

Raspoznavanje visokorizičnih mentalnih stanja pilota na temelju fizioloških značajki i strojnog učenja

Palac, Laura

Master's thesis / Diplomski rad

2025

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Electrical Engineering and Computing / Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:168:408173>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-03-30**



Repository / Repozitorij:

[FER Repository - University of Zagreb Faculty of Electrical Engineering and Computing repository](#)



SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 760

**RASPOZNAVANJE VISOKORIZIČNIH MENTALNIH STANJA
PILOTA NA TEMELJU FIZIOLOŠKIH ZNAČAJKI I STROJNOG
UČENJA**

Laura Palac

Zagreb, veljača 2025.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 760

**RASPOZNAVANJE VISOKORIZIČNIH MENTALNIH STANJA
PILOTA NA TEMELJU FIZIOLOŠKIH ZNAČAJKI I STROJNOG
UČENJA**

Laura Palac

Zagreb, veljača 2025.

DIPLOMSKI ZADATAK br. 760

Pristupnica: **Laura Palac (0036510945)**
Studij: Računarstvo
Profil: Računalno inženjerstvo
Mentor: izv. prof. dr. sc. Siniša Popović

Zadatak: **Raspoznavanje visokorizičnih mentalnih stanja pilota na temelju fizioloških značajki i strojnog učenja**

Opis zadatka:

Proučiti izabranu literaturu o primjenama strojnog učenja u detekciji visokorizičnih mentalnih stanja pilota, kao što su visoka razina stresa, kognitivno preopterećenje, iznenadni prepad (engl. startle) uzrokovan neočekivanom ugrozom itd. Definirati metodologiju analize fizioloških signala primjenom metoda strojnog učenja u cilju raspoznavanja visokorizičnih stanja pilota u raspoloživom podatkovnom skupu snimljenom na simulatoru leta aviona. Eksploratornim i statističkim analizama utvrditi koje fiziološke značajke iz proučene literature i/ili vlastitog predloženog skupa značajki, izlučene iz snimljenih signala primjenom izabranih programskih alata, predstavljaju najsnažnije pojedinačne indikatore promatranih rizičnih stanja pilota. Provesti komparativne analize nekoliko popularnih metoda nadziranog strojnog učenja na problemima binarnog i višeklasnog raspoznavanja promatranih stanja pilota. U komparativnim analizama primijeniti uobičajene evaluacijske metrike modela strojnog učenja, kao što su konfuzijske matrice, točnost, preciznost, odziv, ROC krivulje itd.

Rok za predaju rada: 14. veljače 2025.

Sadržaj

Uvod	1
1. Osvrt na izabranu relevantnu literaturu	3
2. Raspoloživi podatkovni skup s visokorizičnim stanjima pilota	5
3. Metodologija analize i obrade podataka.....	8
3.1. Izračun značajki.....	9
3.2. Podjela podatkovnog skupa na skup za treniranje i testiranje	12
4. Metodologija strojnog učenja	13
5. Rezultati i diskusija	15
5.1. Rezultati metode stabla odluke.....	15
5.2. Rezultati metode slučajne šume	17
5.3. Rezultati KNN metode	18
Zaključak	21
Literatura	23
Sažetak.....	25
Summary.....	26

Uvod

Ljudska pogreška najčešći je uzrok zrakoplovnih nesreća, a mentalno stanje pilota može značajno utjecati na sigurnost leta. Čak 80 posto zrakoplovnih nesreća uzrokovano je ljudskom pogreškom (piloti, kontrolori zračnog prometa, mehaničari itd.), dok je 20 posto posljedica kvara na opremi [14]. Različiti faktori kao što su stres, povećano kognitivno opterećenje i neočekivane situacije mogu izazvati promjene u mentalnim sposobnostima pilota koje dovode do smanjenja koncentracije i ometaju proces donošenja odluka [9]. Razvoj sustava za prepoznavanje takvih visokorizičnih stanja u realnom vremenu može pružiti podršku pilotima te povećati sigurnost zrakoplovnog prometa.

Razvojem tehnika strojnog učenja i prikupljanjem obilnih količina fizioloških podataka omogućeno je provođenje naprednih analiza kognitivnih i emocionalnih stanja pilota tokom leta. U ovom radu istražuju se metode strojnog učenja za klasifikaciju mentalnih stanja pilota koristeći podatke objavljene na portalu *NASA's Open Data Portal*. Podaci su dobiveni korištenjem psihofizioloških senzora koji mjere fiziološke odgovore poput elektrodermalne aktivnosti (*engl. GSR – Galvic Skin Response*), elektrokardiograma (*engl. ECG - electrocardiogram*), elektroencefalograma (*engl. EEG - electroencephalogram*) i disanja pilota [5]. Cilj istraživanja je analizirati može li se pouzdano predvidjeti kada pilot ulazi u potencijalno rizično kognitivno-emocionalno stanje na temelju spomenutih fizioloških značajki. Stoga su u radu primijenjena dva pristupa u području podjele podataka na skup za trening modela te testni skup: stratificirana nasumična podjela i LOSO (*engl. Leave One Subject Out*) podjela. LOSO podjela testira sposobnost modela da generalizira na podacima novih pilota koje nije vidio tijekom treniranja. Također je analizirano kako korištenje samo odabranih fizioloških karakteristika (ECG, GSR i disanje) utječe na performanse modela. U radu su korištene tri metode strojnog učenja za klasifikaciju kognitivno-emocionalnih stanja pilota: stablo odluke (*engl. Decision Tree*), slučajna šuma (*engl. Random Forest*) i metoda k najbližih susjeda (*engl. KNN - K-Nearest Neighbors*). Također, koriste se dvije varijante klasifikacije, binarna koja razlikuje bazno stanje od svih drugih stanja, te višeklasna koja raspoznaje sva stanja. Za procjenu modela korištene su metrike točnosti, preciznosti i odziva te analiza matrica zabune kako bi se identificirale najčešće greške modela.

Rezultati ovog rada su u širem aplikativnom smislu vezani uz multidisciplinarno područje koje ima za cilj smanjenje zrakoplovnih nesreća i poboljšanje sigurnosti u zračnom prometu. Dobiveni rezultati mogu se analizirati za daljnji rad na razvoju budućih sustava koji automatski prepoznaju potencijalno rizična kognitivno-emocionalna stanja pilota, omogućujući pravovremene obavijesti vezane za ulazak pilota u rizično stanje te prilagodbu sustava potpore pilotu, što bi u konačnici moglo doprinijeti sigurnijem zračnom prometu.

