

## **HRZZ Uspostavni istraživački projekti (UIP-09-2014)**

Voditelj: Alan Jović

Ustanova: Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva

### **Programski sustav za paralelnu analizu više heterogenih nizova vremenskih podataka s primjenom u biomedicini (UIP-2014-09-6889)**

Isporuka D1.1.

## **Dokumentacija programskog rješenja**

#### **Autori isporuke:**

- ❖ dr. sc. Alan Jović
- ❖ prof. dr. sc. Mario Cifrek
- ❖ dr. sc. Marko Horvat
- ❖ Krešimir Jozić, dipl. ing.

U Zagrebu, 22. 2. 2016.

## Sadržaj

1. Uvod .....	3
2. Zahtjevi na sustav .....	4
2.1. Pregled temeljnih zahtjeva .....	4
2.2. UML-dijagrami obrazaca uporabe za vizualizaciju zahtjeva sustava .....	5
2.2.1. Odabir vrste analize .....	6
2.2.2. Odabir scenarija .....	7
2.2.3. Odabir ulaznih podataka .....	8
2.2.4. Pregled zapisa .....	9
2.2.5. Predobrada zapisa .....	10
2.2.6. Izlučivanje značajki .....	11
2.2.7. Izgradnja modela .....	12
2.2.8. Izvještavanje .....	14
2.2.9. Korisnički račun .....	15
2.2.10. Administracija platforme .....	16
3. Arhitektura platforme .....	17
3.1. Tehnologije .....	17
3.2. Unutarnje značajke platforme .....	18

## **1. Uvod**

Ovaj dokument predoduje dokumentaciju programskog rješenja, kakva je razrađena u uvodnom dijelu projekta, prema radnom planom za aktivnost A1.1. Dokumentacija uključuje opis zahtjeva na sustav, njihovo grafičko predodavanje te kratki konceptualni opis arhitekture programskog rješenja u smislu korištenih tehnologija i nekih unutarnjih značajki sustava. Namjera dokumentacije je da predodi najbolji pristup u okviru razmatranih pristupa oblikovanju programskog rješenja. Dokumentacija u ovoj uvodnoj fazi ne uključuje opis ostvarenja (implementacije) projekta.

## 2. Zahtjevi na sustav

### 2.1. Pregled temeljnih zahtjeva

Zahtjevi na sustav, koji su dani u projektnoj prijavi i kasnije usuglašeni na prvom sastanku radne skupine projekta, dana 16. 10. 2016. su ukratko sljedeći:

- Cjelovito programsko rješenje za analizu multivarijatnih heterogenih biomedicinskih vremenskih nizova
- Programsko rješenje ostvareno kao web platforma
- Programska logika na poslužitelju pisana u programskom jeziku Java
- Sučelje prema klijentu ostvareno skupom internetskih tehnologija (HTML5, CSS3, TypeScript, ...)
- Više ulaznih formata datoteka: EDF/EDF+, .txt i .ann, slikovni formati
- Predobrada podataka: filtriranje signala, detekcija R zubaca, transformacija podataka: vremenska, frekvencijska, vremensko-frekvencijska
- Izlučivanje značajki – ekspertno odabrane ili po želji korisnika, potrebno je podržati vrlo veliki broj domenskih i općih značajki biomedicinskih vremenskih nizova
- Odabir značajki, klasifikacijski i predikcijski algoritmi strojnog učenja za izgradnju modela
- Izvještavanje o rezultatima

Od biomedicinskih vremenskih nizova, platforma treba nužno podržati:

- EKG – elektrokardiogram – kardiovaskularna aktivnost
- HRV – niz RR-intervalu (srčani otkucaji), promatra se varijabilnost srčanog ritma
- EEG – elektroencefalogram – moždana aktivnost
- EMG – elektromiogram – mišićna aktivnost

Dodatno, po mogućnosti, podržat će se još i:

- EOG, EGG, CTG, ECoG, PPG, tlak (ABP, PAP, CVP), respiracija (impedancija), SO<sub>2</sub>, CO<sub>2</sub>, ritam koračanja, otpor kože...

Potrebno je podržati heterogenu multivarijatnu analizu (npr. EKG + HRV + SO<sub>2</sub>)

Težnja je demonstrirati da platforma najprije radi na manjem skupu problema, a zatim je nadograđivati za veći skup, uz zadržavanje razine kvalitete usluge.

## 2.2. UML-dijagrami obrazaca uporabe za vizualizaciju zahtjeva sustava

Zahtjevi na sustav razrađeni su u obliku UML-dijagrama obrazaca uporabe (engl. *UML Use Case Diagram*), v.1.4+, koji prikazuju kako će se korisnik koristiti sa sustavom. Ovi dijagrami daju ponašajni opis korištenja sustava, no bez mogućnosti uvida u detaljni vremenski redoslijed kojim se interakcija sa sustavom odvija. Bez obzira na nedostatak vremenskog pogleda, razmatranjem ponašanja korisnika identificiraju se svi scenariji korištenja, što nam omogućuje lakši razvoj web platforme. Korisnik s platformom interagira kroz web preglednik, a mogućnosti njegove interakcije uspješno opisuju UML-dijagrami obrazaca uporabe.

Dijagrami su podijeljeni u osam dijelova, koji opisuju faze analize:

