sum_{i=1}^{i=8} t_i = 1$.



Slika 11. Schema zadatka segmentacije bolusa baziranog na SSL pristupu



Slika 12. Virtuelna ploča weighted average ensemble modela

4.2.5. Statistička analiza

U cilju utvrđivanja praktične razlike između primjene SSL i supervised learning pristupa u zadatku segmentacije bolusa, koristili smo Wilcoxon's signed rank test [145], [146]. Korišćena je cross-validation sa 9 različitih nasumično generisanih podjela (splits), pri čemu se svaka podjela (split) sastojala od 70% podskupa za obuku, 20% podskupa za validaciju i 10% podskupa za testiranje.

4.3. Rezultati

Rezultati zadatka segmentacije bolusa na osnovu različitih vrijednosti hiperparametara u okviru pretext zadatka prikazani su u Tabeli 8. Upotreba supervised learning pristupa rezultirala je preciznošću od 0.786 ± 0.0050 F1 score. Međutim, dodatno povećanje performansi supervised learning modela čak i za 1% postaje sve izazovnije kako se približava gornjim granicama performansi. Sprovedeno istraživanje pokazalo je da SSL pristup, pored rješavanja izazova prisutnog kod primjene supervised learning pristupa, ima potencijal da dodatno poboljša performanse zadatka segmentacije bolusa. Ovu tvrdnju potvrđuje povećanje od 0.016 F1 score koristeći SSL pristup u poređenju sa baseline modelom koji se oslanja na ImageNet inicijalizacije. Tabela 8. pokazuje da je pretraining model sa stopom učenja (learning rate) od 0,001, transformacijom frejmova sa color jitter, preskakanjem od 8 frejmova, 3 frejma po klipu i maksimalno 3 klipa po video zapisu postigao najbolje performanse. Kako bi se dodatno poboljšale performanse modela, u ovom doktorskom istraživanju predložena je SSL ensemble strategija. Performanse strategije SSL ensemble, gdje su za članove ansambla odabrana top osam SSL modela, predstavljeni su na Grafikonu 1. Kao što se može vidjeti rezultati pokazuju da korišćenje šest SSL modela sa najboljim performansama omogućavaju poboljšanje performansi modela za 0.008. Pored toga, sproveli smo studiju ablacije kako bismo uporedili ensemble strategije između SSL i supervised learning pristupa. Rezultati pokazuju da primjena ensemble strategija kod supervised learning pristupa donosi poboljšanje performansi za 0,01 koristeći četiri ili pet modela sa najboljim performansama. Međutim, postignuto unapređenje ipak nije dovoljno da nadmaši performanse postignute SSL pristupom. Prethodno pomenuti rezultati su se odnosili na podijeljeni skup podataka koji pokazuje najbolje performanse u zadatku segmentacije

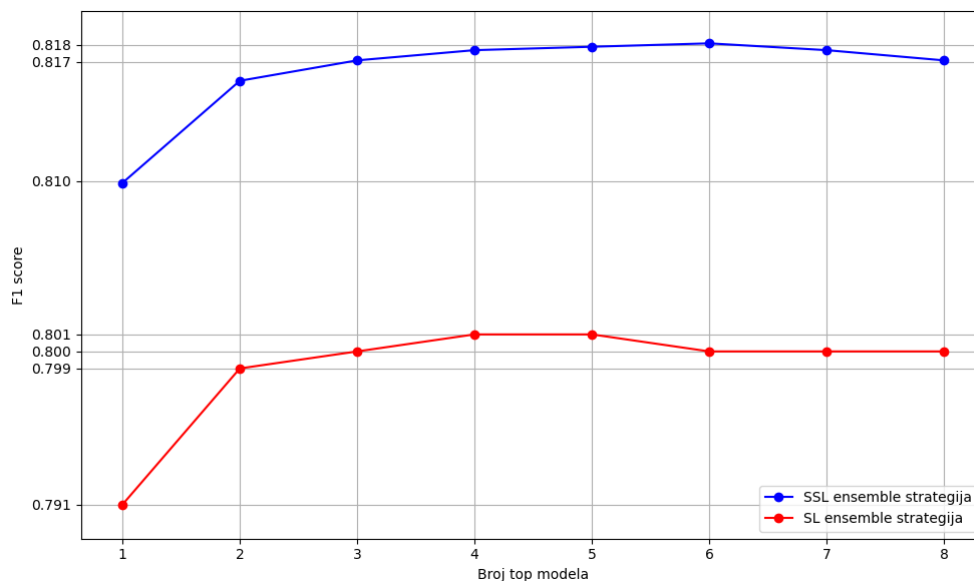
Tabela 8. Performanse zadatka segmentacije bolusa na osnovu različitih vrijednosti hiperparametara u okviru pretext zadatka (F1 score)

		<i>ft</i> <i>color jitter</i>						<i>ft</i> <i>flip</i>					
		<i>fs</i> 6		<i>Fs</i> 7		<i>fs</i> 8		<i>fs</i> 6		<i>fs</i> 7		<i>fs</i> 8	
		<i>cpv</i> 2	<i>cpv</i> 3	<i>cpv</i> 2	<i>cpv</i> 3	<i>cpv</i> 2	<i>cpv</i> 3	<i>cpv</i> 2	<i>cpv</i> 3	<i>cpv</i> 2	<i>cpv</i> 3	<i>cpv</i> 2	<i>cpv</i> 3
<i>Lr</i> 0.001	<i>Le</i> <i>n</i> 3	0.78 5	0.78 7	0.79 5	0.76 8	0.79 1	0.81 0	0.77 6	0.79 4	0.77 9	0.78 3	0.77 2	0.78 2
	<i>Le</i> <i>n</i> 4	0.78 5	0.77 9	0.78 7	0.78 3	0.76 0	0.78 5	0.80 4	0.77 6	0.77 4	0.78 4	0.77 5	0.79 1
	<i>Le</i> <i>n</i> 6	0.80 3	0.77 9	0.79 1	0.75 7	0.78 6	0.78 5	0.78 7	0.80 0	0.78 5	0.78 4	0.79 4	0.78 9
<i>Lr</i> 0.000 1	<i>Le</i> <i>n</i> 3	0.78 7	0.77 3	0.77 0	0.74 0	0.77 8	0.79 6	0.78 1	0.79 4	0.79 2	0.78 0	0.80 2	0.77 5
	<i>Le</i> <i>n</i> 4	0.79 2	0.80 4	0.76 8	0.79 5	0.77 2	0.79 2	0.77 6	0.78 8	0.77 8	0.77 3	0.78 0	0.78 2
	<i>Le</i> <i>n</i> 6	0.77 9	0.78 9	0.79 7	0.78 8	0.77 7	0.78 6	0.79 2	0.77 6	0.79 2	0.79 8	0.77 6	0.78 8

Skraćenice: lr-stopa učenja (learning rate), ft-transformacija frejmova (frame transformation), fs-preskakanje između frejmova (skip between frames), len-broj frejmova po klip (number of frames per clip), cpv- maksimalan broj klipova po videu (maximum number of clip per video), cj- varijacija boja (color jitter)

bolusa. Međutim, statistička analiza između supervised learning i SSL pristupa uključila je podatke dobijene obučavanjem/treniranjem modela na 9 nasumičnih podjela podataka (Tabela 9.). Za SSL pristup, rezultati su kreirani korišćenjem identičnih parametara (weights) inicijalizacije za sve podjele, kreiranih u toku pretext zadatka a koji su identifikovani u Tabeli 8. i to onih sa najboljim F1 score od 0,81. Potvrđivanjem statistički

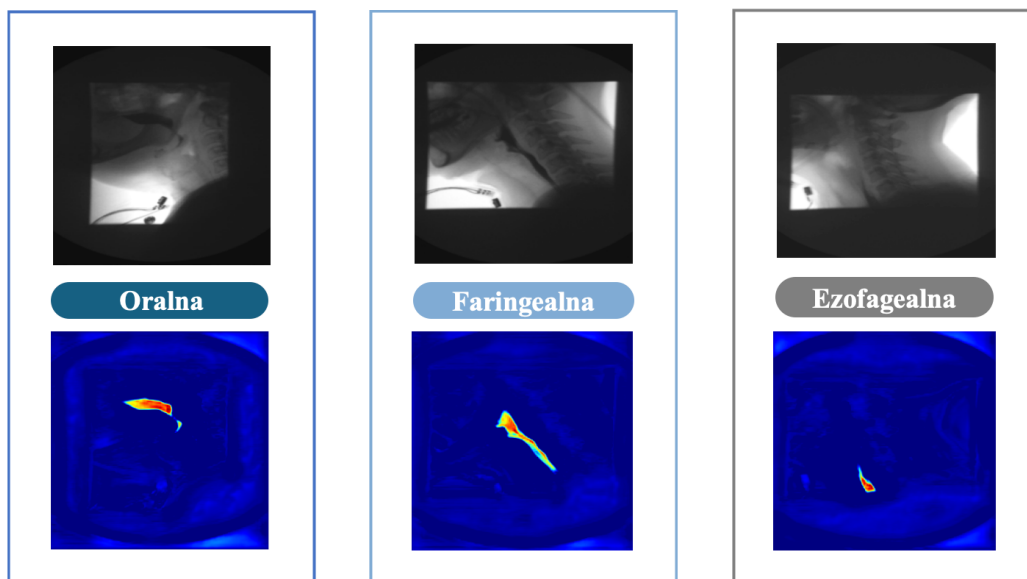
značajnog poboljšanja performansi zadatka segmentacije bolusa ($p = 0,0039$), pretpostavljamo da bi se primjećeno poboljšanje generalizovalo na sve ostale modele pretext zadatka. Da bismo dodatno povećali kredibilitet dobijenih rezultata i ponudili vizuelna objašnjenja za proces donošenja odluka, koristili smo Grad-CAM tehniku [147], koja je pokazala da naš pristup uspješno segmentira bolus kroz sve tri faze gutanja (Slika 13).