1. Osvrt na izabranu relevantnu literaturu

Ovo poglavlje pruža pregled radova povezanih s ovim radom. Točnije, u pregledu radova se fokusiram na rad [1] koji je objavio Američki institut za aeronautiku i astronautiku te na diplomski rad [2] s Odjela za kognitivne znanosti i umjetnu inteligenciju Sveučilišta Tilburg u Nizozemskoj.

Rad [1] se bavi prepoznavanjem i predviđanjem kognitivnih stanja pilota tijekom simulacije leta pomoću psihofizioloških senzora. Cilj rada je razviti sustav koji u stvarnom vremenu prepoznaje ograničavajuća stanja ljudskih performansi povezana s pažnjom (*engl. AHPLS – attention-related human performace limiting states*) koja mogu utjecati na smanjenje svjesnosti o stanju aviona i dovesti do nesreće. Podaci koje rad koristi su prikupljeni na dvadeset i četiri pilota dok su obavljali različite zadatke u simulatoru leta noseći psihofiziološke senzore. Zadaci su uključivali zadatke odmora, zadatke osmišljene da izazovu AHPLS i uvjete velikog/malog radnog opterećenja te eksperimentalne scenarije leta. Psihofiziološki senzori primijenjeni na sudionicima mjerili su elektroencefalogramske signale (EEG), elektrokardiogram (ECG), disanje te galvanske reakcije kože (GSR). Osim toga, pratilo se i vizualno ponašanje te fiziološke reakcije očiju pilota. Također, nakon svakog zadatka, piloti su sudjelovali u tri upitnika za samoprocjenu. Dobiveni podaci analizirani su kako bi se dobile značajke za treniranje modela strojnog učenja. Analizirane su različite frekvencije EEG-a, varijacije srčanog ritma, promjene u disanju i razine provodljivosti kože. Razmatrane su različite metode strojnog učenja: slučajna šuma (*engl. Random Forest*), podizanje gradijenta (*engl. Gradient Boosting*) te stroj s potpornim vektorima (*engl. SVM – Support Vector Machine*). Modeli su obučeni na tzv. *benchmark* zadacima, odnosno zadacima odmora te različitim zadacima osmišljenima da izazovu AHPLS i uvjete velikog/malog kognitivnog opterećenja, dok su testirani na simulacijama leta. Točnost predviđanja stanja ovisila je o tipu klasifikatora, korištenim signalima, značajkama i metodama učenja. Najbolji rezultat, odnosno točnost od 95%, postignut je korištenjem značajki EEG-a, disanja i GSR-a.

Cilj rada [2] je prepoznati i ispravno klasificirati različita kognitivna stanja pilota koja utječu na donošenje odluka tokom leta i na taj način smanjiti broj nesreća uzrokovanih ljudskom greškom. Rad koristi podatke koje je *Kaggle* osigurao za sudionike natjecanja „*Reducing Commercial Aviation Fatalities*“. Podaci su prikupljeni na 18 pilota koji su bili

izloženi različitim zadacima tijekom kojih su nosili psihofiziološke senzore. Skup podataka za treniranje modela strojnog učenja je prikupljen tijekom tzv. *benchmark* zadataka, odnosno zadataka s kontroliranim eksperimentima izvan simulatora leta. Skup podataka za testiranje je prikupljen tijekom cjelovite simulacije leta, tj. polijetanja, leta i slijetanja. Rad analizira različite psihofiziološke značajke pilota tokom izvršavanja zadataka poput ECG-a, EEG-a, GSR-a te disanja. Stanja pilota koja se klasificiraju su usmjerenost pažnje (*engl. CA - Channelized Attention*), skretanje pažnje (*engl. DA – Diverted Attention*), iznenađenje (*engl. SS – Startle/Surprise*) te bazno stanje (*engl. BL - Baseline*). Kao metoda strojnog učenja koristi se metoda podizanja gradijenta. Model postiže F1 rezultat od 0.55 te tvrdi da fiziološke značajke ECG-a, disanja i GSR-a najviše doprinose prepoznavanju stanja u kojemu se pilot nalazi. Također, u radu se diskutira da neki piloti imaju znatno bolje rezultate u raspoznavanju njihovih stanja što ukazuje na individualne razlike u psihofiziološkim odgovorima pilota.

2. Raspoloživi podatkovni skup s visokorizičnim stanjima pilota

Podatkovni skup koji se koristi u ovom radu preuzet je s portala *NASA's Open Data Portal*. Psihofiziološki senzori bilježili su podatke od 18 pilota tokom izvođenja različitih zadataka u simulatoru leta i izvan njega. Cilj prikupljanja ovih podataka jest njihovo korištenje u treniranju modela strojnog učenja kako bi se prepoznala različita stanja pažnje pilota [5]. Cjelokupni skup podataka se sastoji od skupa za treniranje, koji obuhvaća podatke prikupljene tokom niza kontroliranih eksperimentalnih zadataka izvan simulatora leta, te skupa podataka za testiranje (*engl. LOFT – Line Oriented Flight Training*) prikupljenih tokom cjelovite simulacije leta (polijetanje, let i slijetanje). Međutim, ovaj rad koristi isključivo testni podatkovni skup za treniranje i za testiranje modela strojnog učenja. Više o podjeli LOFT skupa na skup za treniranje i testiranje različitih modela strojnog učenja opisano je u poglavlju 3.2. Ovi podaci korišteni su u *Kaggle* natjecanju „*Reducing Commercial Aviation Fatalities*“ kao i podaci u radu [2].

Piloti su tokom zadataka bili izloženi ometanjima s ciljem izazivanja jednog od četiriju kognitivnih, i u određenoj mjeri emocionalnih, stanja opisanih u tablici. LOFT skup sadrži cjelovitu simulaciju leta tijekom koje su piloti mogli iskusiti bilo koje od ovih stanja, ali nikada više od jednog stanja istovremeno.