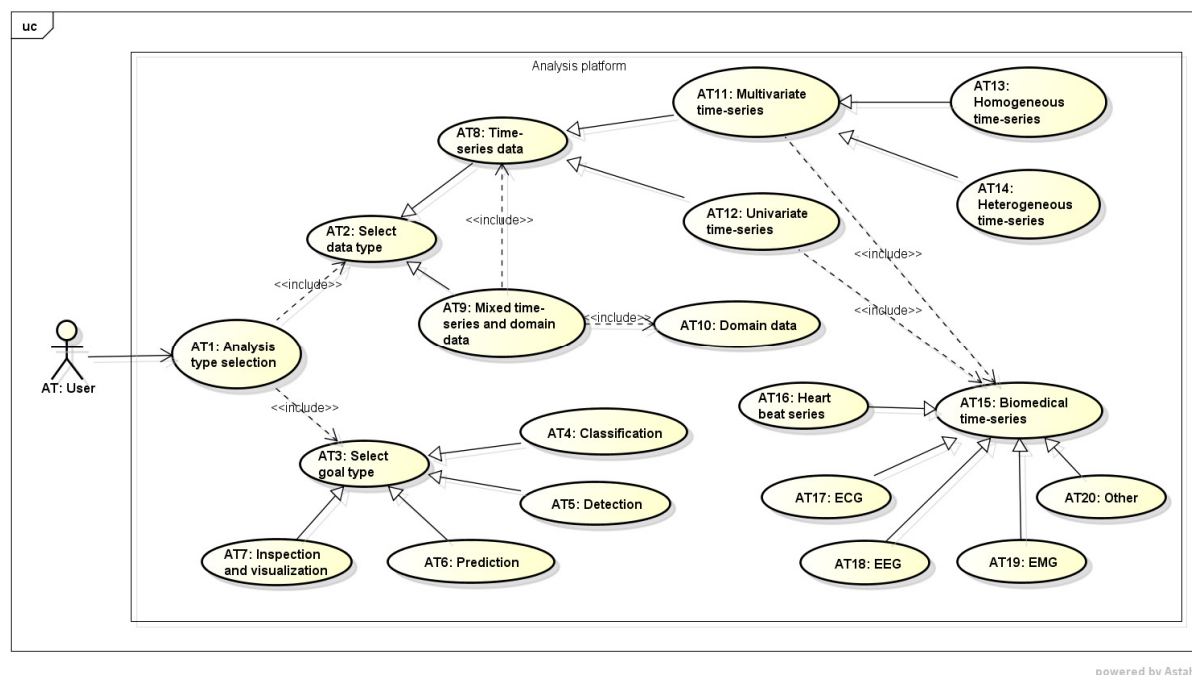
1. Odabir vrste analize (engl. *Analysis type selection*)
2. Odabir scenarija (engl. *Scenario selection*)
3. Odabir ulaznih podataka (engl. *Input data selection*)
4. Pregled zapisa (engl. *Records inspection*)
5. Predobrada zapisa (engl. *Records preprocessing*)
6. Izlučivanje značajki (engl. *Feature extraction*)
7. Izgradnja modela (engl. *Model construction*)
8. Izvještavanje (engl. *Reporting*)

Dodatno, 9. dijagram, Korisnički račun (engl. *User account*), opisuje registraciju, prijavu u sustav i upravljanje svojim profilom.

Konačno, 10. dijagram, Administracija platforme (engl. *Platform administration*) opisuje administratorski pristup platformi.

U nastavku su prikazani dijagrami i dan je njihov detaljniji opis.

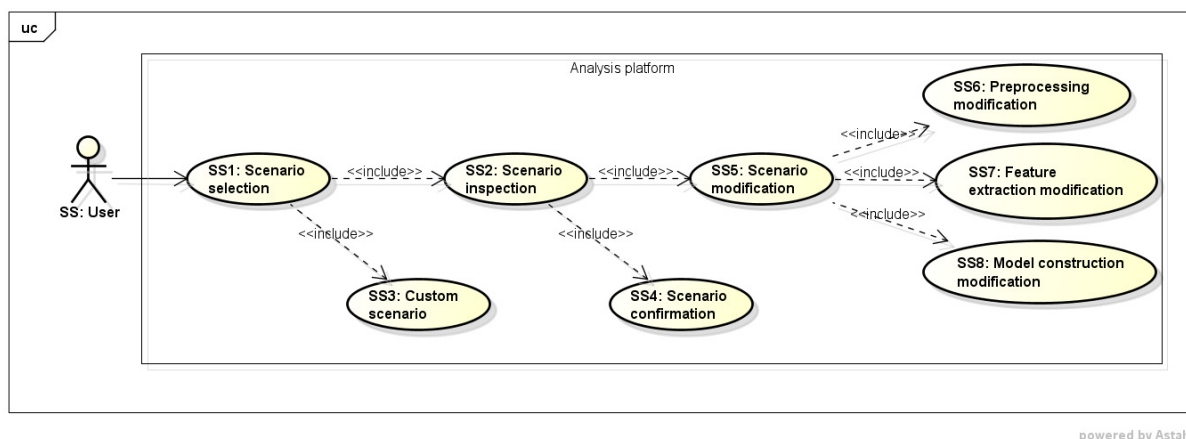
## 2.2.1. Odabir vrste analize



Slika 1. Odabir vrste analize (engl. *Analysis type selection*)

Odabir vrste analize (slika 1) provodi se tako da korisnik izabere cilj analize koju želi provesti (detekcija, klasifikacija, predikcija, pregled i vizualizacija) i tip podataka koji će analiza uključivati. Tip podataka može uključivati isključivo biomedicinske vremenske nizove ili biomedicinske vremenske nizove skupa s dodatnim domenskim podacima (npr. anamneza subjekta, metapodaci o vrsti zapisa i sl.). Biomedicinski vremenski nizovi mogu biti po tipu univarijatni, što znači da imaju samo jedan niz izmjerenih podataka, npr. vremena srčanih otkucaja ili multivarijatni, što znači da imaju u jednom zapisu više nizova izmjerenih podataka. Dodatno, multivarijatni nizovi mogu biti homogeni (od jednog tipa, npr. EEG podaci) ili heterogeni (od više tipova, npr. EKG + srčani otkucaji + sistolički tlak). Tipovi nizova koji su podržani platformom su svi navedeni u poglavlju 2.1.