Grafikon 1. Poređenje enensemble strategije između SSL i supervised learning pristupa (F1 score vs broj korišćenih modela)

Tabela 9. Performanse zadatka segmentacije bolusa kroz dva različita pristupa 9 nasumičnih podjela (split) skupa podataka (F1 score)

Pristup	Run	Split1	Split2	Split3	Split4	Split5	Split6	Split7	Split8	Split9
Supervised	1	0.742	0.689	0.773	0.791	0.749	0.755	0.762	0.762	0.754
	2	0.734	0.682	0.773	0.781	0.746	0.745	0.755	0.761	0.741
	3	0.737	0.683	0.768	0.787	0.733	0.747	0.754	0.753	0.742
		0.738	0.685	0.771	0.786	0.743	0.749	0.757	0.759	0.746
	<i>mean±std</i>	±	±	±	±	±	±	±	±	±
		0.0040	0.0038	0.0029	0.0050	0.0085	0.0053	0.0044	0.0049	0.0072
Self-supervised	1	0.749	0.714	0.783	0.810	0.760	0.782	0.761	0.775	0.774
	2	0.743	0.709	0.770	0.796	0.757	0.770	0.759	0.761	0.762
	3	0.735	0.695	0.766	0.800	0.758	0.754	0.760	0.760	0.764
		0.742	0.706	0.773	0.802	0.758	0.769	0.760	0.765	0.767
	<i>mean±std</i>	±	±	±	±	±	±	±	±	±
		0.0070	0.0098	0.0089	0.0072	0.0015	0.0140	0.0010	0.0084	0.0064



Slika 13. Grad-CAM vizuelizacije koje ilustruju donošenje odluka modela kroz tri faze gutanja u zadatku segmentacije bolusa

4.4. Diskusija

Cilj ovog istraživanja je razvoj pouzdanog SSL modela koji će omogućiti poboljšanje performansi zadatka segmentacije bolusa, prethodno postignutih korišćenjem supervised learning pristupa. Konkretno, razvijeni model za segmentaciju bolusa treba da odgovori na pitanje „*Koje oblasti na snimcima predstavljaju „bolus“ i gdje se na snimcima nalaze te oblasti?*“ i to bez potrebe za ljudskom intervencijom. Zapremina bolusa, do određene mjere, moduliše nekoliko fizioloških odgovora, kao što su šire otvaranje gornjeg ezofagealnog sfinktera (UES) i izraženije anteriorno pomjeranje hoidne kosti u poređenju sa manjim bolusima. Dodatno utiče i na vremenski slijed fizioloških događaja tokom normalnog gutanja. Sve navedeno čini lokalizaciju bolusa i svih njegovih ostataka jednim od ključnih parametara u procjeni disfagije [1]. Poremećaji tokom faza procesa gutanja dovode do ugrožene sigurnosti i efikasnosti. Posebno su kritični poremećaji tokom faringealne faze, jer mogu povećati rizik od ulaska djelova bolusa u disajne puteve. U zavisnosti od vremena i vrste ovih poremećaja, to može rezultirati manjim ili većim aspiracijama, što može dovesti do hitnih medicinskih stanja [134]. Najzastupjenija dijagnostička metoda za procjenu disfagije je videofluoroscopska studija gutanja (VFSS), koja zahtijeva ručnu analizu frejm po frejm, čineći proces analize

dugotrajnim [31]. Stoga postoji jasna potreba za razvojem sistema za automatsku analizu poremećaja gutanja.

Tokom godina u iMED laboratoriji, segmentacija bolusa predstavlja jedan od glavnih pravaca istraživanja. Ranije su se istraživači uglavnom fokusirali na primjenu supervised learning pristupa za automatsku procjenu disfagije u videofluoroskopskoj studiji gutanja. Ove studije su obuhvatale detekciju i praćenje pokreta hioidne kosti [31], [148], [149] ili rješavanje zadatka segmentacije bolusa [26], [28], [150]. U okviru ovih studija, usmjerenih na zadatke segmentacije bolusa, identifikovali smo nekoliko izazova povezanih sa supervised learning pristupom. Glavni izazovi uključuju nedostatak robusnosti i generalizacije, kao i razliku između pretraining skupa podataka i downstream skupa podataka. Potencijalno rješenje za navedene nedostatke moglo bi da podrazumijeva primjenu sofisticiranijeg ML pristupa kao što je SSL. Oslanjajući se na ovu pretpostavku započeli smo prvo istraživanje o mogućnostima primjene SSL pristupa u zadatku automatske procjene poremećaja gutanja. Dodatno, ovo istraživanje je uključivalo poređenje performansi različitih ML pristupa, uključujući: supervised learning, SSL, kao i SSL u kombinaciji sa ensemble strategijom. U tu svrhu za sve pristupe korišćen je isti skup podataka, koji je uključivao frejmove rezolucije 256x256 u kombinaciji sa ResNet18 backbone mrežom.

Algoritam kreiran tokom istraživanja implementiran je na Compute Canada cluster okruženju, zahtijevajući oko tri nedelje za treniranje/obuku svih modela. Rezultati su pokazali važnost pretraining faze u postizanju superiornih performansi. U ovom kontekstu, umjesto korišćenja ImageNet pretrained weights primijenjene su SSL weights, što je rezultiralo statistički značajnim poboljšanjem performansi downstream zadataka ($p=0,0039$). U okviru pretext zadatka, identifikovano je pet ključnih hiperparametara: transformacija frejmova, brzina učenja, preskakanje frejmova, broj frejmova po klipu i maksimalan broj klipova po videu. Istraživanje je pokazalo da efikasna strategija augmentacije skupa podataka fluoroskopije uključuje transformacije poput varijacija boja (color jitter) i horizontalno/vertikalno okretanje (flip transformations). Konkretno, uticaj color jitter transformacije na kvalitet učenja reprezentacija pokazao se kao značajan. Transformacije koje podrazumijevaju isjecanje (crop transformations) namjerno su izostavljene kako bi se spriječio gubitak ključnih informacija iz fluoroskopskih snimaka. Rezultati su takođe pokazali da hiperparametar preskakanje između frejmova (frame skip)

igra značajnu ulogu u pretext zadatku. Ovaj hiperparametar treba da omogući dovoljno preskakanja između frejmova kako bi model mogao efikasno učiti promjene na nivou piksela. Najbolji rezultati postignuti su kada je klip kreiran uzorkovanjem svakog osmog frejma. Što se tiče ostalih hiperparametara, poput broja frejmova po klipu i maksimalnog broja klipova po videu, model je pokazao veću robusnost. Iako sprovedeno istraživanje nailazi na slična ograničenja u pogledu skupova podataka (npr. često nizak kvalitet rendgenskih snimaka, koji mogu da variraju od jednog uređaja do drugog, što dovodi do varijacija u performansama), kao studija sprovedena od strane Z. Zhang i koautori [31], naše istraživanje se oslanjalo na podatke prikupljene tokom rutinske kliničke njege. Time je unaprijeđena spoljašnja validnost kreiranih metoda kako bi bile u skladu sa standardnim kliničkim praksama.

Ovakav pristup imao je za cilj testiranje robusnosti metoda na podacima prikupljenim u stvarnim kliničkim uslovima, umjesto u kontrolisanim laboratorijskim uslovima. Važno je napomenuti da su podaci prikupljeni iz iste bolnice, što potencijalno ograničava dijagnostičku varijabilnost među pacijentima. Buduća istraživanja treba usmjeriti ka validaciji performansi modela na fluoroskopskim setovima podataka sakupljenim iz različitih bolnica i sa različitih uređaja. Takođe, istraživanje federated learning metode za automatizovanu analizu poremećaja gutanja moglo bi približiti primjenu ML modela svakodnevnoj kliničkoj praksi. Dodatno, u ovom istraživanju fokusirali smo se na korišćenje U-Net++ modela u zadatku segmentacije bolusa budući da se u literaturi navodi da je baš ovaj model najpogodniji za situacije u kojima je skup podataka kreiran sa malim brojem uzoraka. U budućem istraživanju performanse postignute korišćenjem U-Net++ modela trebalo bi uporediti sa modelima kao npr. EfficientNet, FastViT.

4.5. Zaključak

U ovom istraživanju razvijena je pouzdana SSL mreža za automatsku segmentaciju bolusa čime su unaprijeđene performanse postignute korišćenjem supervised learning pristupa. Za učenje reprezentacija iskorišćen je model Contrastive Random Walk model, dok je za downstream zadatak korišćen U-Net++ model. SSL pristup omogućio je poboljšanje performansi za 0.016 F1 score u poređenju sa supervised learning pristupom. Posebno je važan doprinos koji se ogleda uvođenjem novog koncepta

SSL ensemble modela, gdje članovi ensemble odgovaraju modelima sa različitim SSL inicijalizacijama. Implementacija pomenutog koncepta omogućila je dodatno poboljšanje performansi za 0.027 F1 score u poređenju sa ImageNet inicijalizacijom. Sprovedeno istraživanje zasigurno predstavlja početni korak u primjeni i istraživanju upotrebe SSL pristupa u fluoroskopskim skupovima podataka.

5. Automatsko mjerenje faringealnog ostatka

Glavni doprinosi prezentovani u ovom poglavlju su sastavni dio istraživačkog rada Martinović I., Dousty M., Li W., Đukanović M., Coyle L. J., Sejdić E. (2025), *Leveraging Ensemble on Self-Supervised Techniques to Enhance Model Performance: A Regime Analysis*.

Disfagija se manifestuje kroz poremećaje u dva ključna funkcionalna aspekta gutanja, sigurnost i efikasnost. Prethodno istraživanje se odnosilo na automatizaciju analize parametara koji su ključni za procjenu sigurnosti gutanja. Međutim, u literaturi je primjetan nedostatak studija koje se bave automatskom analizom parametara povezanih sa efikasnošću gutanja. Poremećaji efikasnosti gutanja manifestuju se kroz sakupljanje ostataka bolusa u farinks, nakon završetka procesa gutanja, što može rezultirati aspiracijom. U ovom istraživanju koristeći mogućnosti ML realizovano je automatsko određivanje količine faringealnog ostatka, koristeći kvantitativnu kliničku metodu $\%(C_2 - C_4)^2$. Kreiranje ML sistema oslanja se na kombinovanjem modela razvijenih u prethodnom istraživanju (*segmentacija bolusa koristeći self-supervised learning pristup*) i modelom za detekciju karakterističnih tačaka (*landmark detection task*) razvijenom u okviru iMED Lab istraživačke grupe. Sprovedeno istraživanje dokazuje da ML ne doprinosi samo unapređenju analize parametara povezanih sa sigurnošću gutanja, već da se isto tako uspješno može iskoristiti i za unapređenje analize parametara povezanih sa efikasnošću gutanja, a sve u okviru videofluoroskopske studije gutanja.

5.1. Uvod

Efikasnost gutanja se odnosi na sposobnost transporta hrane od usne duplje, kroz jednjak, do želuca, bez formiranja ostataka bolusa u farinks po završetku procesa gutanja [151]. Istraživači navode da proces gutanja uglavnom predstavlja fenomen koji se zasniva na pritisku [152]. Iz tog razloga efikasno kretanje bolusa se postiže stvaranjem adekvatne razlike pritisaka u zonama ispred i iza bolusa, što u konačnici rezultira jakim propulzivnim silama [151]. Poremećaji u ovom funkcionalnom aspektu gutanja se

manifestuju kroz sakupljanje djelova bolusa u farinks nakon završetka procesa gutanja. Ostaci bolusa se najčešće mogu locirati u valekule (valleculae) i piriformnim sinusima (pyriform sinuses) [153]. Valekule predstavljaju anatomske strukture koje se nalaze u gornjem dijelu farinksa, između korijena jezika i epiglotisa. U procesu gutanja valekule imaju veoma važnu ulogu, djelujući kao „*privremeni rezervoari*“ za bolus prije transporta ka jednjaku. Sa druge strane, piriformni sinusi predstavljaju dvije uparene šupljine smještene u donjem dijelu farinksa. Ove anatomske strukture predstavljaju dio zaštitnog mehanizma koji omogućava pravilno usmjeravanje hrane dalje od disajnih puteva. Nakupljanje djelova bolusa u pomenute anatomske strukture, nakon završetka procesa gutanja, može izazvati niz posledica za pacijente od kojih je najozbiljnija aspiracija ostataka bolusa u disajne puteve. Aspiracija je česta pojava kod poremećaja u efikasnosti gutanja iz razloga što se nakon završetka procesa gutanja povlači zaštitni mehanizam koji štiti disajne puteve od penetracije bolusa [154]. Ozbiljnost ovog poremećaja najbolje ilustruje istraživanje [155], koje je pokazalo da je oko 56% registrovanih aspiracija direktno povezano sa formiranjem faringealnog ostatka. Iz tog razloga postoji jasna klinička potreba za kvantifikacijom faringealnog ostatka kao mjere ozbiljnosti poremećaja efikasnosti gutanja.