Tablica 1 Kognitivno-emocionalna stanja pilota koja klasificiramo

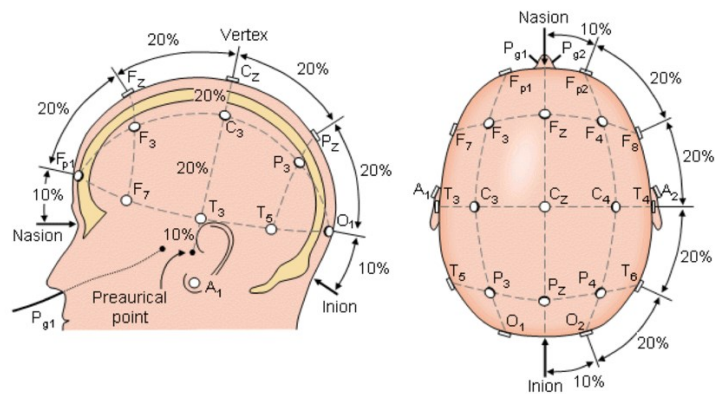
Stanje pilota	Opis	Oznaka/ Dogadaj
Bazno stanje	Stanje bez ometanja	0
Stanje usmjerene pažnje (<i>engl. CA – Channelized Attention</i>)	Stanje u kojem je pilot fokusiran na jedan zadatak te zbog toga ignorira sve ostale zadatke	2
Stanje preusmjerene pažnje (<i>engl. DA – Diverted Attention</i>)	Stanje u kojem je pažnja pilota preusmjerena zbog radnji ili misaonih procesa povezanih s donošenjem odluka	5

<i>Startle</i> /Iznenadenje (<i>engl. SS – Startle/Surprise</i>)	<i>Startle</i> je brza fiziološka reakcija na snažan iznenadni podražaj, a iznenadenje je emocionalni i kognitivni odgovor na širi skup iznenadnih ili neočekivanih situacija, uključujući i događaje koji izazivaju <i>startle</i> [9].	1
--	--	---

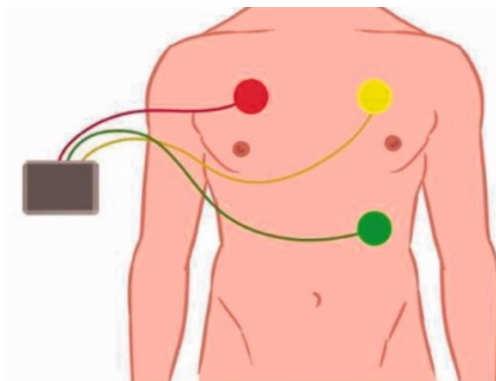
Kako je već gore napomenuto, piloti su za vrijeme obavljanja zadataka te same simulacije leta nosili psihofiziološke senzore koji su bilježili različite fiziološke signale pilota. Svaki senzor radio je s frekvencijom uzorkovanja od 256 Hz. Fiziološki signali pilota koji se prate su EEG, ECG, disanje i GSR.

Polja podataka u podatkovnom skupu [5]:

- TimeSecs – oznaka vremena od početka mjerenja u sekundama
- EEG_FP1, EEG_F7, EEG_F8, EEG_T4, EEG_T6, EEG_T5, EEG_T3, EEG_FP2, EEG_O1, EEG_P3, EEG_Pz, EEG_F3, EEG_Fz, EEG_F4, EEG_C4, EEG_P4, EEG_POz, EEG_C3, EEG_Cz, EEG_O2 – očitavanja različitih kanala EEG-a, alfanumerički znakovi nakon podvlake predstavljaju pozicije elektroda na glavi pilota prema slici (Slika 1).
- ECG – ECG signal električne aktivnosti srca kako je prikazano na slici (Slika 2), senzor razlučivosti od 0,012215 μV po bitu i raspona od -100 mV do +100 mV
- R – disanje (*engl. Respiration*), odnosno mjerenje podizanja i spuštanja prsnog koša, senzor razlučivosti od 0,2384186 μV po bitu i raspona od -2,0 V do +2,0 V
- GSR – mjera elektrodermalne aktivnosti, senzor razlučivosti od 0,2384186 μV po bitu i raspona od -2,0 V do +2,0 V.



Slika 1 Raspored senzora na glavi pilota[3]

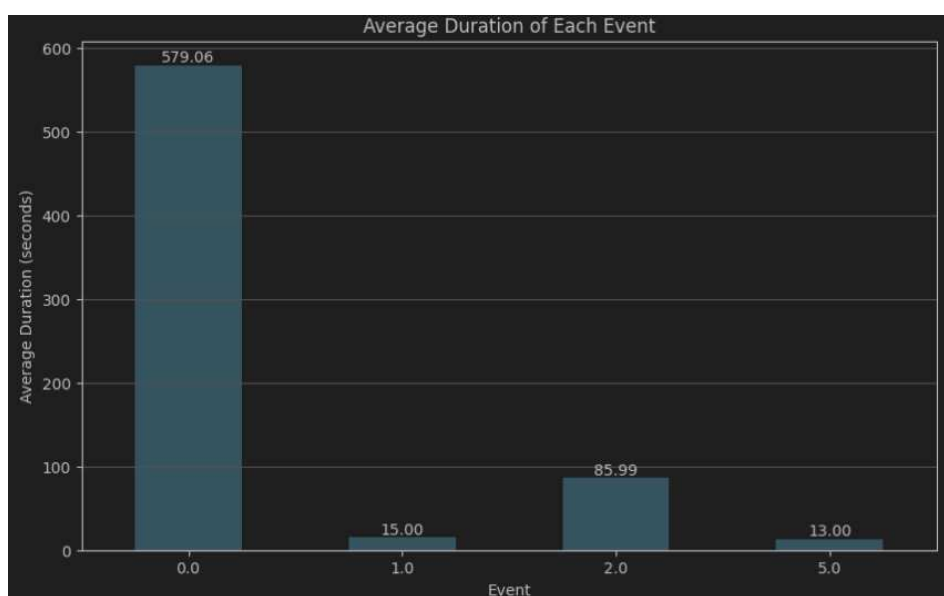


Slika 2 Prijenosni ECG[6]