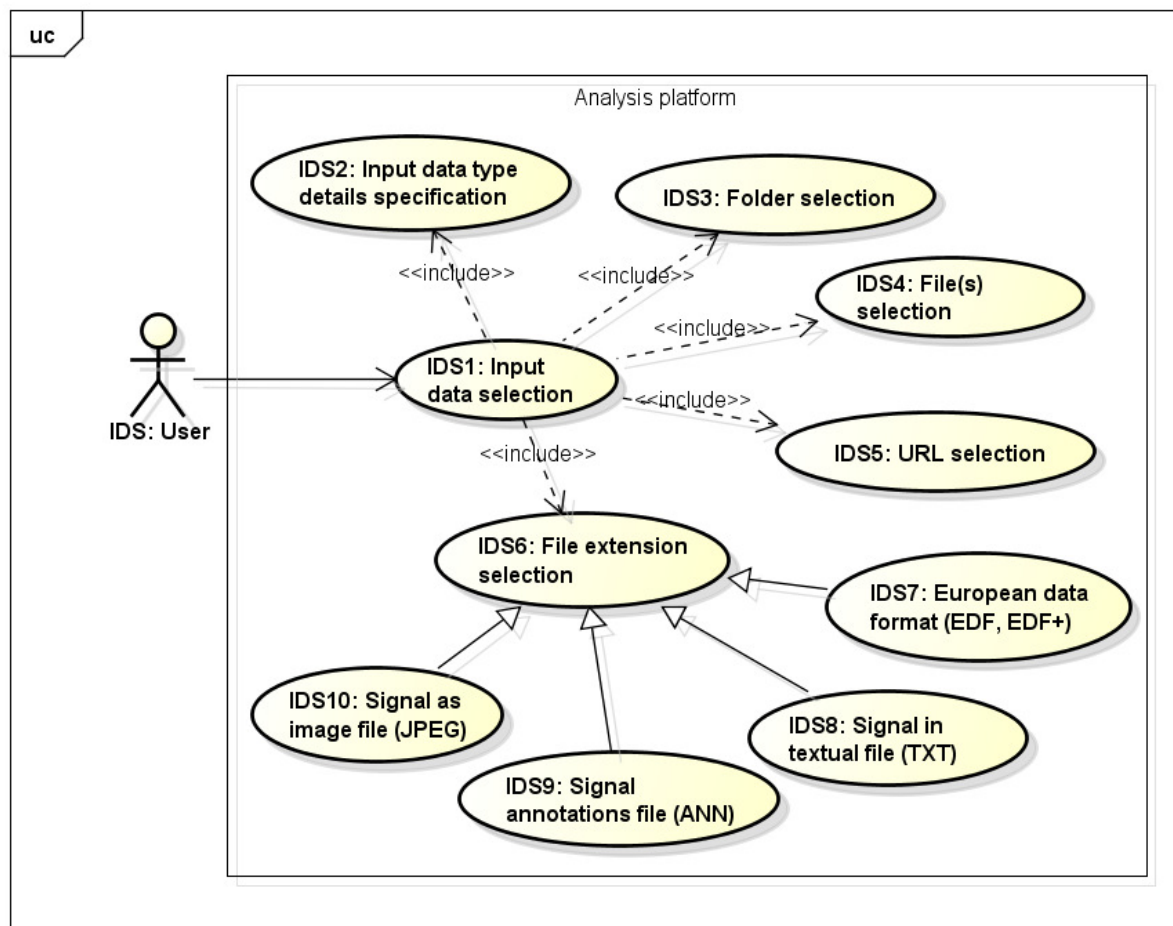
## 2.2.2. Odabir scenarija



Slika 2. Odabir scenarija (engl. *Scenario selection*)

Odabir scenarija (slika 2) omogućuje korisniku da izabere već ponuđene scenarije analize, s obzirom na prethodno definirani tip analize ili da definira svoj vlastiti scenarij. Korisniku se nudi popis mogućih scenarija analize sa svim njihovim karakteristikama. Korisnik može potvrditi određeni scenarij ili izmijeniti neku od karakteristika scenarija, kao što su metode predobrade signala, metode izlučivanja značajki i metode izgradnje modela. Pri definiranju vlastitog scenarija, kao i pri izmjeni nekog od postojećih scenarija, korisnik će moći odabrati koje metode želi koristiti u ostvarenju svojeg scenarija. Izbor metoda za izlučivanje značajki u okviru nekog scenarija bit će ostvaren pomoću ugrađenog ekspertnog sustava za preporuku značajki, a sve na temelju relevantnih znanstvenih radova i smjernica u području medicine i biomedicinskog inženjerstva.

### 2.2.3. Odabir ulaznih podataka



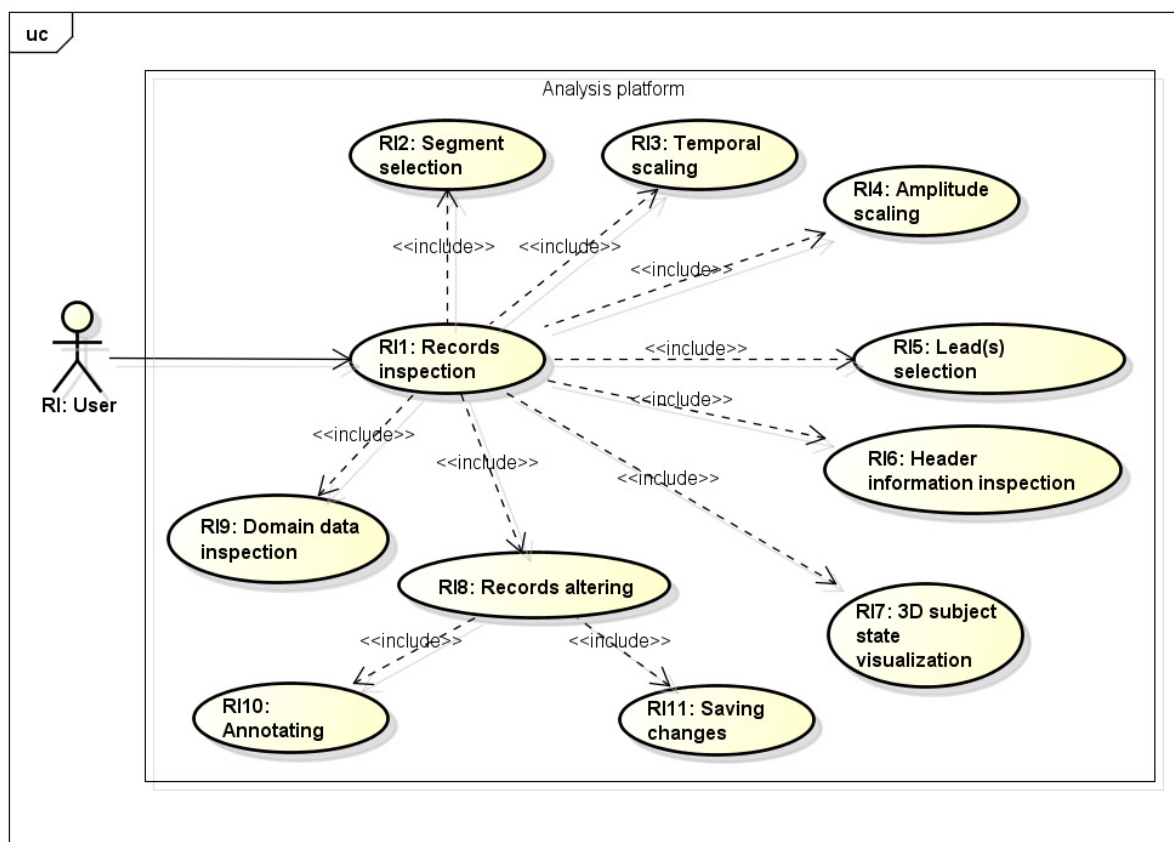
powered by Astah

Slika 3. Odabir ulaznih podataka (engl. *Input data selection*)