U kliničkoj praksi postoje različite metode koje se koriste za kvantifikaciju faringealnog ostatka na fluoroskopskim snimcima [153]. Uspostavljene metode se mogu kategorizovati na: vizuelno-perceptivne metode procjene prisustva (ili odsustva) ostataka u farinksu, vizuelno-perceptivne metode procjene količine ostataka ili eliminacije bolusa u odnosu na bolusa na početku gutanja, vizuelno-perceptivne metode procjene stepena popunjenosti anatomskih struktura u farinksu (valekule ili piriformni sinusa) ostacima bolusa, te kvantitativne metode bazirane na mjerenje površine ostataka u pikselima. U ovom istraživanju fokus će biti na kvantitativnoj metodi baziranoj na mjerenje piksela $%(C_2-C_4)^2$ [156], prije svega jer se ova klinička metoda karakteriše sa izuzetnom pouzdanošću i preciznošću.

Doprinos ovog dijela doktorskog rada ogleda se u uspostavljanju nove oblasti istraživanja, koje se odnosi na kreiranje automatskih metoda za merenje faringealnog ostatka. Ovim pionirskim poduhvatom želi se istraživački fokus sa kreiranja automatskih metoda analize sigurnosti gutanja preusmjeriti na kreiranje automatskih metoda za analizu efikasnosti gutanja. Pregledom baza naučnih publikacija pronađena je samo jedna

studija [154], koja korespondira sa novouspostavljenom istraživačkom oblasti. Studija je sprovedena u okviru iMED Lab istraživačke grupe i oslanja se na supervised learning metode.

Automatizacija $\%(C_2-C_4)^2$ kliničke metode u ovom istraživanju biće realizovana kombinacijom ML modela, baziranog na SSL pristupu, za zadatak segmentacije bolusa i ranije razvijene metode u okviru iMED Lab istraživačke grupe za detekciju karakterističnih tačaka [157], tj. pršljenova C_2 i C_4 . Razvojem automatske metode namijenjena procjeni efikasnosti gutanja želi se kliničarima ponuditi dodatna podrška u cilju efikasnije i tačnije procjene procesa gutanja.

5.2. Materijali i Metode

5.2.1. Kreiranje skupa podataka

Skup podataka korišćen u ovom dijelu doktorskog rada istovjetan je skupu podataka kreiranom za potrebe razvoja ML modela za segmentaciju bolusa. Važno je napomenuti da protokol studije odobren od strane Etičkog komiteta (PRO19040040) institucije odakle podaci potiču, Medicinskog centara Univerziteta u Pittsburgu. U studiju je uključeno 270 studija gutanja od 28 pacijenata, starosti od 19 do 94 godine. Broj frejmova u fluoroskopskim snimcima iznosi 30, sa rezolucijom 720x1080.

5.2.2. Klinička analiza efikasnosti gutanja

Klinička analiza efikasnosti gutanja u videofluoroskopskoj studiji gutanja zasniva se na metodama koje omogućavaju kvantifikaciju faringealnog ostatka. U ovom istraživanju fokus je stavljen na kvantitativne metode bazirane na mjerenje površine ostataka u pikselima [156]. Glavne prednosti ovih metoda u odnosu na vizualno-perceptivne metode ogledaju se u: boljoj preciznosti, tj. smanjenju subjektivnosti; boljoj replikabilnosti, tj. boljoj interrater pouzdanosti; kao i mogućnosti detekcije malih ali kliničkih značajnih promjena koje se ne mogu uočiti na ordinalnim skalama (npr. 0: nema ostataka, 1: tanki sloj ostataka, 2: nakupljanje ostataka) . U Tabeli 10. prikazane su razvijene kvantitativne kliničke metode za procjenu efikasnosti gutanja bazirane na mjerenju površine ostataka u pikselima. Posebni fokus će bit na $\%(C_2-C_4)^2$ metodi jer se

na osnovu dostupnih istraživanja pokazala kao najpouzdanija i najpreciznija metoda za kliničku primjenu [153]. U literaturi [158] se navodi da se rizik od penetracije i aspiracije povećava kada je $\%(C_2-C_4)^2 > 3\%$.

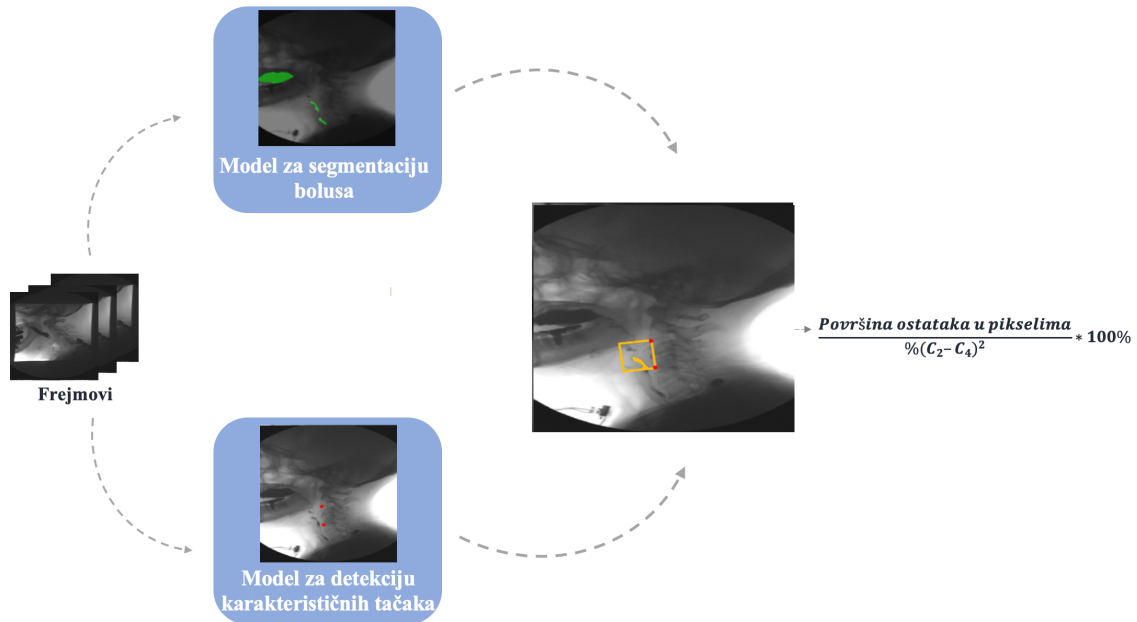
Tabela 10. Kvantitativne metode bazirane na mjerenje površine ostataka u pikselima

<i>Naziv kliničke metode</i>	<i>Pristup za kvantifikaciju faringealnog ostatka</i>
<i>%-Full [159]</i>	$(\text{Površina ostatka} / \text{Prostorna površina}) \times 100$
<i>Normalized Residue Ratio Scale (NRRS_v = valemule, NRRS_p = piriformni sinusi) [160]</i>	$(\%-\text{Full}) \times [(\text{Površina ostatka} / ((\text{Dužina } C_2-C_4)^2) \times 10]$
$\%(C_2-C_4)^2$ [156]	$\text{Površina ostatka} / ((\text{Dužina } C_2-C_4)^2)$

5.2.3. Automatska analiza efikasnosti gutanja

Automatizacija kvantifikacije faringealnog ostatka koristeći $\%(C_2-C_4)^2$ kliničku metodu postignuta je kombinacijom ML modela za segmentaciju bolusa i ML modela za detekciju karakterističnih tačaka (Slika 14). Na osnovu Tabele 10. proističe da se kvantifikacija faringealnog ostatka sprovodi računajući površinu ostatka u odnosu na kvadratni C_2-C_4 referentni skalar. Iz tog razloga, model realizovan za zadatak segmentacije bolusa omogućava računanje površine ostatka u pikselima, dok model za detekciju karakterističnih tačaka omogućava detekciju C_2 i C_4 pršljena. Detaljan opis realizacije modela za segmentiranje bolusa, baziranog na SSL pristupu, dat je u poglavlju 4, dok je realizacija modela za detekciju karakterističnih tačaka realizovana u iMED Lab istraživačkoj grupi i prezentovana u okviru naučne publikacije [157]. Primarna namjena modela za detekciju karakterističnih tačaka je praćenje pomjeranja hioidne kosti u C_2-C_4 koordinatnom sistemu. Realizovani model se zasniva na dubokoj Full-HRNet neuralnoj mreži koja detektuje četiri karakteristične tačke. Prilagođavanjem ovog modela omogućeno je detekcija dvije tačke. Obuka ovog modela sprovedena je na istom skupu podataka kao model iz Poglavlja 4, koji je uključivao fluoroskopske snimke rezolucije

256x256. Međutim treba naglasiti da se kvantifikacija faringealnog ostatka sprovodi nakon završetka procesa gutanja, tj. na poslednjem frejmu videofluoroskopske studije gutanja.



Slika 14. Automatska kvantifikacija faringealnog ostatka koristeći kliničku metodu $\% (C_2 - C_4)^2$

5.2.4. Statistička analiza

Statistička analiza za ML model namijenjen automatskoj segmentaciji bolusa opisana je u prethodnom poglavlju. Procjena performansi ML modela za detekciju karakterističnih tačaka oslanja se na srednjoj prosječnoj grešci (mean average error-MAE), koja se izračunava kao rastojanje između stvarne pozicije N i predviđene pozicije N' u poslednjem frejmu (jednačina 1).

$$MAE = |N - N'| \dots \dots \dots (1)$$

5.3. Rezultati i Diskusija

Detaljan opis rezultata ML modela za automatsku segmentaciju bolusa prikazan je u prethodnom istraživanju. Šta više, performanse modela baziranog na SSL pristupu upoređivane su sa široko rasprostranjenim modelima baziranim na supervised learning

pristupu. Performanse ML modela za segmentaciju bolusa kombinovane su sa performansama modela za detekciju karakterističnih tačaka u cilju kvantifikacije faringealnog ostataka koristeći $\%(C_2-C_4)^2$ kliničku metodu. Performanse modela za detekciju C_2 pršljena je 4,4128 MAE, dok je 3,1075 MAE za detekciju C_4 pršljena.

Cilj ovog istraživanja je uspostavljanje nove istraživačke oblasti, koja će biti usmjerena ka razvoju automatskih metoda za procjenu efikasnosti gutanja. Pregledom literature jasno se uočava da je trenutno glavni fokus na razvoju automatskih metoda za procjenu sigurnosti gutanja. U želji da se fokus usmjeri ka analizi efikasnosti gutanja u ovom dijelu doktorskog rada automatizovana je $\%(C_2-C_4)^2$ klinička metoda. Ova klinička metoda spada u grupu kvantitativnih metoda baziranih na mjerenju površine ostataka u pikselima. Ove metode su ocijenjene kao znatno bolje u odnosu na vizualno-perceptivne metode. Međutim treba naglasiti da kvantitativne metode bazirane na mjerenju površine ostataka u pikselima imaju određena ograničenja, koja se odnose na sve kliničke metode primijenjene za procjenu efikasnosti gutanja u videofluoroskopskoj studiji gutanja. Ovo se primarno odnosi na dvodimenzionalnu prirodu (2D) fluoroskopskih snimaka što u konačnici otežava kvantifikaciju ostatka bolusa, koji je trodimenzionalne (3D) prirode. Ipak istraživanja pokazuju da postoji uska korelacija između kvantifikacije ostatka bolusa koristeći (2D) fluoroskopske snimke i 3D kompjuterizovanu tomografiju ($R^2 = .91$) [161].