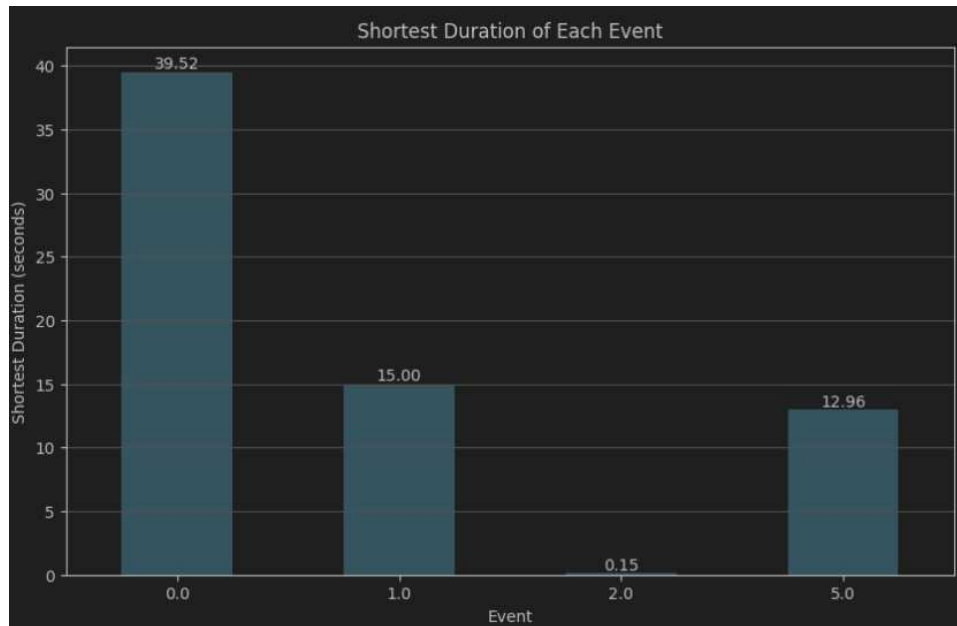
Podaci koji se odnose na EEG signale prikupljeni su pomoću sustava *B-Alert X24*, dok su podaci za ECG, disanje i GSR prikupljeni pomoću *NeXus-10* sustava [5]. Također, bitno je napomenuti da, budući da fiziološki podaci nisu umjetno generirani, već se radi o podacima stvarnih osoba, u samim signalima senzora je prisutan šum [5].

3. Metodologija analize i obrade podataka

Analiza podatkovnog skupa nužan je korak koji nam omogućuje razumijevanje osnovne strukture podataka. Kako je već spomenuto ovaj rad koristi isključivo LOFT skup, tj. skup podataka prikupljen tijekom cjelokupne simulacije leta. Time smo uštedili na vremenu potrebnom za obradu podataka i samoj količini podataka. Količina podataka dodatno je smanjena izdvajanjem prozora od 12 sekundi pri pojavi svakog događaja (0, 1, 2, i 5, kako prikazuje Tablica 1). Dakle, pri svakoj pojavi nekog događaja, ne uzima se njegovo cjelokupno trajanje, već samo prozor od 12 sekundi. Time smo osigurali bolje izbalansirani skup za treniranje i testiranje modela strojnog učenja u odnosu na uzimanje cijelog skupa podataka, što bi ujedno dalo ogroman broj uzoraka za trening s klasom 0 i onemogućilo trening bez primjene superračunala. Ukoliko dođe do pojave događaja koji traje manje od 12 sekundi, pojava tog događaja i prozor se zanemaruju. Graf 1 i Graf 2 prikazuju prosječno i najkraće trajanje svakog događaja. Prozor trajanja 12 sekundi izabran je na temelju analize najkraćeg trajanja, ali i uzimajući u obzir prosječno vrijeme trajanja svakog događaja. Najmanje prozora i podataka se gubi ako uzmemo 12 sekundi kao interval trajanja svakog događaja koji ćemo uzimati u razmatranje jer je to najkraće trajanje događaja, zanemarujući situacije gdje stanja traju djelić sekunde te ih ignoriramo kao stanja prebrze dinamike da bi se uopće registrirala u navedenim fiziološkim signalima (vidjeti Graf 2). Ovako gubimo 24 pojave događaja 2 jer toliko puta on traje manje od 12 sekundi. Nakon ovakve obrade podataka u dobivenom skupu podataka odnos klasa 0:1:2:5 je otprilike 1:0.31:0.45:0.68.



Graf 1 Srednje vrijeme trajanja događaja



Graf 2 Najkraće vrijeme trajanja događaja

Cilj rada je implementirati procjenu opisanih četiriju kognitivno-emocionalnih stanja pilota na temelju izračunatih značajki. Stanje pilota mijenja se kao odgovor na događaje koji se pojavljuju u simulatoru. Procjene se izvode u vremenskim okvirima (*engl. Frame*) gdje svaki okvir ima unaprijed određeno trajanje ovisno o frekvenciji izračuna značajki. Tijekom svakog okvira, kognitivno-emocionalno stanje se procjenjuje iz fizioloških podataka prikupljenih u prethodnom vremenskom prozoru [7]. U ovom radu frekvencija izračuna značajki na segmentu od 12 sekundi trajanja pojedinog događaja, iznosi 2Hz. Ovo je implementirano tako da se uzme svaki gore spomenuti segment od 12 sekundi svakog događaja te se unutar tog segmenta s pomakom od 0.5 sekundi vrši izračun značajki na prozoru širine 5 sekundi dok se ne dosegne kraj segmenta od 12 sekundi. Ovakav pristup sličan je pristupu raspoznavanja emocija na temelju analize fizioloških signala izloženom u radu [7].

3.1. Izračun značajki

Za svakog pilota i za svaki senzorski modalitet, odnosno polje podataka u podatkovnom skupu, izračunavaju se značajke navedene u tablici (Tablica 2) koje se koriste u raspoznavanju kognitivno-emocionalnog stanja pilota.

Tablica 2 Popis značajki te polja podataka iz kojih su izračunate

Polje podataka	Značajka	Broj značajki
ECG	<i>HeartRateMean, HeartRateStd</i>	2
R	<i>RespirationMean, RespirationStd</i>	2
GSR	<i>GSR_std, GSR_mean, GSR_slope</i>	3
EEG	<i>Alpha, Beta, Beta/Alpha, Engagement index, Asymmetry index</i>	5
Ukupno		12

Pomoću funkcije *nk.ecg_process()* iz biblioteke *NeuroKit2* obrađuje se ECG segment signala iz polja podataka, uz frekvenciju uzorkovanja od 256Hz. Ova funkcija detektira vrhove R valova (*engl. R Peaks*), odnosno glavnih vrhova u ECG signalu koji predstavljaju otkucaje srca i iz njih izračunava trenutni srčani ritam [8]. Funkcija vraća podatkovni okvir koji sadrži niz stupaca od kojih se koristi *ECG_Rate* koji prikazuje promjene u srčanom ritmu tijekom zadanog vremenskog okvira (izraženo u otkucajima po minuti). Izračunavanjem srednje vrijednosti *ECG_Rate* stupca dobivamo prosječnu brzinu otkucaja srca pilota tijekom zadanog vremenskog razdoblja, a izračunavanjem standardne devijacije ovog stupca analiziramo koliko se vrijednost srčanog ritma razlikuju od prosječne vrijednosti, tj. veća standardna devijacija ukazuje da postoji veća varijabilnost u srčanom ritmu.