Odabir ulaznih podataka (slika 3) uključuje izbor ekstenzija (formata) datoteka koje se uzimaju u obzir prilikom analize, a one uključuju EDF, EDF+, anotacijske datoteke, signale (bez anotacija) u tekstualnom obliku ili slikovne zapise signala (npr. JPEG, TIFF). Korisnik će trebati izabrati odgovarajući format datoteka kako bi učitavanje zapisa u platformu prošlo bez pogrešaka. Osim odabira formata, korisnik specificira datoteke koje se učitavaju u platformu ili određuje direktorij u kojem se nalaze sve potrebne datoteke. Ako postoje specifični detalji vezani uz učitavanje zapisa (npr. učitavanje samo nekih dijelova datoteka, ograničenja za slikovne zapise i sl.) to također korisnik izabire prilikom ove faze.



## 2.2.4. Pregled zapisa

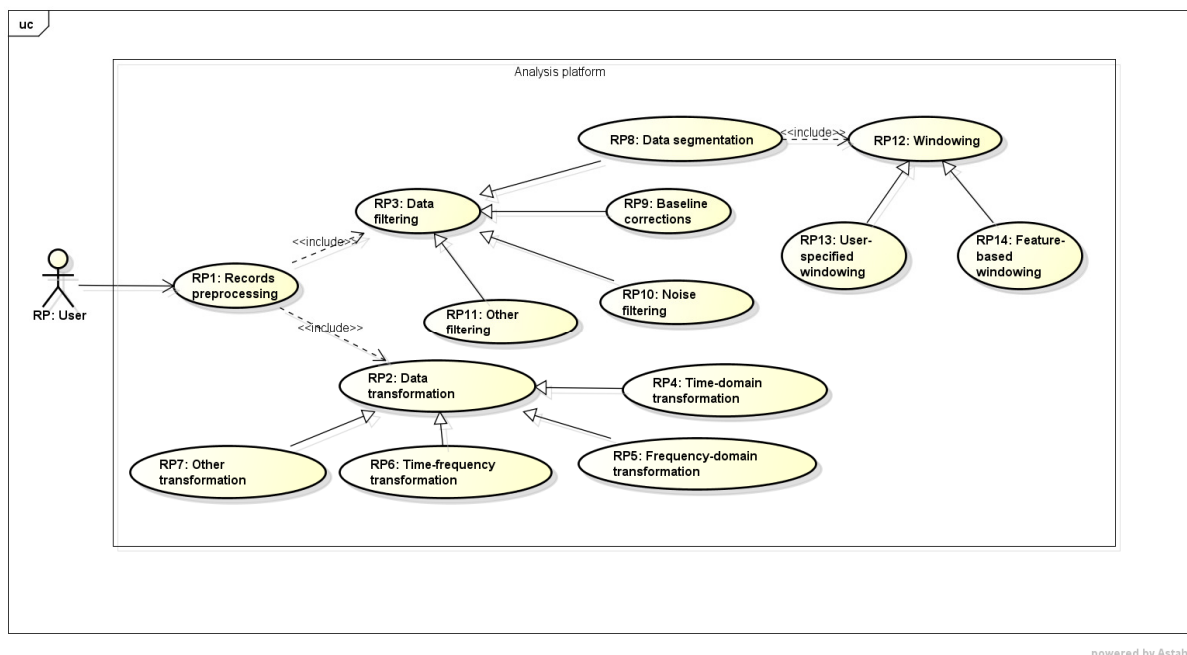


powered by Astah

Slika 4. Pregled zapisa (engl. *Records inspection*)

Nakon učitavanja, bit će moguće svaki učitani zapis promotriti i proanalizirati (slika 4) kroz vizualno sučelje. Pritom će se omogućiti odabir određenog segmenta zapisa, vremensko skaliranje, amplitudno skaliranje, odabir tragova signala, pregled zaglavlja zapisa i pregled domenskih podataka (npr. anamneze subjekta, meta-podataka o vrsti zapisa i sl.), ako oni postoje. 3D vizualizacija stanja subjekta bit će omogućena za neka određena stanja u okviru određenih područja primjene (npr. vizualizacija lokacije infarkta miokarda na temelju EKG-a). 3D vizualizacija će se ostvariti korištenjem WebGL tehnologije na strani klijenta. Dodatno, korisnik će moći anotirati neke signale ili izmijeniti postojeće anotacijske datoteke i te promjene u zapisu pohraniti.