Sprovedeno istraživanje je realizovano oslanjajući se na mogućnosti Compute Canada cluster infrastrukture a sve to zahvaljujući saradnji sa iMED Lab istraživačkom grupom sa Univerziteta u Toronto. Postignuti rezultati jasno ukazuju da se analiza efikasnosti gutanja može jednako uspješno automatizovati kao i analiza sigurnosti gutanja. Dodatno, ovo istraživanje je pokazalo na koji način se kombinacijom već postojećih ML modela unutar istraživačke grupe mogu uspješno kreirati inovativni sistemi koji imaju jasnu kliničku opravdanost. Takođe, baš u ovoj primjeni demonstrirano je da povećanje performansi korišćenjem SSL pristupa za F1 rezultat od 0,016 (0,027) u zadatku segmentacije bolusa ima veliki klinički značaj. Iako ovo poboljšanje može predstavljati samo nekoliko procenata volumetrijske razlike u poređenju sa rezultatom dobijenim korišćenjem supervised learning pristupa, ono je i dalje značajno u kontekstu automatske procjene efikasnosti gutanja, gdje netačna procjena faringealnog ostatka može dovesti do pogrešne dijagnoze. Predložena metoda u ovom poglavlju predstavlja novi pristup koji otvara novo istraživačko polje u proceni efikasnosti gutanja.

Nedostaci predloženog sistema za automatsko mjerenje faringealnog ostatka su prije svega povezani sa korišćenim skupom podataka kreiranim u okviru iste zdravstvene ustanove i na istom fluoroskopskom aparatu. Ovo u konačnici može rezultirati degradacijom performansi modela u slučaju da se sistem primijeni na podacima koji potiču iz druge zdravstvene ustanove ili od drugog fluoroskopskog uređaja. Takođe, dodatan problem pričinjava i relativno mali skup podataka koji bi u budućnosti mogao biti proširen koristeći mogućnosti difuzionih modela.

5.4. Zaključak

Poremećaji u gutanju zahvataju dva usko povezana funkcionalna aspekta i to sigurnost i efikasnost. Automatska analiza sigurnosti gutanja je realizovana u prethodnom poglavlju gdje je razvijen ML model baziran na SSL pristupu za automatsku segmentaciju bolusa. U ovom istraživanju kreirani ML model za segmentaciju bolusa je u kombinaciji sa ML modelom za detekciju karakterističnih tačaka omogućio automatizaciju $\%(C_2-C_4)^2$ kliničke metode. Ova klinička metoda omogućava kvantizaciju faringealnog ostatka na nivou piksela. Postignutim rezultatima jasno je pokazano da se ML može uspješno iskoristiti za automatsku analizu cjelokupnog procesa gutanja. Iako ovo predstavlja pionirski korak u automatskoj analizi efikasnosti gutanja jasno se uočava prostor za buduća istraživanja, koja treba da uključe najsavremenija dostignuća u oblasti vještačke inteligencije. Jedan od mogućih budućih pravaca bi mogao da bude i primjena multi-modalnih sistema gdje će se u procjeni procesa gutanja uključiti i drugi tipovi podataka, a ne samo fluoroskopski snimci.

6. Zaključak

Gutanje predstavlja izuzetno složen neuromišićni proces koji uključuje više od 30 nerava i mišića. Složenost ovog procesa ogleda se u činjenici da je gutanje usko povezano sa disanjem, što zahtijeva preciznu koordinaciju. U slučaju da proces gutanja nije precizno koordinisan dolazi do poremećaja u gutanju, koji se naziva disfagija. Disfagija se manifestuje kroz uticaj na sigurnost i efikasnost gutanja. Pojavom disfagije pacijenti se suočavaju sa posledicama koje se kreću od blagih do veoma ozbiljnih kakva je penetracija i/ili aspiracija bolusa u disajne puteve. Ovo doktorsko istraživanje sprovedeno je u iMED Lab istraživačkoj grupi na Univerzitetu u Toronto i rezultiralo je kreiranjem automatskih metoda za procjenu sigurnosti i efikasnosti procesa gutanja. Prva razvijena metoda se odnosi na automatsku segmentaciju bolusa, dok je u drugom dijelu dokorskog rada fokus stavljen na automatskoj detekciji faringealnog ostatka. Postignuti rezultati jasno ukazuju da se vještačka inteligencija sa velikim uspjehom može primijeniti u analizi ključnih parametara koji se odnose na procjenu poremećaja gutanja. Primjenom razvijenih metoda u svakodnevnoj kliničkoj praksi vjeruje se da se tačnost i efikasnost rada ljekara može dodatno biti unapređena. Međutim da bi metode bazirane na ML postale sastavni dio kliničke prakse potrebno je da razvijene metode posjeduju adekvatnu generalizaciju i robusnost. Ova dva funkcionalna aspekta potrebno je dodatno procijeniti budući da podaci korišćeni u ovom doktorskom radu potiču iz jedne zdravstvene ustanove i sa jednog dijagnostičkog uređaja. U budućnosti bi trebalo proširiti set podataka a jedna od mogućih metoda je federated learning metoda koja ne zahtijeva direktno dijeljenje povjerljivih medicinskih podataka sa drugim zdravstvenim ustanovama. Takođe, budući pravci istraživanja se tiču primjene najsavremenijih ML metoda koje se uglavnom oslanjaju na multimodalni pristup. Pored istraživačkog pravca na polju primjene vještačke inteligencije u procesu analize poremećaja gutanja ovo doktorsko istraživanje je podstaklo istraživačke aktivnosti u pravcu razvoja inovativne minimalno invazivne dijagnostičke metode.

Finansijska podrška

Istraživanja prikazana u doktorskoj tezi realizovana su uz podršku Ministarstva nauke Crne Gore, kroz grant za realizaciju doktorskih istraživanja na univerzitetima u Crnoj Gori (br:02/2-062/20-1909). Dodatno, boravak u iMED Lab istraživačkoj grupi na Univerzitetu u Torontu podržan je od strane kompanije Crnogorski elektroprenosni sistem AD (br:1-1-D/23-2385/14).

Partneri

Sprovedeno doktorsko istraživanje realizovano je u iMED Lab istraživačkoj grupi u partnerstvu sa Univerzitetom u Torontu, Univerzitetom u Pitsburgu, Kliničkim Centrom Crne Gore i kompanijom Glosarij doo.



Literature

- [1] S. Hadjikitoutis, “Abnormal patterns of breathing during swallowing in neurological disorders,” *Brain*, vol. 123, no. 9, pp. 1863–1873, Sep. 2000.
- [2] K. Matsuo and J. B. Palmer, “Coordination of mastication, swallowing and breathing,” *Japanese Dental Science Review*, vol. 45, no. 1, pp. 31–40, May 2009.
- [3] Cichero, JULIE A Y and Murdoch, Bruce E., *Dysphagia: Foundation, Theory and Practice*, 1st ed. Wiley, 2006.
- [4] S. Mao, “A Study of Deep Neural Networks Associated Computational Deglutition for Swallowing Assessment,” PhD, University of Pittsburgh, 2022.
- [5] R. Ding and J. A. Logemann, “Pneumonia in Stroke Patients: A Retrospective Study,” *Dysphagia*, vol. 15, no. 2, pp. 51–57, Mar. 2000.
- [6] P. E. Marik, “Aspiration Pneumonitis and Aspiration Pneumonia,” *N Engl J Med*, vol. 344, no. 9, pp. 665–671, Mar. 2001.
- [7] C. Takizawa, E. Gemmell, J. Kenworthy, and R. Speyer, “A Systematic Review of the Prevalence of Oropharyngeal Dysphagia in Stroke, Parkinson’s Disease, Alzheimer’s Disease, Head Injury, and Pneumonia,” *Dysphagia*, vol. 31, no. 3, pp. 434–441, Jun. 2016.
- [8] J. Curran and M. E. Groher, “Development and dissemination of an aspiration risk reduction diet,” *Dysphagia*, vol. 5, no. 1, pp. 6–12, Mar. 1990.
- [9] L. Barnes, L. Bergstrom, D. Clarke, and M. Petersen, “Speech Pathology Learner Guide Provide support in dysphagia management.” The State of Queensland (Queensland Health), Apr-2017.
- [10] J. Leira, A. Maseda, L. Lorenzo-López, N. Cibeira, R. López-López, L. Lodeiro, and J. C. Millán-Calenti, “Dysphagia and its association with other health-related risk factors in institutionalized older people: A systematic review,” *Archives of Gerontology and Geriatrics*, vol. 110, p. 104991, Jul. 2023.
- [11] G. Lanzoni, C. Sembenini, S. Gastaldo, L. Leonardi, V. P. Bentivoglio, G. Faggian, L. Bosa, P. Gaio, and M. Cananzi, “Esophageal Dysphagia in Children: State of the Art and Proposal for a Symptom-Based Diagnostic Approach,” *Front. Pediatr.*, vol. 10, p. 885308, Jun. 2022.
- [12] Gastrointestinal Society, “Dysphagia.” Canadian Society of Intestinal Research.
- [13] N. Audag, M. Toussaint, G. Liistro, L. Vandervelde, E. Cugy, and G. Reychler, “European Survey: Dysphagia Management in Patients with Neuromuscular Diseases,” *Dysphagia*, vol. 37, no. 5, pp. 1279–1287, Oct. 2022.
- [14] M. Aslam and M. F. Vaezi, “Dysphagia in the elderly,” *Gastroenterol Hepatol (N Y)*, vol. 9, no. 12, pp. 784–795, Dec. 2013.
- [15] J. Sotirović, “Swallowing disorders,” *Galenika Med J*, vol. 1, no. 4, pp. 77–83, 2022.
- [16] K. Matsuo and J. B. Palmer, “Anatomy and Physiology of Feeding and Swallowing: Normal and Abnormal,” *Physical Medicine and Rehabilitation Clinics of North America*, vol. 19, no. 4, pp. 691–707, Nov. 2008.
- [17] J. Robbins, J. W. Hamilton, G. L. Lof, and G. B. Kempster, “Oropharyngeal swallowing in normal adults of different ages,” *Gastroenterology*, vol. 103, no. 3, pp. 823–829, Sep. 1992.