Za izračun srednje vrijednosti disanja te standardnu devijaciju disanja koristi se *NumPy* biblioteka. *RespirationMean* predstavlja prosječnu dubinu disanja tijekom analiziranog vremenskog segmenta. Veće vrijednosti ukazuju na dublje disanje, dok manje vrijednosti ukazuju na pliće disanje. *RespirationStd* pokazuje koliko se amplituda disanja razlikuje od prosječne vrijednosti. Visoka vrijednost ukazuje na prisutne značajne promjene u dubini disanja (npr. variranje između dubokih i plitkih udisaja). Niska vrijednost ukazuje na stabilno i ujednačeno disanje. S obzirom na to da respiracijski signal oscilira dinamikom broja učinjenih udisaja po minuti, ako je vremenski prozor na kojemu se ove značajke promatraju kraći od jednog respiracijskog ciklusa, vrijednosti ovih značajki dakako ovise o tome koji je dio respiracijskog ciklusa zahvaćen u dotičnom prozoru.

Srednja vrijednost, standardna devijacija te nagib GSR segmenta ukazuju na elektrodermalnu aktivnost pilota. Ukoliko je srednja vrijednost viša, kao i standardna devijacija to može označavati stanje povećane emocionalne pobuđenosti, dok niže vrijednosti mogu ukazivati na opušteno, tj. Bazno stanje pilota. Ukoliko je nagib pozitivan, GSR signal raste što može ukazivati na povećanu pobuđenost. S druge strane, negativne vrijednosti nagiba možemo tumačiti kao stanje opuštanja nakon stresne situacije.

EEG spektar je podijeljen na 5 intervala frekvencija:

1. δ - 0.5-4 Hz
2. θ - 4-8 Hz
3. α - 8-13 Hz
4. β - 13-30 Hz
5. γ - 30-70 Hz.

Za vrijeme baznog stanja i stanja odmora dominiraju α valovi, dok za vrijeme veće pobuđenosti i stresa dominiraju β valovi [15]. Nadalje, istraživanja pokazuju da je u stanju stresa vidljiva povećana aktivnost α valova na desnoj strani mozga u odnosu na lijevu [16].

Značajka *Alpha* označava prosječnu snagu α frekvencija, dok *Beta* označava prosječnu snagu β frekvencija kroz sljedeće EEG kanale: EEG_FP1, EEG_FP2, EEG_F3, EEG_F4, EEG_F7 te EEG_F8. Značajka *Alpha/Beta* pokazuje odnos snage beta i alfa valova. Viši omjer pokazatelj je povećanog mentalnog opterećenja, dok niži omjer obično ukazuje na opuštenije stanje i nižu kognitivnu aktivnost. *Engagement index* je pokazatelj stupnja mentalne angažiranosti ili pažnje, a računa se prema formuli:

$$\text{Engagement index} = \frac{\text{Beta}}{\text{Alpha} + \text{Theta}}$$

Više vrijednosti upućuju na veću razinu pozornosti i mentalne angažiranosti, dok niže vrijednosti ukazuju na stanje niže pažnje ili opuštenosti. *Asymmetry index* predstavlja asimetriju alfa aktivnosti između lijeve (EEG kanal EEG_F3) i desne (EEG kanal EEG_F4) hemisfere mozga prema formuli

$$\text{Asymmetry index} = \ln \text{Alpha}_L - \ln \text{Alpha}_D$$

Pozitivne vrijednosti ukazuju na veću aktivaciju desne hemisfere, dok negativne vrijednosti ukazuju na veću aktivaciju lijeve hemisfere. Snage α β i θ valova izračunate su pomoću funkcije `nk.eeg_power()` iz biblioteke *NeuroKit2*.

3.2. Podjela podatkovnog skupa na skup za treniranje i testiranje

U radu su implementirane dvije podjele na skup podataka za treniranje triju modela strojnog učenja te na skup za njihovo testiranje:

1. Stratificirana nasumična podjela – Ova metoda omogućava nasumičnu podjelu podataka na skup za trening (80% podataka) te na testni skup (20% podataka). Početna ideja ove metode bila je zadržati proporciju klasa, tj. različitih događaja u podatkovnom skupu, kako bi oba skupa podataka imala uravnoteženu podjelu događaja. Međutim, primijećeno je da je podatkovni skup neuravnotežen, odnos klasa 0:1:2:5 je otprilike 1:0.31:0.45:0.68. Ovo može rezultirati modelom strojnog učenja koji „favorizira“ najučestalije klase, tj. u ovom slučaju događaj 0. Navedeni problem riješen je korištenjem tehnike SMOTE (*engl. Synthetic Minority Oversampling Technique*) nad skupom za treniranje. SMOTE generira sintetičke primjere manjinskih klasa u skupu za treniranje i na taj način smanjuje neuravnoteženost ovog skupa te povećava šansu za boljim učenjem manjinskih klasa.
2. *Leave One Subject Out* (LOSO) podjela – LOSO metoda se koristi za generalizaciju modela na nove subjekte, odnosno osigurava da se model testira na podacima pilota koje nije vidio u treningu. Umjesto slučajne podjele podataka, u svakoj iteraciji podaci jednog pilota se koriste za testiranje, dok se podaci svih ostalih pilota koriste za treniranje modela. U ovoj podjeli se također koristi tehnika SMOTE kako bi se izbalansirale klase koje predviđamo, tj. događaji.

4. Metodologija strojnog učenja

U ovom radu istražene su tri različite metode strojnog učenja za klasifikaciju događaja: stablo odluke (*engl. Decision Tree*), slučajna šuma (*engl. Random Forest*) i metoda k najbližih susjeda (*engl. KNN - K-Nearest Neighbors*).