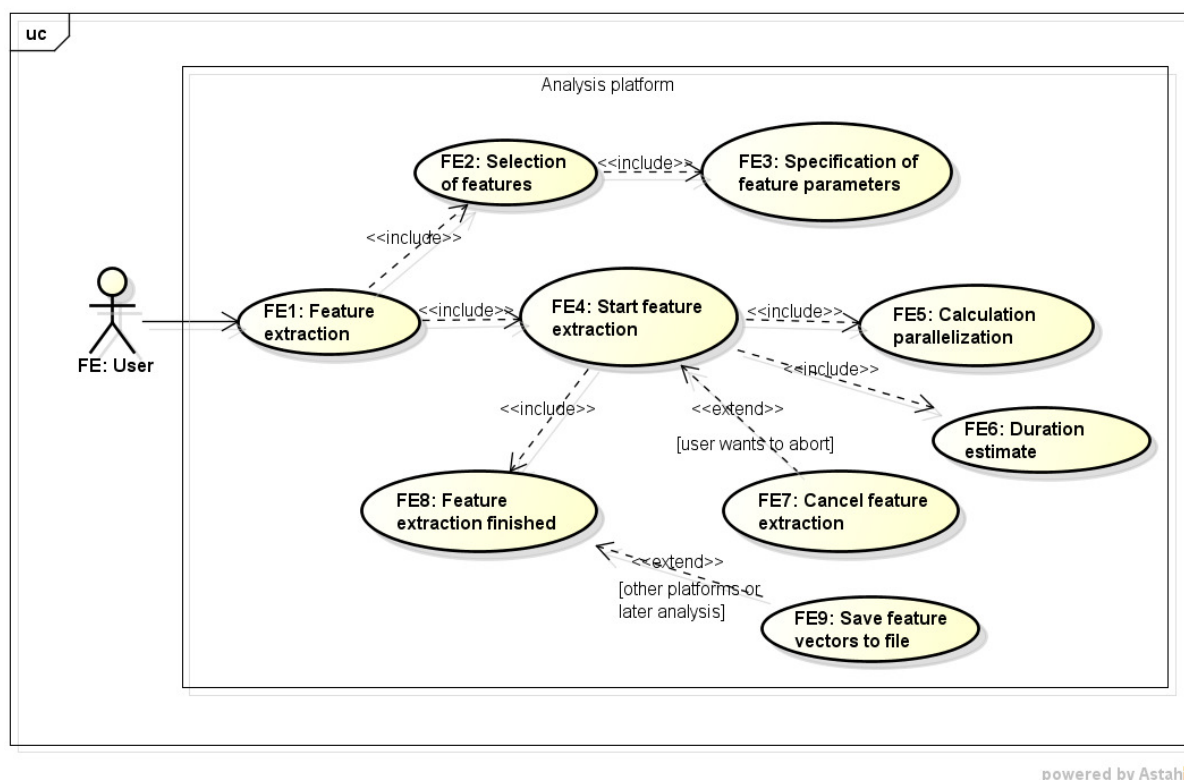
## 2.2.5. Predobrada zapisa



Slika 5. Predobrada zapisa (engl. *Records preprocessing*)

Predobrada zapisa (slika 5) uključuje različite postupke filtriranja zapisa i transformacije nad podacima. Filtriranje uključuje uklanjanje šuma (npr. *notch* filter za 50 i 60 Hz), korekcije bazne linije signala (npr. kod EKG-a nastale zbog disanja ili pokreta tijela), segmentaciju zapisa na prozore određene širine za potrebe daljnje analize. Širinu prozora može odabrati korisnik ili je ona određena samim značajkama koje se izlučuju (npr. jedan RR-interval). Ostali postupci filtriranja podataka uključuju različite linearne ili nelinearne postupke filtriranja u cilju dobivanja kvalitetnijeg zapisa. Transformacije nad podacima u zapisu rade se nakon postupaka filtriranja i uključuju različite vremenske (npr. PCA), frekvencijske (npr. FFT, Hilbert-Huangova transformacija), vremensko-frekvencijske (npr. transformacija valićima) i druge vrste transformacija. Cilj transformacije podataka je dobiti podatke u obliku iz kojeg se lakše izračunavaju precizne domenske ili opće značajke za opis određenog stanja subjekta.

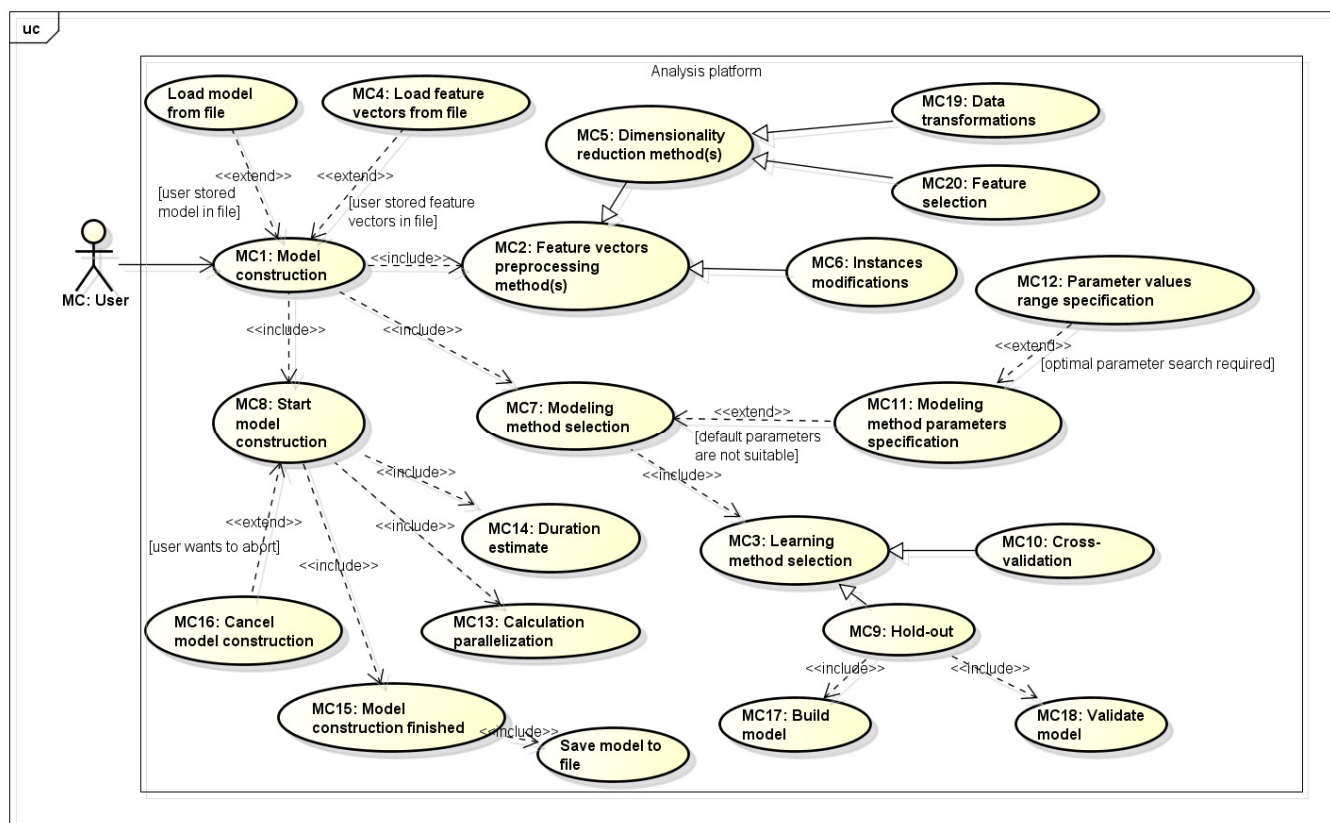
## 2.2.6. Izlučivanje značajki



Slika 6. Izlučivanje značajki (engl. *Feature extraction*)