- [18] K. L. DePippo, M. A. Holas, and M. J. Reding, "Validation of the 3-oz Water Swallow Test for Aspiration Following Stroke," *Archives of Neurology*, vol. 49, no. 12, pp. 1259–1261, Dec. 1992.
- [19] D. W. Vander Linden, J. H. Cauraugh, and T. A. Greene, "The Effect of Frequency of Kinetic Feedback on Learning an Isometric Force Production Task in Nondisabled Subjects," *Physical Therapy*, vol. 73, no. 2, pp. 79–87, Feb. 1993.
- [20] N. H. Zaidi, H. A. Smith, S. C. King, C. Park, P. A. O'Neill, and M. J. Connolly, "Oxygen Desaturation on Swallowing as a Potential Marker of Aspiration in Acute Stroke," *Age Ageing*, vol. 24, no. 4, pp. 267–270, 1995.
- [21] L. Sp, "Dysphagia: Approach to Assessment and Treatment," *JHNSS*, vol. 1, no. 1, Jun. 2017.
- [22] M. S. De Lima, L. D. Mangilli, F. C. Sassi, and C. R. F. De Andrade, "Functional magnetic resonance and swallowing: critical literature review," *Brazilian Journal of Otorhinolaryngology*, vol. 81, no. 6, pp. 671–680, Nov. 2015.
- [23] S. Roldan-Vasco, A. Orozco-Duque, and J. R. Orozco-Arroyave, "Swallowing disorders analysis using surface EMG biomarkers and classification models," *Digital Signal Processing*, vol. 133, p. 103815, Mar. 2023.
- [24] Y. Inamoto, R. Ueha, and M. Gonzalez-Fernandez, "Use of CT for Dysphagia Evaluation: Advantages and Disadvantages in the Study of Swallowing," *Curr Phys Med Rehabil Rep*, vol. 12, no. 3, pp. 250–255, Jun. 2024.
- [25] A. Alexandropoulou, E. Magkouti, A. Despoti, N. Leventakis, and S. Nanas, "Studying the Swallow using Surface Electroencephalography: A Systematic Review," *HRJ*, vol. 9, no. 2, pp. 104–114, Apr. 2023.
- [26] H. Caliskan, A. S. Mahoney, J. L. Coyle, and E. Sejdic, "Automated Bolus Detection in Videofluoroscopic Images of Swallowing Using Mask-RCNN," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, 2020, pp. 2173–2177.
- [27] K.-C. Wei, S.-H. Cheng, M.-Y. Hsiao, Y.-C. Wang, C.-H. Weng, J.-Y. Chen, and T.-G. Wang, "Swallowing kinematic analysis might be helpful in predicting aspiration and pyriform sinus stasis," *Sci Rep*, vol. 12, no. 1, p. 1354, Jan. 2022.
- [28] A. Bandini, S. Smaoui, and C. M. Steele, "Automated pharyngeal phase detection and bolus localization in videofluoroscopic swallowing study: Killing two birds with one stone?" arXiv, 26-May-2022.
- [29] E. Sejdic, "iMED Lab."
- [30] E. Sejdić, Y. Khalifa, A. S. Mahoney, and J. L. Coyle, "ARTIFICIAL INTELLIGENCE AND DYSPHAGIA: NOVEL SOLUTIONS TO OLD PROBLEMS," *Arq. Gastroenterol.*, vol. 57, no. 4, pp. 343–346, Dec. 2020.
- [31] Z. Zhang, J. L. Coyle, and E. Sejdić, "Automatic hyoid bone detection in fluoroscopic images using deep learning," *Sci Rep*, vol. 8, no. 1, p. 12310, Dec. 2018.
- [32] L. Wynants, B. Van Calster, G. S. Collins, R. D. Riley, G. Heinze, E. Schuit, E. Albu, B. Arshi, V. Bellou, M. M. J. Bonten, D. L. Dahly, J. A. Damen, T. P. A. Debray, V. M. T. De Jong, M. De Vos, P. Dhiman, J. Ensor, S. Gao, M. C. Haller, M. O. Harhay, L. Henckaerts, P. Heus, J. Hoogland, M. Hudda, K. Jenniskens, M. Kammer, N. Kreuzberger, A. Lohmann, B. Levis, K. Luijken, J. Ma, G. P. Martin, D. J. McLernon, C. L. A. Navarro, J. B. Reitsma, J. C. Sergeant, C. Shi, N. Skoetz, L. J. M. Smits, K. I. E. Snell, M. Sperrin, R. Spijker, E. W. Steyerberg, T. Takada, I. Tzoulaki, S. M. J. Van Kuijk, B. C. T. Van Bussel, I. C. C. Van Der Horst, K.

- Reeve, F. S. Van Royen, J. Y. Verbakel, C. Wallisch, J. Wilkinson, R. Wolff, L. Hooft, K. G. M. Moons, and M. Van Smeden, "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal," *BMJ*, p. m1328, Apr. 2020.
- [33] M. R. Hosseinzadeh Taher, F. Haghighi, R. Feng, M. B. Gotway, and J. Liang, "A Systematic Benchmarking Analysis of Transfer Learning for Medical Image Analysis," in *Domain Adaptation and Representation Transfer, and Affordable Healthcare and AI for Resource Diverse Global Health*, vol. 12968, S. Albarqouni, M. J. Cardoso, Q. Dou, K. Kamnitsas, B. Khanal, I. Rekik, N. Rieke, D. Sheet, S. Tsafaris, D. Xu, and Z. Xu, Eds. Cham: Springer International Publishing, 2021, pp. 3–13.
- [34] S. Imai, M. Akahane, Y. Konishi, and T. Imamura, "Benefits of computed tomography in reducing mortality in emergency medicine," *Open Medicine*, vol. 13, no. 1, pp. 394–401, Sep. 2018.
- [35] A. B. Crummy, C. M. Strother, and C. A. Mistretta, "The History of Digital Subtraction Angiography," *Journal of Vascular and Interventional Radiology*, vol. 29, no. 8, pp. 1138–1141, Aug. 2018.
- [36] R. Wake, M. Yoshiyama, H. Iida, H. Takeshita, T. Kusuyama, H. Kanamitsu, H. Mitsui, Y. Yamada, S. Shimodozono, and K. Haze, "History of Coronary Angiography," in *Advances in the Diagnosis of Coronary Atherosclerosis*, S. Kirac, Ed. InTech, 2011.
- [37] J. D. Howell, "Coronary Heart Disease and Heart Attacks, 1912–2010," *Med. Hist.*, vol. 55, no. 3, pp. 307–312, Jul. 2011.
- [38] The National Lung Screening Trial Research Team, "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," *N Engl J Med*, vol. 365, no. 5, pp. 395–409, Aug. 2011.
- [39] E. Bercovich and M. C. Javitt, "Medical Imaging: From Roentgen to the Digital Revolution, and Beyond," *Rambam Maimonides Med J*, vol. 9, no. 4, p. e0034, Oct. 2018.
- [40] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, "Deep learning for chest X-ray analysis: A survey," *Medical Image Analysis*, vol. 72, p. 102125, Aug. 2021.
- [41] K. Pierre, A. G. Haneberg, S. Kwak, K. R. Peters, B. Hochegger, T. Sananmuang, P. Tunlayadechanont, P. J. Tighe, A. Mancuso, and R. Forghani, "Applications of Artificial Intelligence in the Radiology Roundtrip: Process Streamlining, Workflow Optimization, and Beyond," *Seminars in Roentgenology*, vol. 58, no. 2, pp. 158–169, Apr. 2023.
- [42] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." arXiv, 25-Dec-2017.
- [43] C. Leibig, M. Brehmer, S. Bunk, D. Byng, K. Pinker, and L. Umutlu, "Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis," *The Lancet Digital Health*, vol. 4, no. 7, pp. e507–e519, Jul. 2022.
- [44] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies

- With Progress Highlights, and Future Promises,” *Proc. IEEE*, vol. 109, no. 5, pp. 820–838, May 2021.
- [45] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G. S. Corrado, A. Darzi, M. Etemadi, F. Garcia-Vicente, F. J. Gilbert, M. Halling-Brown, D. Hassabis, S. Jansen, A. Karthikesalingam, C. J. Kelly, D. King, J. R. Ledsam, D. Melnick, H. Mostofi, L. Peng, J. J. Reicher, B. Romera-Paredes, R. Sidebottom, M. Suleyman, D. Tse, K. C. Young, J. De Fauw, and S. Shetty, “International evaluation of an AI system for breast cancer screening,” *Nature*, vol. 577, no. 7788, pp. 89–94, Jan. 2020.
- [46] X. Zhuang, Y. Li, Y. Hu, K. Ma, Y. Yang, and Y. Zheng, “Self-supervised Feature Learning for 3D Medical Images by Playing a Rubik’s Cube,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, vol. 11767, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, pp. 420–428.
- [47] Z. Zhou, V. Sodha, M. M. Rahman Siddiquee, R. Feng, N. Tajbakhsh, M. B. Gotway, and J. Liang, “Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, vol. 11767, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, pp. 384–393.
- [48] H. Li, F.-F. Xue, K. Chaitanya, S. Luo, I. Ezhov, B. Wiestler, J. Zhang, and B. Menze, “Imbalance-Aware Self-supervised Learning for 3D Radiomic Representations,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12902, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 36–46.
- [49] S. Shurrab and R. Duwairi, “Self-supervised learning methods and applications in medical imaging analysis: a survey,” *PeerJ Computer Science*, vol. 8, p. e1045, Jul. 2022.
- [50] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang, “Self-supervised Learning: Generative or Contrastive,” *IEEE Trans. Knowl. Data Eng.*, pp. 1–1, 2021.
- [51] A. Newell and J. Deng, “How Useful Is Self-Supervised Pretraining for Visual Tasks?,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 7343–7352.
- [52] N. Tomasev, I. Bica, B. McWilliams, L. Buesing, R. Pascanu, C. Blundell, and J. Mitrovic, “Pushing the limits of self-supervised ResNets: Can we outperform supervised learning without labels on ImageNet?” arXiv, 03-Nov-2022.
- [53] F. Navarro, C. Watanabe, S. Shit, A. Sekuboyina, J. C. Pecken, S. E. Combs, and B. H. Menze, “Evaluating the Robustness of Self-Supervised Learning in Medical Imaging.” arXiv, 14-May-2021.
- [54] M.-K. N. Huu, V. Q. Ngo, T.-D. Nguyen, V.-T. Nguyen, and T. D. Ngo, “Antique Photo Restoration and Colorization via Generative Model,” in *2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, Phu Quoc, Vietnam, 2022, pp. 1–6.
- [55] O. H. Nagoor, J. Whittle, J. Deng, B. Mora, and M. W. Jones, “Sampling strategies for learning-based 3D medical image compression,” *Machine Learning with Applications*, vol. 8, p. 100273, Jun. 2022.

- [56] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, “Pixel Recurrent Neural Networks.” arXiv, 19-Aug-2016.
- [57] T. Salimans, A. Karpathy, X. Chen, and D. P. Kingma, “PixelCNN++: Improving the PixelCNN with Discretized Logistic Mixture Likelihood and Other Modifications.” arXiv, 19-Jan-2017.
- [58] S. Reed, A. van den Oord, N. Kalchbrenner, S. G. Colmenarejo, Z. Wang, D. Belov, and N. de Freitas, “Parallel Multiscale Autoregressive Density Estimation.” arXiv, 10-Mar-2017.
- [59] J. Ehrhardt and M. Wilms, “Autoencoders and variational autoencoders in medical image analysis,” in *Biomedical Image Synthesis and Simulation*, Elsevier, 2022, pp. 129–162.
- [60] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, “Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction,” in *Artificial Neural Networks and Machine Learning – ICANN 2011*, vol. 6791, T. Honkela, W. Duch, M. Girolami, and S. Kaski, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 52–59.
- [61] A. Singh and T. Ogunfunmi, “An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications,” *Entropy*, vol. 24, no. 1, p. 55, Dec. 2021.
- [62] S. Dehghan and M. F. Amasyali, “SelfCCL: Curriculum Contrastive Learning by Transferring Self-Taught Knowledge for Fine-Tuning BERT,” *Applied Sciences*, vol. 13, no. 3, p. 1913, Feb. 2023.
- [63] G. B. Ramani and N. R. Scalise, “It’s more than just fun and games: Play-based mathematics activities for Head Start families,” *Early Childhood Research Quarterly*, vol. 50, pp. 78–89, 2020.
- [64] P. H. Le-Khac, G. Healy, and A. F. Smeaton, “Contrastive Representation Learning: A Framework and Review,” *IEEE Access*, vol. 8, pp. 193907–193934, 2020.
- [65] P. Kumar, P. Rawat, and S. Chauhan, “Contrastive self-supervised learning: review, progress, challenges and future research directions,” *Int J Multimed Info Retr*, vol. 11, no. 4, pp. 461–488, Dec. 2022.
- [66] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, “A Survey on Contrastive Self-Supervised Learning,” *Technologies*, vol. 9, no. 1, p. 2, Dec. 2020.
- [67] Z. Lu, “Brief Introduction to Contrastive Learning Pretext Tasks for Visual Representation.” arXiv, 06-Oct-2022.
- [68] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, P. Bachman, A. Trischler, and Y. Bengio, “Learning deep representations by mutual information estimation and maximization.” arXiv, 22-Feb-2019.
- [69] P. Bachman, R. D. Hjelm, and W. Buchwalter, “Learning Representations by Maximizing Mutual Information Across Views.” arXiv, 08-Jul-2019.
- [70] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, “Deep Clustering for Unsupervised Learning of Visual Features.” arXiv, 18-Mar-2019.
- [71] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, “Unsupervised Learning of Visual Features by Contrasting Cluster Assignments.” arXiv, 08-Jan-2021.
- [72] Y. Tian, D. Krishnan, and P. Isola, “Contrastive Multiview Coding.” arXiv, 18-Dec-2020.