Stablo odluke je hijerarhijski model koji se temelji na donošenju odluka koristeći uvjete postavljene na značajkama podataka. Model gradi stablo gdje svaki čvor predstavlja uvjet, a grane predstavljaju moguće ishode [10]. U ovom radu hiperparametar *max_depth*, tj. maksimalna dubina stabla, optimiziran je korištenjem metode *GridSearchCV* na skupu za treniranje kako bi se pronašla najbolja vrijednost dubine stabla. Slučajna šuma je metoda koja koristi više stabala odluke kako bi poboljšala robusnost modela. Svako stablo u šumi trenira se na podskupu podataka korištenjem tehnike *bootstrap* uzorkovanja (*engl. bootstrap sampling*), a konačna odluka donosi se glasanjem svih stabala [11]. Kako bi se smanjilo vrijeme treniranja, fiksno su postavljene vrijednosti hiperparametara: *max_depth=10* (maksimalna dubina stabla) i *n_estimators=50* (broj stabala). Metoda K najbližih susjeda (KNN) klasificira nove primjere na temelju sličnosti s poznatim podacima. Model izračunava udaljenost između novih podataka i postojećih primjera te donosi odluku na temelju većinske klase među k najbližih susjeda [12]. Optimalna vrijednost k u rasponu od 1 do 21 odabrana je korištenjem metode *GridSearchCV* na skupu za treniranje kako bi se pronašao najbolji broj susjeda za klasifikaciju.

U radu su implementirane i dvije varijante klasifikacije:

- Višeklasna klasifikacija - model trenira predviđanje više različitih događaja (0, 1, 2, 5).
- Binarna klasifikacija – model razlikuje događaj 0 od svih ostalih događaja. Pri ovakvoj klasifikaciji nije bilo potrebno balansirati klase u skupu za trening jer je skup manje neuravnotežen (1:2).

U modelu koji koristi stratificiranu nasumičnu podjelu na testni skup i skup za trening, implementirana je varijanta u kojoj su se koristile i samo značajke vezane za ECG, GSR i disanje (*HeartRateMean*, *HeartRateStd*, *RespirationMean*, *RespirationStd*, *GSR_std*, *GSR_mean*, *GSR_slope*). Ovakva analiza nije provedena na modelu s LOSO podjelom jer bi zbog većeg broja iteracija trajala predugo. Rezultati ove analize omogućuju usporedbu između modela koji koristi sve dostupne značajke i modela koji koristi samo spomenute

značajke te pruža uvid u doprinos spomenutih značajki na točnosti modela i je li moguće postići dobru klasifikaciju koristeći samo ograničen skup fizioloških značajki. Sve navedeno implementirano je u razvojnom okruženju *PyCharm*, korištenjem sučelja *Jupyter Notebook* i programskog jezika *Python 3.11.1*.

Za procjenu uspješnosti modela korištene su sljedeće metrike evaluacije: točnost (*engl. accuracy*), preciznost (*engl. precision*), odziv (*engl. recall*), te matrica zabune (*engl. confusion matrix*). Preciznost i odziv u višeklasnoj klasifikaciji izračunati su korištenjem makro prosjeka (*engl. macro-average*), što znači da su ove metrike izračunate zasebno za svaku klasu, a zatim je uzet prosjek vrijednosti tih metrika preko svih klasa.

5. Rezultati i diskusija

Ovo poglavlje donosi pregled rezultata po modelu strojnog učenja, varijanti klasifikacije (više-klasno/binarno) te načinu podjele podataka na skup za trening i test. Kako je već spomenuto, modeli se uspoređuju na temelju njihove točnosti, preciznosti, odaziva te matrice zabune. Točnost *majority-class* klasifikatora računa se prema formuli:

$$Accuracy_{majority} = \frac{\text{broj primjera najčešće klase}}{\text{ukupan broj primjera}}$$

U slučaju više-klasne klasifikacije pri stratificiranoj nasumičnoj podjeli ova točnost iznosi 40.94%, dok u slučaju binarne klasifikacije pri istoj podjeli iznosi 59.06%.

5.1. Rezultati metode stabla odluke

U slučaju stratificirane nasumične podjele podataka, binarna klasifikacija ove metode donosi nešto bolje rezultate od više-klasne podjele. Tablica 3 pokazuje rezultate modela uzimajući u obzir varijantu klasifikacije te značajke korištene za raspoznavanje kognitivno-emocionalnog stanja pilota. Najbolji rezultati u ovom slučaju postižu se korištenjem binarne klasifikacije nad odabranim fiziološkim značajkama. Iako je razlika između rezultata nad svim značajkama i odabranim značajkama mala, korištenjem manjeg broja značajki smanjujemo količinu podataka za trening, a tako i samo trajanje treniranja modela.

Tablica 3 Rezultati modela stabla odluke na stratificiranoj nasumičnoj podjeli podataka

		Sve značajke	<i>HeartRateMean, HeartRateStd, RespirationMean, RespirationStd, GSR_std, GSR_mean, GSR_slope</i>
Više-klasna klasifikacija	Točnost	0.75	0.77
	Preciznost	0.74	0.76
	Odziv	0.75	0.76

Binarna klasifikacija	Točnost	0.84	0.88
	Preciznost	0.85	0.89
	Odziv	0.89	0.91

Iz konfuzijskih matrica (Slika 3 i Slika 4) očitano je da pri višeklasnoj klasifikaciji model koji koristi odabrane značajke nešto bolje predviđa događaj 0 i 5 (za događaj 0: 3 više točnih predviđanja; za događaj 5: 18 više točnih predviđanja), dok model sa svim značajkama nešto bolje predviđa događaje 1 i 2 (za događaj 1: 6 više točnih predviđanja; za događaj 2: 3 više točnih predviđanja). Ove razlike su zanemarive.

235	12	23	35
10	74	4	8
15	7	105	10
27	10	17	153

Slika 3 Matrica zabune modela stabla odluke s višeklasnom klasifikacijom korištenjem svih značajki pri stratificiranoj nasumičnoj podjeli podataka

238	15	21	31
11	68	3	14
14	5	102	16
16	11	9	171

Slika 4 Matrica zabune modela stabla odluke s višeklasnom klasifikacijom korištenjem odabranih značajki pri stratificiranoj nasumičnoj podjeli podataka

U slučaju modela koji koristi LOSO podjelu podatkovnog skupa, rezultati su daleko lošiji. Tablica 4 prikazuje srednju vrijednost, standardnu devijaciju i medijan točnosti, preciznosti i odziva 18 modela.

Tablica 4 Rezultati modela stabla odluke na LOSO podjeli podataka

		Srednja vrijednost	Standardna devijacija	Medijan
Višeklasna klasifikacija	Točnost	0.31	0.067	0.31
	Preciznost	0.27	0.084	0.28
	Odziv	0.27	0.064	0.24
Binarna klasifikacija	Točnost	0.52	0.12	0.53
	Preciznost	0.62	0.31	0.71
	Odziv	0.28	0.18	0.30

Binarnom klasifikacijom postižu se bolji rezultati, ali veće standardne devijacije ukazuju na veliku varijabilnost kroz iteracije. Tako npr. model koji za testni skup uzima podatke od pilota br. 12 daje rezultate: točnost 0.70, preciznost 0.87 i odziv 0.51, dok model koji testira podatke pilota 6 daje rezultate: točnost 0.31, preciznost 0.00 i odziv 0.00.