Izlučivanje značajki (slika 6) središnji je korak u analizi biomedicinskih vremenskih nizova. Korisniku se već pri izboru scenarija nudi popis značajki koje će se izlučiti. Bez obzira na sam scenarij, u fazi izlučivanja značajki, moguće je izabrati dodatne značajke i specificirati parametre pri izlučivanju pojedine značajke, za one značajke koje su parametarske. Prilikom pokretanja izlučivanja značajki, korisniku će se omogućiti pregled podataka o procjeni trajanja analize i izbor vrste paralelizacije računanja koja se provodi. Korisnik će moći u svakom trenutku prekinuti izlučivanje značajki. Ako želi, korisnik može po završetku analize pohraniti vektore značajki u datoteku za kasnije korištenje ili samo nastaviti s analizom.

## 2.2.7. Izgradnja modela



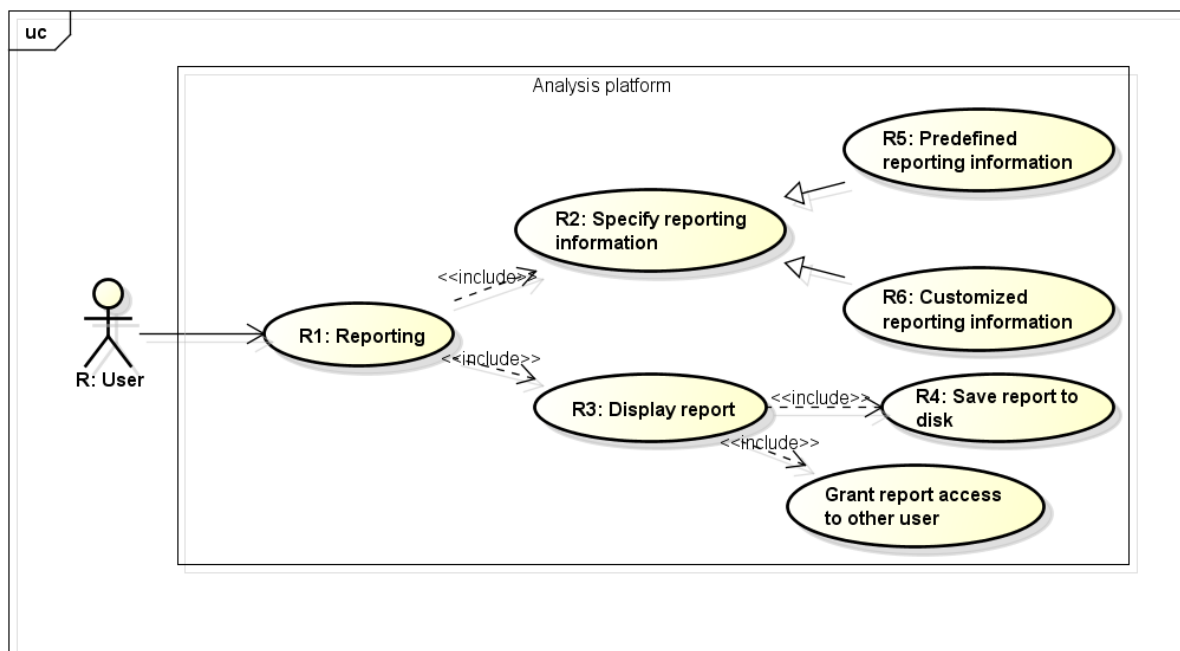
powered by Astah

Slika 7. Izgradnja modela (engl. *Model construction*)

Izgradnja modela (slika 7) složeni je korak analize biomedicinskih nizova koji nastupa nakon izlučivanja značajki i koji uključuje različite postupke strojnog učenja za obradu izlučenih vektora značajki. Korisnik će moći učitati zapis s vektorima značajki koje je već ranije izračunao i pohranio ili će moći nastaviti s analizom tek nedavno dobivenog i nepohranjenog zapisa. Pri izgradnji modela najprije je omogućen odabir metoda predobrade vektora značajki. Tu je moguće koristiti metode za smanjenje dimenzionalnosti vektora značajki kao što su različite transformacije (npr. PCA) ili metode odabira značajki kojima se nastoji zadržati u skupu značajki samo one koje su relevantne i neredundantne. U tu svrhu mogu se koristiti metode filtera, omotača, ugrađene, hibridne i druge metode. Smanjenje dimenzionalnosti je potrebno u slučaju postojanja velikog broja izlučenih značajki (bilo izborom korisnika bilo automatskom preporukom ekspertnog sustava). Osim smanjenja dimenzionalnosti prostora značajki, moguće je koristiti i neke postupke izmjene samih vektora značajki (npr. ponovno uzorkovanje, popunjavanje nedostajućih vrijednosti i slični).

Nakon predobrade zapisa slijedi odabir metode za izgradnju modela. Pritom se odabire naziv metode i definiraju se po potrebi vrijednosti parametara metode, ako je metoda za izgradnju modela parametarska (npr. stablo odluke C4.5). Prije početka izgradnje modela, potrebno je izabrati metodu učenja koja se koristi: postupak izdvajanja dijela skupa podataka za ispitivanje ili postupak unakrsne validacije. Moguće je i samo provesti ispitivanje modela nad novim primjerima, ako već postoji naučeni model od ranije. Početak izgradnje modela omogućit će korisniku pregled podataka o procjeni trajanja analize i izbor vrste paralelizacije računanja koja se provodi. Po završetku izgradnje modela bit će omogućeno izgrađeni model pohraniti u datoteku.