- [73] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum Contrast for Unsupervised Visual Representation Learning,” 2019.
- [74] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A Simple Framework for Contrastive Learning of Visual Representations,” 2020.
- [75] Y. Tian, C. Sun, B. Poole, D. Krishnan, C. Schmid, and P. Isola, “What Makes for Good Views for Contrastive Learning?” arXiv, 18-Dec-2020.
- [76] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. H. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. D. Guo, M. G. Azar, B. Piot, K. Kavukcuoglu, R. Munos, and M. Valko, “Bootstrap your own latent: A new approach to self-supervised Learning,” 2020.
- [77] X. Chen and K. He, “Exploring Simple Siamese Representation Learning.” arXiv, 20-Nov-2020.
- [78] H. Sowrirajan, J. Yang, A. Y. Ng, and P. Rajpurkar, “MoCo-CXR: MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models,” 2020.
- [79] H.-Y. Zhou, S. Yu, C. Bian, Y. Hu, K. Ma, and Y. Zheng, “Comparing to Learn: Surpassing ImageNet Pretraining on Radiographs by Comparing Image Representations,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, vol. 12261, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 398–407.
- [80] W. Liao, H. Xiong, Q. Wang, Y. Mo, X. Li, Y. Liu, Z. Chen, S. Huang, and D. Dou, “MUSCLE: Multi-task Self-supervised Continual Learning to Pre-train Deep Models for X-Ray Images of Multiple Body Parts,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13438, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 151–161.
- [81] T. Truong, S. Mohammadi, and M. Lenga, “How Transferable Are Self-supervised Features in Medical Image Classification Tasks?” arXiv, 30-Nov-2021.
- [82] S. Azizi, B. Mustafa, F. Ryan, Z. Beaver, J. Freyberg, J. Deaton, A. Loh, A. Karthikesalingam, S. Kornblith, T. Chen, V. Natarajan, and M. Norouzi, “Big Self-Supervised Models Advance Medical Image Classification.” arXiv, 01-Apr-2021.
- [83] P. Müller, G. Kaissis, C. Zou, and D. Rueckert, “Radiological Reports Improve Pre-training for Localized Imaging Tasks on Chest X-Rays,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13435, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 647–657.
- [84] A. K. Tanwani, J. Barral, and D. Freedman, “RepsNet: Combining Vision with Language for Automated Medical Reports,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13435, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 714–724.
- [85] B. Bozorgtabar, D. Mahapatra, G. Vray, and J.-P. Thiran, “SALAD: Self-supervised Aggregation Learning for Anomaly Detection on X-Rays,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, vol. 12261, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 468–478.

- [86] J. Park, I.-Y. Kwak, and C. Lim, “A Deep Learning Model with Self-Supervised Learning and Attention Mechanism for COVID-19 Diagnosis Using Chest X-ray Images,” *Electronics*, vol. 10, no. 16, p. 1996, Aug. 2021.
- [87] M. Gazda, J. Plavka, J. Gazda, and P. Drotar, “Self-Supervised Deep Convolutional Neural Network for Chest X-Ray Classification,” *IEEE Access*, vol. 9, pp. 151972–151982, 2021.
- [88] A. Sriram, M. Muckley, K. Sinha, F. Shamout, J. Pineau, K. J. Geras, L. Azour, Y. Aphinyanaphongs, N. Yakubova, and W. Moore, “COVID-19 Prognosis via Self-Supervised Representation Learning and Multi-Image Prediction.” arXiv, 24-Jan-2021.
- [89] Z. Xie, Y. Lin, Z. Zhang, Y. Cao, S. Lin, and H. Hu, “Propagate Yourself: Exploring Pixel-Level Consistency for Unsupervised Visual Representation Learning.” arXiv, 09-Mar-2021.
- [90] L. Schöckel, G. Jost, P. Seidensticker, P. Lengsfeld, P. Palkowitsch, and H. Pietsch, “Developments in X-Ray Contrast Media and the Potential Impact on Computed Tomography,” *Invest Radiol*, vol. 55, no. 9, pp. 592–597, Sep. 2020.
- [91] P. Pape, A. H. Jensen, O. Bergdal, T. N. Munch, S. S. Rudolph, and L. S. Rasmussen, “Time to CT scan for patients with acute severe neurological symptoms: a quality assurance study,” *Sci Rep*, vol. 12, no. 1, p. 15269, Sep. 2022.
- [92] Cleveland Clinic, “CT (Computed Tomography) Scan,” *CT (Computed Tomography) Scan*, 2023. .
- [93] F. Haghghi, M. R. Hosseinzadeh Taher, Z. Zhou, M. B. Gotway, and J. Liang, “Learning Semantics-Enriched Representation via Self-discovery, Self-classification, and Self-restoration,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, vol. 12261, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 137–147.
- [94] X. Tao, Y. Li, W. Zhou, K. Ma, and Y. Zheng, “Revisiting Rubik’s Cube: Self-supervised Learning with Volume-Wise Transformation for 3D Medical Image Segmentation,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, vol. 12264, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 238–248.
- [95] M. Blendowski, H. Nickisch, and M. P. Heinrich, “How to Learn from Unlabeled Volume Data: Self-supervised 3D Context Feature Learning,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, vol. 11769, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, pp. 649–657.
- [96] A. Bitarafan, M. F. Azampour, K. Bakhtari, M. Soleymani Baghshah, M. Keicher, and N. Navab, “Vol2Flow: Segment 3D Volumes Using a Sequence of Registration Flows,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13434, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 609–618.
- [97] F. Matzkin, V. Newcombe, S. Stevenson, A. Khetani, T. Newman, R. Digby, A. Stevens, B. Glocker, and E. Ferrante, “Self-supervised Skull Reconstruction in Brain CT Images with Decompressive Craniectomy,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, vol. 12262, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D.

- Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 390–399.
- [98] R. Bhalodia, L. Kavan, and R. T. Whitaker, “Self-supervised Discovery of Anatomical Shape Landmarks,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, vol. 12264, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 627–638.
- [99] D. Zeng, Y. Wu, X. Hu, X. Xu, H. Yuan, M. Huang, J. Zhuang, J. Hu, and Y. Shi, “Positional Contrastive Learning for Volumetric Medical Image Segmentation,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12902, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 221–230.
- [100] Y. Wu, D. Zeng, Z. Wang, Y. Shi, and J. Hu, “Federated Contrastive Learning for Volumetric Medical Image Segmentation,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12903, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 367–377.
- [101] F. Liu, K. Yan, A. P. Harrison, D. Guo, L. Lu, A. L. Yuille, L. Huang, G. Xie, J. Xiao, X. Ye, and D. Jin, “SAME: Deformable Image Registration Based on Self-supervised Anatomical Embeddings,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12904, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 87–97.
- [102] V. Sandfort, K. Yan, P. M. Graffy, P. J. Pickhardt, and R. M. Summers, “Use of Variational Autoencoders with Unsupervised Learning to Detect Incorrect Organ Segmentations at CT,” *Radiology: Artificial Intelligence*, vol. 3, no. 4, p. e200218, Jul. 2021.
- [103] X. Chen, L. Yao, T. Zhou, J. Dong, and Y. Zhang, “Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images,” *Pattern Recognition*, vol. 113, p. 107826, May 2021.
- [104] J. Li, G. Zhao, Y. Tao, P. Zhai, H. Chen, H. He, and T. Cai, “Multi-task contrastive learning for automatic CT and X-ray diagnosis of COVID-19,” *Pattern Recognition*, vol. 114, p. 107848, Jun. 2021.
- [105] Y. Ye, J. Zhang, Z. Chen, and Y. Xia, “DeSD: Self-Supervised Learning with Deep Self-Distillation for 3D Medical Image Segmentation,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13434, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 545–555.
- [106] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “VoxelMorph: A Learning Framework for Deformable Medical Image Registration,” *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019.
- [107] M. Amirian, J. A. Montoya-Zegarra, J. Gruss, Y. D. Stebler, A. S. Bozkir, M. Calandri, F. Schwenker, and T. Stadelmann, “PrepNet: A Convolutional Auto-Encoder to Homogenize CT Scans for Cross-Dataset Medical Image Analysis,” Oct. 2021.

- [108] Y. N. Truong Vu, T. Tsue, J. Su, and S. Singh, “An improved mammography malignancy model with self-supervised learning,” in *Medical Imaging 2021: Computer-Aided Diagnosis*, Online Only, United States, 2021, p. 29.
- [109] J. Liu, L. Cao, O. Akin, and Y. Tian, “Robust and accurate pulmonary nodule detection with self-supervised feature learning on domain adaptation,” *Front. Radio*, vol. 2, p. 1041518, Dec. 2022.
- [110] M. Zheng, B. Planche, X. Gong, F. Yang, T. Chen, and Z. Wu, “Self-supervised 3D Patient Modeling with Multi-modal Attentive Fusion,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13437, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 115–125.
- [111] W.-C. Wang, E. Ahn, D. Feng, and J. Kim, “A Review of Predictive and Contrastive Self-supervised Learning for Medical Images.” arXiv, 09-Feb-2023.
- [112] F. C. Ghesu, B. Georgescu, A. Mansoor, Y. Yoo, D. Neumann, P. Patel, R. S. Vishwanath, J. M. Balter, Y. Cao, S. Grbic, and D. Comaniciu, “Self-supervised Learning from 100 Million Medical Images.” arXiv, 04-Jan-2022.
- [113] Z. Li, Z. Cui, S. Wang, Y. Qi, X. Ouyang, Q. Chen, Y. Yang, Z. Xue, D. Shen, and J.-Z. Cheng, “Domain Generalization for Mammography Detection via Multi-style and Multi-view Contrastive Learning,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12907, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 98–108.
- [114] K. You, S. Lee, K. Jo, E. Park, T. Kooi, and H. Nam, “Intra-class Contrastive Learning Improves Computer Aided Diagnosis of Breast Cancer in Mammography,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13433, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 55–64.
- [115] J. D. Miller, V. A. Arasu, A. X. Pu, L. R. Margolies, W. Sieh, and L. Shen, “Self-Supervised Deep Learning to Enhance Breast Cancer Detection on Screening Mammography.” arXiv, 15-Mar-2022.
- [116] J. Hu, Z. Feng, Y. Mao, J. Lei, D. Yu, and M. Song, “A Location Constrained Dual-Branch Network for Reliable Diagnosis of Jaw Tumors and Cysts,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12907, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 723–732.
- [117] A. Taleb, C. Rohrer, B. Bergner, G. De Leon, J. A. Rodrigues, F. Schwendicke, C. Lippert, and J. Krois, “Self-Supervised Learning Methods for Label-Efficient Dental Caries Classification,” *Diagnostics*, vol. 12, no. 5, p. 1237, May 2022.
- [118] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, and R. Collins, “UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age,” *PLoS Med*, vol. 12, no. 3, p. e1001779, Mar. 2015.
- [119] R. Windsor, A. Jamaludin, T. Kadir, and A. Zisserman, “Self-supervised Multi-modal Alignment for Whole Body Medical Imaging,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12902, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 90–101.