5.2. Rezultati metode slučajne šume

Najbolji rezultati ove metode postižu se korištenjem binarne klasifikacije što pokazuje Tablica 5. Razlika u rezultatima pri korištenju svih značajki za treniranje ili samo onih vezanih za ECG, GSR i disanje je vrlo mala kao i u prethodnoj metodi. To ukazuje na to da su ove značajke najkorisnije pri predviđanju stanja pilota, tj. događaja. Rezultati modela s LOSO podjelom podataka (Tablica 6) su puno slabiji te, iako binarna klasifikacija bolje predviđa, visoke standardne devijacije ukazuju na veliku varijabilnost među pilotima te da model ne generalizira dobro na nove subjekte.

Tablica 5 Rezultati modela slučajne šume na stratificiranoj nasumičnoj podjeli podataka

		Sve značajke	<i>HeartRateMean, HeartRateStd, RespirationMean, RespirationStd, GSR_std, GSR_mean, GSR_slope</i>
Višeklasna klasifikacija	Točnost	0.78	0.75
	Preciznost	0.79	0.75
	Odziv	0.78	0.75
Binarna klasifikacija	Točnost	0.86	0.86
	Preciznost	0.90	0.87
	Odziv	0.85	0.89

Tablica 6 Rezultati modela slučajne šume na LOSO podjeli podataka

		Srednja vrijednost	Standardna devijacija	Medijan
Višeklasna klasifikacija	Točnost	0.38	0.079	0.39
	Preciznost	0.31	0.11	0.30
	Odziv	0.33	0.065	0.32
Binarna klasifikacija	Točnost	0.57	0.11	0.60
	Preciznost	0.66	0.13	0.69
	Odziv	0.61	0.21	0.64

5.3. Rezultati KNN metode

U slučaju KNN metode strojnog učenja, Tablica 7 pokazuje da su značajke vezane za EEG ipak korisne za preciznije raspoznavanje kognitivno-emocionalnog stanja pilota, tj. situacije u kojoj se nalazi. Uspoređujući sve tri metode po navedenim metrikama, ova

metoda pruža najbolje rezultate pri stratificiranoj podjeli podataka. Također, pri ovoj podjeli podataka, višeklasna i binarna klasifikacija daju gotovo jednake rezultate.

Tablica 7 Rezultati modela KNN na stratificiranoj nasumičnoj podjeli podataka

		Sve značajke	<i>HeartRateMean, HeartRateStd, RespirationMean, RespirationStd, GSR_std, GSR_mean, GSR_slope</i>
Višeklasna klasifikacija	Točnost	0.87	0.79
	Preciznost	0.87	0.78
	Odziv	0.87	0.80
Binarna klasifikacija	Točnost	0.90	0.85
	Preciznost	0.91	0.85
	Odziv	0.93	0.91

Tablica 8 Rezultati modela KNN na LOSO podjeli podataka

		Srednja vrijednost	Standardna devijacija	Medijan
Višeklasna klasifikacija	Točnost	0.33	0.057	0.32
	Preciznost	0.28	0.060	0.30
	Odziv	0.29	0.069	0.29
Binarna klasifikacija	Točnost	0.56	0.080	0.54
	Preciznost	0.63	0.097	0.62
	Odziv	0.64	0.15	0.66

Kao i kod prethodnih metoda, rezultati modela pri LOSO podjeli su znatno lošiji (Tablica 8). Međutim, za razliku od prethodne dvije metode, KNN je konzistentniji u različitim iteracijama, ali lošiji rezultati ukazuju na to da ni ovaj model ne generalizira dobro.

Zaključak

U ovom radu istražene su mogućnosti prepoznavanja visokorizičnih kognitivno-emocionalnih stanja pilota korištenjem strojnog učenja na temelju fizioloških značajki povezanih s EEG-om, ECG-om, GSR-om i disanjem. Implementirane su tri metode strojnog učenja: stablo odluke, slučajna šuma i metoda KNN. Također, primijenjena su dva različita načina podjele podataka na skup za trening i testni skup: stratificirana nasumična podjela i LOSO podjela. Dodatno, analiziran je i utjecaj korištenja samo odabranih fizioloških značajki (ECG, GSR i disanje) na rezultate modela.

Rezultati pokazuju da modeli trenirani nad podacima podijeljenim stratificiranom nasumičnom podjelom postižu daleko bolje rezultate u odnosu na modele koji koriste LOSO podjelu. Ovo je slučaj kod svakog implementiranog modela strojnog učenja što ukazuje na izazov generalizacije modela na nove pilote zbog njihovih individualnih razlika. Izvori individualnih razlika između pilota mogu biti raznoliki, od njihovih bazičnih osobina i dobi do razlika u njihovom domenskom znanju i pilotskom iskustvu [13]. Također, binarna klasifikacija svakog modela daje bolje rezultate od višeklasne, što ukazuje na to da je pouzdanije prepoznati kada pilot ulazi u bilo koje potencijalno rizično stanje nego precizno odrediti o kojem specifičnom stanju se radi. Kod višeklasne klasifikacije, očite su značajne greške modela pri razdvajanju određenih stanja, što upućuje na potrebu za daljnjim unaprjeđenjem modela.

Općenito, najbolje rezultate dao je KNN model s binarnom klasifikacijom i korištenjem svih značajki za učenje pri stratificiranoj nasumičnoj podjeli podatkovnog skupa. Što se tiče višeklasne klasifikacije pri istoj podjeli podataka, najbolje rezultate je također dao model KNN s korištenjem svih značajki za učenje. Najbolje rezultate pri LOSO podjeli u višeklasnoj klasifikaciji ostvarila je metoda slučajne šume, ali s većom standardnom devijacijom pri računanju preciznosti. U binarnoj klasifikaciji pri LOSO podjeli najbolje rezultate dala je KNN metoda jer ima najniže standardne devijacije za sve metrike pa je i najstabilnija kroz iteracije po pilotima. Također, pokazalo se da korištenje samo fizioloških značajki ECG-a, GSR-a i disanja ne dovodi do značajnog smanjenja performansi modela, što sugerira da su ove značajke najrelevantnije za raspoznavanje kognitivno-emocionalnih stanja pilota zastupljenih u raspoloživom podatkovnom skupu.