## 2.2.8. Izvještavanje

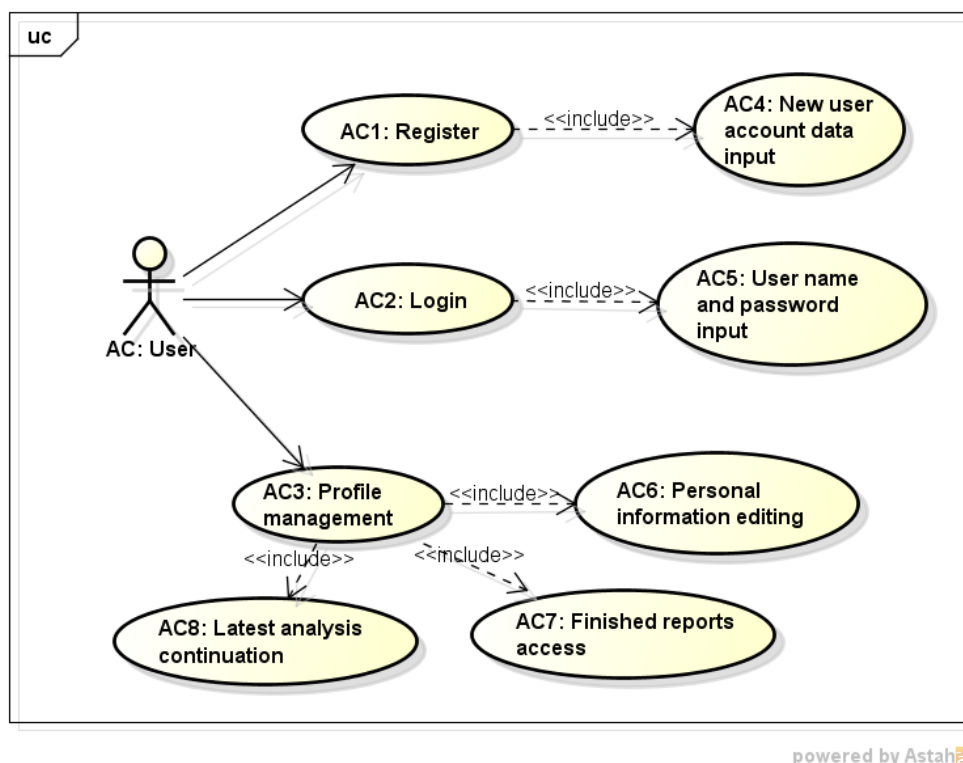


powered by Astah

Slika 8. Izvještavanje (engl. *Reporting*)

Izvještavanje (slika 8) je korak u kojem korisnik bira način izvještavanja, a zatim mu se izvještaj prikazuje u odgovarajućem obliku (na web stranici, u PDF, Excel, i/ili Word/ODF obliku). Bit će omogućena pohrana izvještaja na lokalno računalo, a kopija izvještaja će biti pohranjena i na poslužitelju. Dodatno, korisnik će moći omogućiti pristup izvještaju i drugim registriranim korisnicima sustavima kojima to želi.

### 2.2.9. Korisnički račun



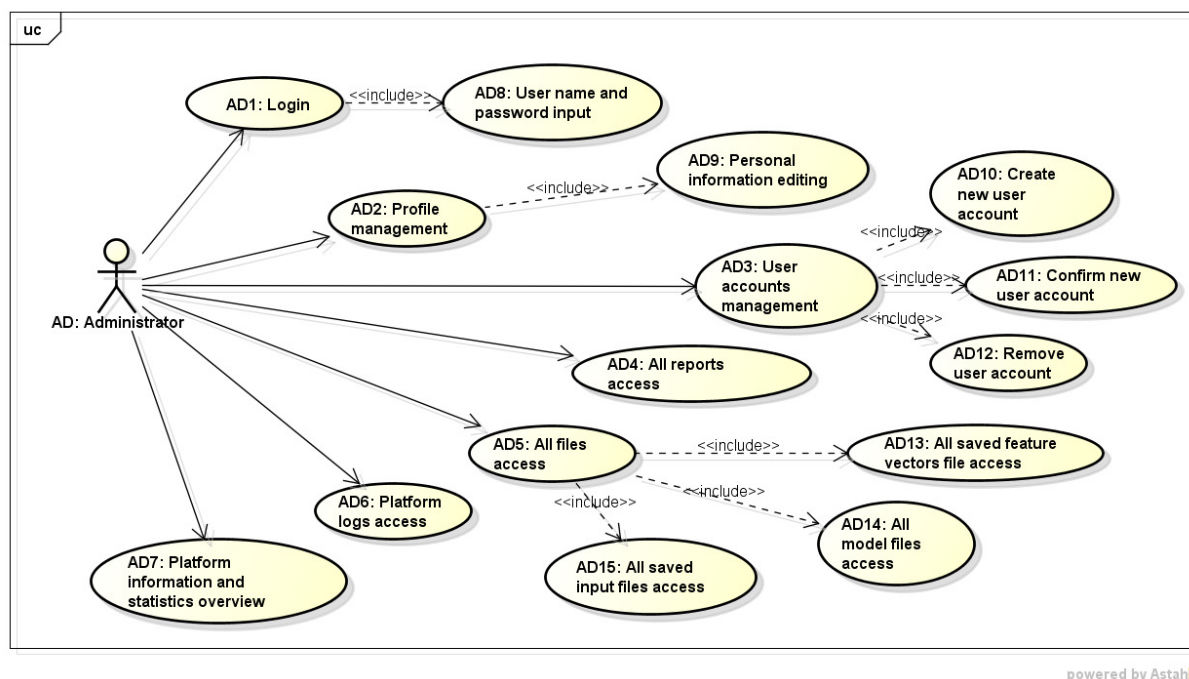
Slika 9. Korisnički račun (engl. *User account*)

Račun za novog korisnika otvarat će se registracijom korisnika na platformu, koja će uključivati unos podataka o novom korisniku (osnovni podaci: korisničko ime, lozinka i e-mail, dodatni podaci: ime i prezime, zvanje, ustanova, adresa ustanove, telefon), slika 9.

Prijava na sustav bit će omogućena nakon potvrde da je registracija uspješno provedena.

Upravljanje profilom bit će omogućeno nakon uspješne prijave u sustav, a ono će uključivati izmjenu korisničkih podataka, pristup generiranim izvještajima koji su povezani s korisnikom ili nastavak zadnje analize koju je korisnik provodio.

## 2.2.10. Administracija platforme



Slika 10. Administracija platforme (engl. *Platform administration*)

Administrator platforme će se trebati prijaviti na platformu na isti način kao i korisnik, unoseći svoje korisničko ime i lozinku (slika 10). Na raspolaganju će mu biti uređivanje njegovih osobnih informacija profila. Administrator će moći upravljati korisničkim računima ostalih korisnika platforme, što će uključivati otvaranje novog korisničkog računa, potvrđivanje registracije novog korisničkog računa i uklanjanje postojećeg korisničkog računa. Za svaki račun koji postoji u platformi, administrator će moći vidjeti sve osobne informacije (osim lozinke), ali ih neće moći mijenjati.