- [120] H. Zheng, J. Han, H. Wang, L. Yang, Z. Zhao, C. Wang, and D. Z. Chen, "Hierarchical Self-supervised Learning for Medical Image Segmentation Based on Multi-domain Data Aggregation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12901, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 622–632.
- [121] X. Zhang, S. Feng, Y. Zhou, Y. Zhang, and Y. Wang, "SAR: Scale-Aware Restoration Learning for 3D Tumor Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12902, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 124–133.
- [122] B. Liu, L.-M. Zhan, and X.-M. Wu, "Contrastive Pre-training and Representation Distillation for Medical Visual Question Answering Based on Radiology Images," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12902, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 210–220.
- [123] Z. Chen, Y. Du, J. Hu, Y. Liu, G. Li, X. Wan, and T.-H. Chang, "Multi-modal Masked Autoencoders for Medical Vision-and-Language Pre-training," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13435, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 679–689.
- [124] D. I. Kim, "PREFACE: How Dangerous Are X-ray Studies That We Undertake Every Day?," *J Korean Med Sci*, vol. 31, no. Suppl 1, p. S2, 2016.
- [125] M. Uffmann and C. Schaefer-Prokop, "Digital radiography: The balance between image quality and required radiation dose," *European Journal of Radiology*, vol. 72, no. 2, pp. 202–208, Nov. 2009.
- [126] R. Zha, Y. Zhang, and H. Li, "NAF: Neural Attenuation Fields for Sparse-View CBCT Reconstruction," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13436, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 442–452.
- [127] W. Roger, P. Lambin, S. Keek, M. Beuque, S. Primakov, E. Lavrova, G. Wu, C. Yan, S. Sanduleanu, H. Gietema, R. Casale, M. Occhipinti, A. Jochems, and H. Woodruff, "Towards Texture Accurate Slice Interpolation of Medical Images Using PixelMiner," In Review, preprint, Jun. 2021.
- [128] C.-K. Ta, A. Aich, A. Gupta, and A. K. Roy-Chowdhury, "Poisson2Sparse: Self-supervised Poisson Denoising from a Single Image," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13438, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 557–567.
- [129] J. Xu and E. Adalsteinsson, "Deformed2Self: Self-supervised Denoising for Dynamic Medical Imaging," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, vol. 12902, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 25–35.
- [130] Barone, Federico, "Compressing Medical Images with Minimal Information Loss," Master in High Performance Computing.
- [131] N. Dong, M. Kampffmeyer, and I. Voiculescu, "Learning Underrepresented Classes from Decentralized Partially Labeled Medical Images," in *Medical Image*

- Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13438, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 67–76.
- [132] Warren, Charles, “An Analysis of Compr An Analysis of Compressive Convolutional A olutional Autoencoders for oencoders for Image Archiving In Medical Informatics,” MASTER’S THESES, Michigan Technological University, United States, 2022.
- [133] Kwon,Young Joon (Fred), Toussie,Danielle, Reina,G Anthony, Tang,Ping Tak Peter, Doshi,Amish H, and Oermann, Eric K, “Autoencoder Image Compression Algorithm for Reduction of Resource Requirements,” presented at the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada., Vancouver, Canada, 2020, p. 5.
- [134] K. W. Altman, G.-P. Yu, and S. D. Schaefer, “Consequence of Dysphagia in the Hospitalized Patient: Impact on Prognosis and Hospital Resources,” *Arch Otolaryngol Head Neck Surg*, vol. 136, no. 8, p. 784, Aug. 2010.
- [135] P. Rouzrokh, B. Khosravi, S. Faghani, M. Moassefi, D. V. Vera Garcia, Y. Singh, K. Zhang, G. M. Conte, and B. J. Erickson, “Mitigating Bias in Radiology Machine Learning: 1. Data Handling,” *Radiology: Artificial Intelligence*, vol. 4, no. 5, p. e210290, Sep. 2022.
- [136] G. Varoquaux and V. Cheplygina, “Machine learning for medical imaging: methodological failures and recommendations for the future,” *npj Digit. Med.*, vol. 5, no. 1, p. 48, Apr. 2022.
- [137] I. Martinović, S. Mao, M. Dousty, W. Li, M. Đukanović, E. Colak, and E. Sejdić, “X-ray Modalities and Self-Supervised Learning: Tackling Small Sample Sizes and Uncurated Datasets,” 2023.
- [138] S.-C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, “Self-supervised learning for medical image classification: a systematic review and implementation guidelines,” *npj Digit. Med.*, vol. 6, no. 1, p. 74, Apr. 2023.
- [139] R. Krishnan, P. Rajpurkar, and E. J. Topol, “Self-supervised learning in medicine and healthcare,” *Nat. Biomed. Eng*, vol. 6, no. 12, pp. 1346–1352, Aug. 2022.
- [140] S. Azizi, L. Culp, J. Freyberg, B. Mustafa, S. Baur, S. Kornblith, T. Chen, N. Tomasev, J. Mitrović, P. Strachan, S. S. Mahdavi, E. Wulczyn, B. Babenko, M. Walker, A. Loh, P.-H. C. Chen, Y. Liu, P. Bavishi, S. M. McKinney, J. Winkens, A. G. Roy, Z. Beaver, F. Ryan, J. Krogue, M. Etemadi, U. Telang, Y. Liu, L. Peng, G. S. Corrado, D. R. Webster, D. Fleet, G. Hinton, N. Houlsby, A. Karthikesalingam, M. Norouzi, and V. Natarajan, “Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging,” *Nat. Biomed. Eng*, vol. 7, no. 6, pp. 756–779, Jun. 2023.
- [141] A. Jabri, A. Owens, and A. A. Efros, “Space-Time Correspondence as a Contrastive Random Walk.” arXiv, 03-Dec-2020.
- [142] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation.” arXiv, 18-May-2015.
- [143] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation.” arXiv, 18-Jul-2018.
- [144] K. Shaga Devan, H. A. Kestler, C. Read, and P. Walther, “Weighted average ensemble-based semantic segmentation in biological electron microscopy images,” *Histochem Cell Biol*, vol. 158, no. 5, pp. 447–462, Nov. 2022.

- [145] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, p. 80, Dec. 1945.
- [146] M. Krzywinski and N. Altman, "Nonparametric tests," *Nat Methods*, vol. 11, no. 5, pp. 467–468, May 2014.
- [147] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *Int J Comput Vis*, vol. 128, no. 2, pp. 336–359, Feb. 2020.
- [148] M.-Y. Hsiao, C.-H. Weng, Y.-C. Wang, S.-H. Cheng, K.-C. Wei, P.-Y. Tung, J.-Y. Chen, C.-Y. Yeh, and T.-G. Wang, "Deep Learning for Automatic Hyoid Tracking in Videofluoroscopic Swallow Studies," *Dysphagia*, vol. 38, no. 1, pp. 171–180, Feb. 2023.
- [149] H.-I. Kim, Y. Kim, B. Kim, D. Y. Shin, S. J. Lee, and S.-I. Choi, "Hyoid Bone Tracking in a Videofluoroscopic Swallowing Study Using a Deep-Learning-Based Segmentation Network," *Diagnostics*, vol. 11, no. 7, p. 1147, Jun. 2021.
- [150] Y. Ariji, M. Gotoh, M. Fukuda, S. Watanabe, T. Nagao, A. Katsumata, and E. Ariji, "A preliminary deep learning study on automatic segmentation of contrast-enhanced bolus in videofluorography of swallowing," *Sci Rep*, vol. 12, no. 1, p. 18754, Nov. 2022.
- [151] R. Robison, L. DiBiase, A. Ashley, K. McElheny, A. Anderson, J. P. Wymer, and E. K. Plowman, "Swallowing Safety and Efficiency Impairment Profiles in Individuals with Amyotrophic Lateral Sclerosis," *Dysphagia*, vol. 37, no. 3, pp. 644–654, Jun. 2022.
- [152] J. A. Logemann, "Critical Factors in the Oral Control Needed for Chewing and Swallowing," *Journal of Texture Studies*, vol. 45, no. 3, pp. 173–179, Jun. 2014.
- [153] C. M. Steele, M. Peladeau-Pigeon, A. Nagy, and A. A. Waito, "Measurement of Pharyngeal Residue From Lateral View Videofluoroscopic Images," *J Speech Lang Hear Res*, vol. 63, no. 5, pp. 1404–1415, May 2020.
- [154] Z. Zhang, E. Lucatoro, J. Coyles, and E. Sejdic, "Deep Learning-Based Auto-Segmentation and Evaluation of Vallecular Residue in Videofluoroscopy," *SSRN Journal*, 2021.
- [155] R. Leonard, "Two Methods for Quantifying Pharyngeal Residue on Fluoroscopic Swallow Studies: Reliability Assessment," 2017.
- [156] C. M. Steele, M. Peladeau-Pigeon, C. A. E. Barbon, B. T. Guida, A. M. Namasivayam-MacDonald, W. V. Nascimento, S. Smaoui, M. S. Tapson, T. J. Valenzano, A. A. Waito, and T. S. Wolkin, "Reference Values for Healthy Swallowing Across the Range From Thin to Extremely Thick Liquids," *J Speech Lang Hear Res*, vol. 62, no. 5, pp. 1338–1363, May 2019.
- [157] W. Li, S. Mao, A. S. Mahoney, J. L. Coyle, and E. Sejdic, "Automatic Tracking of Hyoid Bone Displacement and Rotation Relative to Cervical Vertebrae in Videofluoroscopic Swallow Studies Using Deep Learning," *J Digit Imaging. Inform. med.*, vol. 37, no. 4, pp. 1922–1932, Feb. 2024.
- [158] C. M. Steele, M. Peladeau-Pigeon, E. Barrett, and T. S. Wolkin, "The Risk of Penetration–Aspiration Related to Residue in the Pharynx," *Am J Speech Lang Pathol*, vol. 29, no. 3, pp. 1608–1617, Aug. 2020.
- [159] C. M. Steele, R. Mukherjee, J. M. Kortelainen, H. Pölönen, M. Jedwab, S. L. Brady, K. B. Theimer, S. Langmore, L. F. Riquelme, N. B. Swigert, P. M. Bath, L. B. Goldstein, R. L. Hughes, D. Leifer, K. R. Lees, A. Meretoja, and N. Muehleman, "Development of a Non-invasive Device for Swallow Screening in

- Patients at Risk of Oropharyngeal Dysphagia: Results from a Prospective Exploratory Study,” *Dysphagia*, vol. 34, no. 5, pp. 698–707, Oct. 2019.
- [160] W. G. Pearson, S. M. Molfenter, Z. M. Smith, and C. M. Steele, “Image-based Measurement of Post-Swallow Residue: The Normalized Residue Ratio Scale,” *Dysphagia*, vol. 28, no. 2, pp. 167–177, Jun. 2013.
- [161] R. W. Mulheren, Y. Inamoto, C. A. Odonkor, Y. Ito, S. Shibata, H. Kagaya, M. Gonzalez-Fernandez, E. Saitoh, and J. B. Palmer, “The Association of 3-D Volume and 2-D Area of Post-swallow Pharyngeal Residue on CT Imaging,” *Dysphagia*, vol. 34, no. 5, pp. 665–672, Oct. 2019.