Budući rad mogao bi se usmjeriti na poboljšanje modela višeklasne klasifikacije te istraživanje dubljih uzroka varijabilnosti modela među pilotima. Također, mogla bi se istražiti mogućnost poboljšanja točnosti prepoznavanja visokorizičnih kognitivno-emocionalnih stanja pilota na način da se, umjesto LOSO podjele, u skup za trening uključi i dio (primjerice prvih 50 posto) podataka preostalog pojedinca na kojem se model evaluira.

Literatura

- [1] Harrivel, Angela & Stephens, Chad & Milletich, Robert & Heinich, Christina & Last, Mary & Napoli, Nicholas & Abraham, Nijo & Prinzel, Lance & Motter, Mark & Pope, Alan. (2017). *Prediction of Cognitive States during Flight Simulation using Multimodal Psychophysiological Sensing*. 10.2514/6.2017-1135.
- [2] Crijnen, J. *Predicting a Pilot's Cognitive State from Physiological Measurements*. Diplomski rad, Tilburg University, 2019.
- [3] Lin, Y. *Reducing Aviation Fatalities by Monitoring Pilots' Cognitive States Using Psychophysiological Measurements*. Diplomski rad, Naval Postgraduate School, 2021.
- [4] Amie Johnson, Jack Sarle, Maggie, Paul Terwilliger, Shannon Walker, Will Cukierski, and William Espenschied. *Reducing Commercial Aviation Fatalities*. <https://kaggle.com/competitions/reducing-commercial-aviation-fatalities>, 2018. Kaggle.
- [5] NASA's Open Data Portal. *Flight Crew Physiological Data for Crew State Monitoring*. https://data.nasa.gov/dataset/Flight-Crew-Physiological-Data-for-Crew-State-Moni/8a5d-xhfr/about_data, 2024.
- [6] Kristensen, A. N., Jeyam, B., Riahi, S., & Jensen, M. B. (2016). *The use of a portable three-lead ECG monitor to detect atrial fibrillation in general practice*. *Scandinavian journal of primary health care*, 34(3), 304–308.
- [7] Kukolja, D., Popović, S., Dropuljić, B., Horvat, M., & Ćosić, K. (2009). *Real-Time Emotional State Estimator for Adaptive Virtual Reality Stimulation*. Interacción.
- [8] Makowski, D. (2021). *Neurophysiological Data Analysis with NeuroKit2*. NeuroKit. <https://neuropsychology.github.io/NeuroKit/>
- [9] Diarra, M., Marchitto, M., Bressolle, M. C., Baccino, T., & Draï-Zerbib, V. (2023). *A narrative review of the interconnection between pilot acute stress, startle, and surprise effects in the aviation context: Contribution of physiological measurements*. *Frontiers in neuroergonomics*, 4, 1059476
- [10] IBM. *What is a decision tree?*. <https://www.ibm.com/think/topics/decision-trees.2025>.
- [11] IBM. *What is random forest?*. <https://www.ibm.com/think/topics/random-forest.2025>.
- [12] IBM. *What is the k-nearest neighbors (KNN) algorithm?*. <https://www.ibm.com/think/topics/knn.2025>.
- [13] Hardy, D. J., & Parasuraman, R. (1997). *Cognition and flight performance in older pilots*. *Journal of Experimental Psychology: Applied*, 3(4), 313.
- [14] Mathavara, K., & Ramachandran, G. (2022). *Role of Human Factors in Preventing Aviation Accidents: An Insight*. IntechOpen. doi: 10.5772/intechopen.106899
- [15] Hayashi, Takuto & Okamoto, Eika & Nishimura, Haruhiko & Mizuno-Matsumoto, Yuko & Ishii, Ryouhei & Ukai, Satoshi. (2013). *Beta Activities in EEG Associated*

with Emotional Stress. International Journal of Intelligent Computing in Medical Sciences & Image Processing. 3. 57-68.

- [16] Acharya, U & Subbhuraam, Vinitha Sree & Ang, Peng & Yanti, Ratna & Suri, Jasjit. *Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals*. International journal of neural systems. 22. 2012.

Sažetak

Raspoznavanje visokorizičnih mentalnih stanja pilota na temelju fizioloških značajki i strojnog učenja

Ljudska greška je jedan od glavnih uzroka zrakoplovnih nesreća, stoga pravovremeno prepoznavanje visokorizičnih mentalnih stanja pilota može značajno doprinijeti sigurnosti letova. Ovaj rad istražuje metode strojnog učenja za klasifikaciju kognitivno-emocionalnih stanja pilota reprezentiranih u javno dostupnom podatkovnom skupu, koristeći fiziološke značajke vezane za elektroencefalogram (EEG), elektrokardiogram (EKG), galvansku reakciju kože (GSR) i disanje. Implementirane su i analizirane tri metode strojnog učenja: stablo odluke, slučajna šuma i KNN. Koriste se dvije vrste klasifikacije: višeklasna i binarna. Učinkovitosti modela uvelike su ovisile o načinu podjele podatkovnog skupa na skup za trening i skup za test. Slučajna šuma bila je najbolji klasifikator u uvjetima višeklasne klasifikacije pri Leave-One-Subject-Out (LOSO) podjeli, a KNN u uvjetima binarne klasifikacije pri LOSO podjeli te binarne i višeklasne klasifikacije pri stratificiranoj nasumičnoj podjeli skupa podataka.

Ključne riječi: pilot, mentalno stanje, fiziološke značajke, senzor, strojno učenje, simulator leta, pažnja.

Summary

Recognition of pilots' high-risk mental states based on physiological features and machine learning

Human error is one of the leading causes of aviation accidents; therefore, timely recognition of high-risk mental states in pilots can significantly contribute to flight safety. This study explores machine learning methods for classifying pilots' cognitive-emotional states represented in a publicly available dataset, using physiological features related to electroencephalography (EEG), electrocardiography (ECG), galvanic skin response (GSR), and respiration. Three machine learning methods were implemented and analyzed: decision tree, random forest, and KNN. Two types of classification were used: multiclass and binary. The effectiveness of the models largely depended on the method used for splitting the dataset into training and test sets. Random forest was the best classifier in case of multi-class classification under the Leave-One-Subject-Out (LOSO) split, while KNN performed best in binary classification under the LOSO split, as well as in both binary and multi-class classification under the stratified random split of the dataset.

Keywords: pilot, mental state, physiological features, sensor, machine learning, flight simulator, attention.