Administrator će moći pregledati i pretraživati sva napravljena izvješća u platformi. Također, bit će mu omogućen pristup do svih datoteka pohranjenih u sustavu, uključujući sve datoteke zapisa pacijenata koje su pohranjene na poslužitelju, sve datoteke s vektorima značajki i sve datoteke s izgrađenim i snimljenim modelima. Osim toga, administrator će moći pristupiti svim dnevnicima platforme koji će biti poredani po datumu. U dnevnicima će biti pohranjene sve relevantne informacije o akcijama korisnika tijekom korištenja platforme. Informacije o platformi i pregled statistike korištenja platforme bit će mu također na raspolaganju.



## 3. Arhitektura platforme

### 3.1. Tehnologije

Cilj projekta je napraviti kvalitetan sustav za obradu i analizu biomedicinskih vremenskih nizova. Predviđeno je da ovu obradu radi medicinsko osoblje, biomedicinski inženjeri ili računarski znanstvenici, ovisno o njihovom cilju. Medicinsko osoblje može iskoristiti platformu za dobivanje moguće dijagnoze za određeni poremećaj ili stanje organizma i to na temelju već postojećeg znanja pohranjenog u platformi, dok ga ostali istraživači mogu koristiti za znanstvene pokuse nad većim brojem zapisa raznih subjekata.

Arhitektura sustava osmišljena je kao web portal, pri čemu će korisnici pristupati portalu putem svojeg web preglednika. Ova arhitektura je odabrana, budući da je potrebno omogućiti pristup platformi iz bilo kojeg udaljenog mjesta. Time se maksimalno širi baza potencijalnih korisnika, što je bitno za prepoznatljivost platforme i širinu primjene.

S klijentske (korisničke) strane, pri izradi platforme koristit će se jezici HTML5, CSS3 i Typescript za oblikovanje web stranica. Kao pomoć pri razvoju web stranica koristit će se radni okviri Angular 2 za programski dio i Bootstrap za ljepši vizualni dio i bolju responzivnost platforme. Također, bit će podržana tehnologija WebGL za 3D vizualizaciju određenih dijelova organizma koji su od interesa za korisnika.

Na poslužiteljskoj strani koristit će se programski jezik Java. Ovaj jezik je odabran za poslužiteljsku stranu zbog velikog broja postojećih knjižnica za obradu signala, parsiranje podataka, strojno učenje i paralelizaciju računanja. Budući da će korisničko sučelje biti prikazano u web pregledniku, nije potrebno koristiti Javin aplikacijski poslužitelj sa svim svojim značajkama. Bit će potrebno iskoristiti samo manji broj komponenata Javinog aplikacijskog poslužitelja, kao što je RESTful poslužitelj za uspješnu komunikaciju s klijentskom stranom, knjižnica za ovjeru korisnika i knjižnica za objektno-relacijsko preslikavanje (npr. Hibernate), što je nužno za pohranu Javinih objekata u relacijsku bazu podataka koju će platforma koristiti.

Na poslužiteljskoj strani isprobat će se dvije varijante: 1) standardni Javin aplikacijski poslužitelj s minimalnim skupom opcija i 2) samostojeći radni okvir koji podržava tražene funkcionalnosti (npr. Spring Boot). Obje varijante će se isprobati i ispitati s obzirom na brzinu izvođenja i korištenje memorijskih resursa.

Za izradu izvještaja koristit će se slobodno dostupna knjižnica JasperReports® Library.

### 3.2. Unutarnje značajke platforme

Razvijena arhitektura treba biti modularna i skalabilna kako bi podržala istovremeno izvođenje za veći broj korisnika. S tim u svezi potrebno je podržati dvije značajne karakteristike programske potpore: 1) paralelizaciju i 2) dinamičko učitavanje modula po potrebi.

Paralelizacija služi za ubrzanje izračunavanja, a posebno je učinkovita za one algoritme koji se bez većih problema mogu raščlaniti u jednostavnije postupke koji nemaju potrebu za međusobnom komunikacijom. Paralelizacija će biti omogućena u dvije faze analize: za izlučivanje značajki i za izgradnju modela. U oba slučaja pokušat će se maksimalno iskoristiti kapaciteti poslužitelja u vidu zauzeća procesorskih jezgri, ali i grafičke kartice. Paralelizacija korištenjem grafičke kartice za podatke koji inače nisu grafički (GPGPU paralelizacija) postići će se korištenjem JCuda ili jocl radnih okvira, ovisno o dostupnom sklopovlju. Vrstu paralelizacije će moći eksplicitno izabrati i korisnik, pri čemu će sustav sam procijeniti ima li na raspolaganju takve resurse kakve korisnik traži.

Dinamičko učitavanje modula po potrebi je korisna značajka platforme, koja je značajna iz perspektive štednje resursa. Ako primjerice korisnik želi klasificirati EKG korištenjem stabla odluke C4.5, tada nema potrebe za učitavanjem u memoriju računala svih ostalih modula.

Ovo će se ostvariti tako što će se oblikovati moduli koji će imati visoku koheziju usluga unutar njih, a malu međuovisnost o ostalim modulima. Korisnik će biti u mogućnosti definirati scenarij izvođenja (ili izabrati neki već ponuđeni) u kojem će točno odrediti redoslijed modula (komponenti) koji se trebaju učitati da bi se postigao zadani cilj. Pomoćni paketi sadržavat će samo popis dostupnih komponenti za određenu fazu analize.

Platforma će sadržavati veliki broj značajki biomedicinskih vremenskih nizova, kako općih značajki, tako i domensko-specifičnih, tj. vezanih uz određenu vrstu vremenskog niza (npr. EKG, EEG). U platformi se planiraju preuzeti značajke iz radnih okvira HRVFrame, EEGFrame i Comp-Engine, uz prevođenje algoritama u programski jezik Java, prilagodbu i ispitivanja algoritama, tamo gdje je to potrebno.

---

dr. sc. Alan Jović, voditelj projekta

---

prof. dr. sc. Mario Cifrek, suradnik

---

dr. sc. Marko Horvat, suradnik

---

Krešimir Jozić, dipl. ing., suradnik