Biografija



Mr Ivan Martinović rođen je 03. avgusta 1995. godine na Cetinju, gdje je završio Osnovnu školu „Njegoš“ i Cetinjsku Gimnaziju kao dobitnik diplome „Luča“ i priznanja Prijestonice Cetinje. Tokom osnovnog i srednjoškolskog obrazovanja intenzivno se bavio košarkom, sve do upisa na fakultet.

Elektrotehnički fakultet u Podgorici, odsjek Elektronika, telekomunikacije i računari, upisao je 2014. godine. U toku studija bio je polaznik većeg broja ljetnjih škola, naučnih konferencija, panel diskusija, kao i IAESTE studentske prakse na Fakultetu elektrotehnike, računarstva i informacijskih tehnologija Sveučilišta u Osijeku. Nakon završetka osnovnih studija, kao jedan od najboljih studenata, upisuje postdiplomske studije gdje stiče zvanje magistra. Specijalistički i magistarski rad realizovao je pod mentorstvom prof. dr Zorana Mijanovića i prof. dr Rade Dragović-Ivanović na istraživačkim projektima koji su našli primjenu u industriji. Ljeta 2019. godine učestvovao je na međunarodnom istraživačkom kampu “Summer Camp 2019” u organizaciji kompanije Ericsson Nikola Tesla Zagreb, Hrvatska. Svoj prvi radni angažman zasnovao je na Univerzitetu Crne Gore u prvom crnogorskom centru izvrsnosti BIO-ICT, kao saradnik u istraživanju. Krajem 2019. godine izabran je od strane Agencije za kontrolu i obezbjeđenje kvaliteta Visokog obrazovanja na listi eksperata iz reda studenata, koji se angažuju u komisijama za sprovođenje postupka akreditacije i reakreditacije ustanova visokog obrazovanja u Crnoj Gori. Od maja 2020 do upisa na doktorske studije radno je bio angažovan kao inženjer medicinske opreme u kompaniji Glosarij, biznis partneru njemačke kompanije Siemens Healthiness.

Doktorske studije upisuje u decembru 2020. godine na Univerzitetu Crne Gore kao stipendista Vlade Crne Gore. Doktorsko istraživanje realizovao je na temu „*Primjena vještačke inteligencije za automatsku analizu fluoroskopskih rendgenskih snimaka*“ pod mentorstvom prof. dr Milene Đukanović i prof. dr Ervina Sejdića. Tokom studija realizovao je studijske posjete Univerzitetu u Toronto, njemačkom Centru za istraživanje vještačke inteligencije, Silicijumskoj dolini, te je pohađao ljetnje škole na Univerzitetu u Oksfordu i Institut Necker Univerziteta Pariz Cité u Parizu. Od strane Evropskog instituta za inovacije i tehnologiju odabran je među 50 doktoranada iz Evrope za sprovođenje istraživanja na Aarhus Univerzitetu u Danskoj.

Pored naučno-istraživačkih aktivnosti, posvećen je promociji nauke kao izvršni direktor Crnogorskog naučno tehnološkog hub-a epiSTEMe i rukovodilac departmana za elektroniku i računarske nauke u Fondaciji PRONA. Aktivno je učestvovao u projektu Škola za 21. vijek (British Council), dok je trenutno angažovan od strane Ministarstva prosvjete, nauke i inovacija kao član radne grupe za izradu plana rada EdTech laboratorije/hub-a u crnogorskim obrazovnim institucijama. Dio je tima Montenegro

Makers fondacije koji je pokrenuo prvu robotičku ligu za osnovce u Crnoj Gori. Saraduje sa Vladom Crne Gore u kreiranju prvog programa obuke na temu vještačke inteligencije za državne službenike i namještenike. Dodatno, Ivan Martinović je bio mentor crnogorskom timu na robotičkoj olimpijadi First Global Challenge 2024.

Autor je 18 naučnih publikacija objavljenih u međunarodnim časopisima i konferencijama. Saradnik je Crnogorske akademije nauka i umjetnosti na izradi Leksikona zdravstva Crne Gore. Dodatno, angažovan je kao recezent u nekoliko vodećih međunarodnih naučnih časopisa i konferencija. Učestvovao je na preko 35 međunarodnih-nacionalnih, naučnih, edukativnih i inovacionih projekata. Član je IEEE-a i fokus grupe ITU-a za vještačku inteligenciju u zdravstvu. Dobitnik je studentske 13. Novembarske nagrade Prijestonice Cetinje (2023) za izuzetne rezultate postignute tokom studija, kao i nagrade Ministarstva prosvjete nauke i inovacija za naučna dostignuća u 2024. godini – Specijalno priznanje za značajan doprinos promociji nauke među mladima. Od strane Crnogorskog udruženja za vještačku inteligenciju nagrađen je priznanjem za najbolji naučni rad u oblasti vještačke inteligencije za akademske godine 2023/24 i 2024/25. Nosilac je jednog od najznačajniji svjetskih priznanja, WSIS Prize 2024 Champion dodijeljeno od strane Međunarodne Unije za Telekomunikacije, krovne organizacije Ujedinjenih nacija za digitalne tehnologije.

IZJAVA O AUTORSTVU

Ime i prezime autora: Ivan Martinović

Broj indeksa/upisa: 5/20

IZJAVLJUJEM

da je doktorska disertacija pod naslovom:

Primjena vještačke inteligencije za automatsku analizu fluoroskopskih rendgenskih snimaka

- rezultat sopstvenog istraživačkog rada,
- da predložena disertacija ni u cjelini ni u djelovima nije bila predložena za dobijanje bilo koje diplome prema studijskim programima drugih ustanova visokog obrazovanja,
- da su rezultati korektno navedeni, i
- da nijesam povrijedio autorska i druga prava intelektualne svojine koja pripadaju trećim licima.

Cetinje,
Jun, 2025. godine



mr Ivan Martinović

IZJAVA O ISTOVJETNOSTI ŠTAMPANE I ELEKTRONSKE VERZIJE DOKTORSKOG RADA

Ime i prezime autora: Ivan Martinović

Broj indeksa/upisa: 5/20

Studijski program: Doktorske studije elektrotehnike

Naslov rada: *Primjena vještačke inteligencije za automatsku analizu fluoroskopskih rendgenskih snimaka*

Mentori: prof. dr Milena Đukanović i akademik prof. dr Ervin Sejdić

Izjavljujem da je štampana verzija mog doktorskog rada istovjetna elektronskoj verziji koju sam predao za objavljivanje u Digitalni arhiv Univerziteta Crne Gore.

Istovremeno izjavljujem da dozvoljavam objavljivanje mojih ličnih podataka u vezi sa dobijanjem akademskog naziva doktora nauka, kao što su ime i prezime, godina i mjesto rođenja, naziv doktorske disertacije i datum odbrane rada.

Cetinje,
Jun, 2025. godine



mr Ivan Martinović

IZJAVA O KORIŠĆENJU

Ovlašćujem Univerzitetsku biblioteku da u Digitalni arhiv Univerziteta Crne Gore pohrani moju doktorsku disertaciju pod naslovom:

Primjena vještačke inteligencije za automatsku analizu fluoroskopskih rendgenskih snimaka.

koja je moje autorsko djelo.

Disertaciju sa svim prilogima predao sam u elektronskom formatu pogodnom za trajno arhiviranje.

Moju doktorsku disertaciju pohranjenu u Digitalni arhiv Univerziteta Crne Gore mogu da koriste svi koji poštuju odredbe sadržane u odabranom tipu licence Kreativne zajednice (Creative Commons) za koju sam se odlučio.

1. Autorstvo
2. Autorstvo – nekomercijalno
- 3. Autorstvo – nekomercijalno – bez prerade**
4. Autorstvo – nekomercijalno – dijeliti pod istim uslovima
5. Autorstvo – bez prerade
6. Autorstvo – dijeliti pod istim uslovima

Cetinje,
Jun, 2025. godine



mr Ivan Martinović

1. **Autorstvo.** Licenca sa najširim obimom prava korišćenja. Dozvoljavaju se prerade, umnožavanje, distribucija i javno saopštavanje djela, pod uslovom da se navede ime izvornog autora (onako kako je izvorni autor ili davalac icence odredio). Djelo se može koristiti i u komercijalne svrhe.
2. **Autorstvo - nekomercijalno.** Dozvoljavaju se prerade, umnožavanje, distribucija i javno saopštavanje djela, pod uslovom da se navede ime izvornog autora (onako kako je izvorni autor ili davalac licence odredio). Komercijalna upotreba djela nije dozvoljena.
3. **Autorstvo - nekomercijalno - bez prerade.** Licenca koju se u najvećoj mjeri ograničavaju prava korišćenja djela. Dozvoljava se umnožavanje, distribucija i javno saopštavanje djela, pod uslovom da se navede ime izvornog autora (onako kako je izvorni autor ili davalac licence odredio). Djelo se ne može mijenjati, preoblikovati ili koristiti u drugom djelu. Komercijalna upotreba djela nije dozvoljena.
4. **Autorstvo - nekomercijalno - dijeliti pod istim uslovima.** Dozvoljava se umnožavanje, distribucija, javno saopštavanje i prerada djela, pod uslovom da se navede ime izvornog autora (onako kako je izvorni autor ili davalac licence odredio). Ukoliko se djelo mijenja, preoblikuje ili koristi u drugom djelu, prerada se mora distribuirati pod istim ili sličnom licencom. Djelo i prerade se ne mogu koristiti u komercijalne svrhe.
5. **Autorstvo - bez prerade.** Dozvoljava se umnožavanje, distribucija i javno saopštavanje djela, pod uslovom da se navede ime izvornog autora (onako kako je izvorni autor ili davalac licence odredio). Djelo se ne može mijenjati, preoblikovati ili koristiti u drugom djelu. Licenca dozvoljava komercijalnu upotrebu djela.
6. **Autorstvo - dijeliti pod istim uslovima.** Dozvoljava se umnožavanje, distribucija i javno saopštavanje djela, pod uslovom da se navede ime izvornog autora (onako kako je izvorni autor ili davalac licence odredio). Ukoliko se djelo mijenja, preoblikuje ili koristi u drugom djelu, prerade se moraju distribuirati pod istim ili sličnom licencom. Ova licenca dozvoljava komercijalnu upotrebu djela i prerada. Slična je softverskim licencama, odnosno licencama otvorenog koda.

